Modelling approaches for histology-independent cancer drugs to inform NICE appraisals

|  |  |
| --- | --- |
| **Produced by** | CRD and CHE Technology Assessment Group, University of York, Heslington, York YO10 5DD  ScHARR Technology Assessment Group, University of Sheffield, Sheffield, S1 4DA |
| **Authors**  **Correspondence to**  **Date completed** | Peter Murphy, Research Fellow, CRD, University of York  David Glynn, Research Fellow, CHE, University of York  Sofia Dias, Professor, CRD, University of York  Rob Hodgson, Research Fellow, CRD, University of York  Lindsay Claxton, Research Fellow, CRD, University of York  Lucy Beresford, NIHR Research Training Fellow, CRD, University of York  Katy Cooper, Senior Research Fellow, ScHARR, University of Sheffield  Paul Tappenden, Professor, ScHARR, University of Sheffield  Kate Ennis, Research Associate, ScHARR, University of Sheffield  Alessandro Grosso, Research Fellow, CHE, University of York  Kath Wright, Information Service Manager, CRD, University of York  Anna Cantrell, Information Specialist, ScHARR, University of Sheffield  Matt Stevenson, Professor, ScHARR, University of Sheffield  Stephen Palmer, Professor, CHE, University of York  Stephen Palmer, CHE, University of York  Email: [stephen.palmer@york.ac.uk](mailto:stephen.palmer@york.ac.uk)  28/02/2020 |

**Source of funding**

**This report was commissioned by the National Institute for Health Research (NIHR) HTA Programme as project number 127582.**

Declared competing interests of the authors

All authors have no personal or pecuniary conflict of interests.

Acknowledgements

We acknowledge and thank Jacoline Bouvy (Senior Scientific Adviser, NICE) and Sophie Cooper (Scientific Adviser, NICE) for their advice throughout the project and comments on the report. The report is based on an assessment of a hypothetical case-study and therefore the data generated is not suitable for sharing beyond that contained within the report. Further information can be obtained from the corresponding author.

Disclaimer

This report presents independent research funded by the NIHR. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

This version of the report has been submitted to the NIHR and is awaiting peer review. The final version of this report will be available from the NIHR website upon publication: https://www.journalslibrary.nihr.ac.uk/hta/#/

**Contribtion of authors**

Peter Murphy contributed to the review of issues for cost-effectiveness, the application of the Bayeian hierarchical approach, the exemplar case study and wrote sections of the report.

David Glynn developed the decision-framework, undertook the analyses using the case study and wrote sections of the report.

Sofia Dias had overall responsibility for the clinical effectiveness sections of the report. She contributed to the protocol, the review of statistical literature addressing the design and analysis of histology-independent trials and the writing of the report.

Rob Hodgson contributed to the protocol, the review of issues for cost-effectiveness, diagnostic testing issues, the development of the exemplar case study and the writing of the report.

Lindsay Claxton contributed to the review of issues for cost-effectiveness and the writing of the report.

Lucy Beresford contributed to the review of diagnostic testing issues and the writing of the report.

Katy Cooper contributed to the systematic review of surrogate endpoints and the writing of the report.

Paul Tappenden contributed to the protocol, the systematic review of surrogate endpoints and the writing of the report.

Kate Ennis contributed to the systematic review of surrogate endpoints and the writing of the report.

Alessandro Grosso contributed to the review of regulatory guidance and the review of previous NICE appraisals.

Kath Wright undertook the searching for studies and managing references and wrote sections of the report.

Anna Cantrell undertook the seaching for studies and managing references for the review of surrogates and wrote sections of the report.

Matt Stevenson contributed to the protocol and commented on all sections of the report.

Stephen Palmer had overall responsibility for the project. He contributed to the protocol and to all aspects of the work including the writing of the report.

Abbreviations

|  |  |
| --- | --- |
| AD | Aggregate data |
| ALK | Anaplastic lymphoma kinase |
| AML | Acute myeloid leukameia |
| BHM | Bayesian hierarchical model |
| BICR | Blinded independent central radiologist |
| BRMA | Bivariate random effects meta-analysis |
| BSES | Biomarker-Surrogate Evaluation Schema criteria |
| BSWP | Biostatistics working party |
| BTC | Biliary tract cancer |
| CADTH | Canadian Agency for Drugs and Technologies in Health |
| CDF | Cancer Drugs Fund |
| CHE | Centre for Health Economics |
| CHF | Congestive heart failure |
| CHMP | Committee for Medicinal Products for Human Use |
| CI | Confidence interval |
| CMN | Congenital mesoblastic nephroma |
| CNS | Central nervous system |
| CR | Complete response |
| CRC | Colorectal cancer |
| CRD | Centre for Reviews and Dissemination |
| CrI | Credible interval |
| ctDNA | Circulating tumour DNA |
| dMMR | deficient Mismatch repair |
| DoR | Duration of response |
| DSU | Decision Support Unit |
| EAP | Early Access Programme |
| ECOG | Eastern Cooperative Oncology Group |
| EGFR | Epidermal Growth Factor Receptor |
| EMA | European Medicines Agency |
| ePAS | Extended patient analaysis set |
| ERG | Evidence Review Group |
| ESMO | European Society for Medical Oncology |
| EVPI | Expected value of perfect information |
| EVPPI | Expected value of partial perfect information |
| EVSI | Expected value of sample information |
| FDA | Food and Drug Administration |
| FISH | Fluorescence *in situ* hybridization |
| FL | First line |
| GI | Gastrointestinal |
| GIST | Gastrointestinal stromal tumour |
| GMI | Growth modulation index |
| HR | Hazard ratio |
| HRQoL | Health-related quality of life |
| HTA | Health technology assessment |
| ICER | Incremental cost-effectiveness ratio |
| IFS | Infantile fibrosarcoma |
| IHC | Immunohistochemistry |
| IMF | International Myeloma Foundation |
| IPD | Individual patient data |
| IQWiG | Institute for Quality and Efficiency in Health Care |
| IRC | Independent Review Committee |
| ITT | Intention-to-treat |
| KRAS | Kirsten rat sarcoma |
| LYG | Life years gained |
| MAA | Managed access agreement |
| MAIC | Matching-adjusted indirect comparison |
| MASC | Mammary analogue secretory carcinoma |
| MSI-H | Microsatellite instability high |
| NE | Not estimable |
| NET | Neuroendocrine tumour |
| NGS | Next generation sequencing |
| NHB | Net health benefit |
| NHL | Non-Hodgkin’s lymphoma |
| NHS | National Health Service |
| NICE | National Institute for Health and Care Excellence |
| NIH | National Institutes of Health |
| NMB | Net monetary benefit |
| NNS | Number needed to screen |
| NR | Not reported |
| NSCLC | Non-small cell lung cancer |
| NTRK | Neurotrophic tyrosine receptor kinase |
| ODAC | Oncologic Drugs Advisory Committee (ODAC) |
| OR | Odds ratio |
| ORR | Overall response rate |
| OS | Overall survival |
| PAS | Primary analysis set |
| PCR | Polymerase chain reaction |
| PD | Progressive disease |
| PFS | Progression-free survival |
| PR | Partial response |
| PSA | Probabilistic sensitivity analysis |
| PSM | Partioned survival model |
| QALY | Quality-adjusted life year |
| RCC | Renal cell carcinoma |
| RCP | Royal College of Pathologists |
| RCT | Randomised controlled trial |
| RECIST | Response Evaluation Criteria in Solid Tumours |
| RR | Relative risk |
| RRcHL | Relapsed or refractory classical Hodgkin lymphoma |
| RTK | Receptor tyrosine kinase |
| RT-PCR | Reverse transcription polymerase chain reaction |
| RWD | Real world data |
| SA | Single arm |
| SAG | Scientific advisory group |
| SAP | Statistical analysis plan |
| ScHARR | School of Health and Related Research |
| SCLC | Small cell lung cancer |
| SmPC | Summary of product characteristics |
| SOC | Standard of care |
| STC | Soft tissue sarcoma/Simulated Treatment Comparison |
| STE | Surrogate threshold effect |
| TA | Technology Appraisal |
| TKI | Tyrosine kinase inhibitor |
| TSD | Technical support document |
| TTE | Time to event |
| TTP | Time to progression |
| TTR | Time to response |
| VGPR | Very good partial response |
| VoH | Value of heterogeneity |
| VOI | Value of information |
| WGS | Whole genome sequencing |

Abstract

The first histology-independent marketing authorisation was granted by the European Medicines Agency in 2019. This is the first time a cancer treatment has been approved based on a common biomarker rather than the location in the body where the tumour originated. This represents an important paradigm shift, meaning that oncological diseases can now be classified by either tumour biomarker status or tumour histogenesis. This research aims to expore the implications for the National Institute for Health and Care Excellence (NICE) appraisal process.

A series of targeted reviews were undertaken to determine the type of evidence that is likely to be available at the point of marketing authorisation and to consider the analyses likely to be required to support NICE appraisals. Several challenges were identified concerning the design and conduct of trials used to support histology-independent products, the greater levels of heterogeneity within the licensed population and the use of surrogate endpoints. We identified possible approaches to address these challenges by reviewing key statistical literature addressing the design and analysis of histology-independent trials and undertaking a systematic review to evalue the use of response endpoints as surrogate outcomes for progression-free (PFS) and overall survival (OS).

We found that the potential for heterogeneity in treatment effects, either across tumour types or across other characteristics, is likely to be a central issue for NICE appraisals. Bayesian hierarchical methods may provide a useful vehicle with which to explore any heterogeneity. Our review suggests that response endpoints may not be reliable surrogates for PFS or OS. However, we concluded that a surrogate-based modelling approach, which captures all relevant uncertainty, may be preferable to the use of immature survival data.

Several additional sources of heterogeneity were also identified as presenting potential challenges to NICE appraisal, including: the cost of testing; baseline risk; quality of life and routine management costs. We concluded that a range of alternative approaches will be required to address different sources of heterogeneity to support NICE appraisals. An exemplar case study was developed to illustrate the nature of the assessments that may be required.

We developed a decision-framework that could be used to inform approval and research policies for histology-independent products. The framework explored the uncertainties and risks associated with different approval policies. Alternative approaches to managing risk were identified, including the role of further data collection, the use of pricing schemes and stratified decision making. We concluded that routine presentation of the scale of the consequences of heterogeneity and decision uncertainty may provide an important additional approach to the assessments specified in the current NICE methods guide.

We identified several key areas requiring further research. Firstly, further exploration of Bayesian hierarchical methods could help inform decision makers determine whether these is sufficient evidence of homogeneity to support pooled analyses. Secondly, further research is required to determine the appropriate basis for apportioning genomic testing costs where there are multiple targets. Finally, further research is required concerning the challenges of uncontrolled Phase II studies and specifically the role and use of surrogate endpoints.

Scientific Summary

**Background**

In May 2017, the U.S. Food and Drug Administration (FDA) granted accelerated approval to pembrolizumab for treating solid tumours with the microsatellite instability high (MSI-H) or mismatch repair deficient (dMMR) biomarker. This is the first time a cancer treatment has been approved based on a common biomarker rather than the location in the body where the tumour originated. This represents an important paradigm shift, meaning that oncological diseases can now be classified by either tumour biomarker status or tumour histogenesis. The first histology-independent marketing authorisation was granted by the European Medicines Agency (EMA) in 2019.

A histology-independent marketing authorisation will include a large number of individual tumour sites. It is unlikely to be feasible or desirable for the National Institute of Health and Care Excellence (NICE) to conduct separate appraisals for each individual tumour site to inform whether approval of these products represents an efficient use of NHS resources. However, the scope of histology-independent indications and the nature of the evidence base for this type of appraisal will pose important challenges to the appropriate quantifion of their value to the NHS and the effective mitigation of any additional risks. NICE needs to consider how to develop a process/approach that will allow a single, biomarker-driven appraisal for histology-independent cancer drugs.

This research aims to inform future NICE policy on how to appraise cancer drugs with histology-independent indications.

**Objectives**

We saught to explore the potential implications of histology-indepdent products for the NICE technology appraisal (TA) process. The specific objectives were:

1. To identify the nature of the evidence likely to be available at initial marketing authorisation.
2. To determine the types of evidence and analyses required to support NICE appraisal.
3. To develop a case study to highlight methods and evidence challenges and to explore alternative ways of addressing these.
4. To develop a conceptual framework to establish the evidence and analyses required to inform cost-effectiveness analyses and to guide NICE decision-making and potential Cancer Drugs Fund (CDF) data collection requirements.
5. To suggest any specific changes to the current NICE methods guide for technology appraisals or additional requirements relating to histology-independent drugs.
6. To make recommendations for further methodological research.

**Methods**

We undertook a series of targeted reviews to determine the type of evidence that is likely to be available at initial marketing authorisation and to consider the analyses likely to be required to support a NICE appraisal. These reviews included:

1. A review of FDA and EMA websites to identify relevant documents relating to regulatory issues and benefit-risk approaches relevant for histology-independent indications.
2. An overview of key statistical literature addressing the design and analysis of histology-independent trials.
3. A systematic review to identify published meta-analyses evaluating the use of overall response rate (ORR) and duration of response (DoR) as surrogate endpoints for progression-free (PFS) and overall survival (OS).

These reviews were used to identify specific challenges for histology-independent appraisals and to identify alternative approaches which might be used to investigate and account for different sources of uncertainty and heterogeneity.

We developed an exemplar economic model to illustrate the nature of the assessments that could be used to assess the cost-effectiveness of a new histology-independent treatment and to inform NICE decision-making. A framework to inform approval and research policies for histology-independent technologies was also proposed to help determine the appropriateness of different policy recommendations and to identify key uncertainties which might be used to inform and prioritise the value of further data collection.

Based on the findings of these findings, a series of recommendations were made concerning whether changes in the current NICE methods guide are required for the appraisal of histology-independent products, alongside a series of recommendations for further methodological research.

**Results**

*Review of regulatory issues and benefit-risk approaches relevant for histology-independent indications*

Our review found that histology-independent products are likely to be evaluated using more complex and innovative study designs intended to increase the efficiency of the drug development process, specifically basket trials with master protocols. Master protocols are used to evaluate multiple drugs and/or multiple cancer subpopulations in parallel, using a single protocol. Basket trials are used to evaluate a single investigational drug or drug combination in different populations (defined by disease stage, histology or treatment history) and are usually designed as single-arm activity-estimating trials with overall response rate (ORR) as the primary endpoint.

Our review of highlighted the importance placed by the regulators in the underyling biologic rationale and strength of existing clinical evidence to support the assumption that a biomarker-defined population is sufficient to establish clinically relevant activity, independent of tumour histology. Impotantly, neither the US or EU regulators concluded that the evidence for the existing histology-independent products was currently sufficient to support a routine approval decision. While the treatment effect observed in the overall population was considered to be clinically important, the initial approvals were also conditional on additional requirements for further evidence generation to increase the precision of the effect estimates and to extend the length of follow-up. As a result, important new evidence will likely emerge over time.

*Overview of key statistical literature addressing the design and analysis of histology-independent trials*

A critical consideration in the design of histology-independent trials is the potential for heterogeneity in prognosis across the different histologies, therefore standardised response rates, reflecting tumour shrinkage, are typically used instead of survival outcomes such as PFS or OS. Also, randomisation to a control arm is rare in basket trials due to the differences in standard of care across the different tumour types. The reliance on surrogate outcomes and lack of a concurrent, randomised, control arm remains a key limitation of these trial designs, and in particular, for the interpretation of such trials for NICE appraisal and HTA processes more generally.

Heterogeneity of effect across different baskets is a key consideration in the analysis of histology-indepedent trials. Once a decision has been made on whether heterogeneity is present or not, the analysis typically proceeds as either separate, independent studies for each basket, or as a single, aggregate, study combining all the baskets. Thus, either complete homogeneity or completely unrelated effects is often assumed. A less restrictive assumption is that efficacy is similar (rather than equal or completely different) across baskets. Bayesian hierarchical models (BHM) are particularly suited for this situation as they estimate the heterogeneity and allow borrowing of information on the effects of the treatment across baskets. However, the BHM is only advantageous if it is considered reasonable to allow such borrowing. Alternatives to complete pooling or borrowing across all baskets have been proposed, which extend the BHM to allow borrowing of information across similar baskets while avoiding too optimistic borrowing for extreme baskets.

Although it is challenging to determine the correct level of borrowing of information across baskets, BHM approaches provide an explicit basis to allow the treatment effect in any basket to be informed by the effects in all other baskets therefore maximising the information available.

*A systematic review to identify published meta-analyses evaluating the use of response rates and duration of response as surrogate endpoints for progression-free and overall survival*

In the context of histology-independent treatments, data on OS, and potentially also other time-to-event outcomes such as PFS, are likely to be immature. Consequently, there may be a need to rely on surrogate outcomes, such as response rate, using data from external sources, in order to estimate other more clinically meaningful final outcomes for NICE appraisal. We undertook a systematic review to assess the strength of the association between response outcomes and PFS, time to progression (TTP) or OS across different types of cancer (primarily advanced or metastatic), based on meta-analyses or meta-regression studies assessing the statistical relationship between these outcomes. Alternative sets of criteria were used to assess the strength of association between surrogate and final endpoints.

A total of 63 studies were included in the review, across 20 different cancer types. The most commonly analysed relationships were between ORR and either PFS or OS. The association between response outcomes and PFS/TTP/OS was found to vary widely between studies and generally scored low to medium when assessed using existing criteria. No clear pattern for strength of association was identified by cancer type.

Our findings that response endpoints may not be reliable surrogates for PFS or OS. However, despite the potentially weak validity of response as a surrogate for PFS and OS, we concluded that it might still be preferable for NICE appraisals to adopt a surrogate-based modelling approach informed by predictions from meta-analyses which capture all relevant uncertainty, than to ignore potential surrogate relationships and extrapolate heavily censored PFS and OS data.

*Exemplar case study*

We also identified a number of additional potential challenges of histology-independent products for NICE appraisals including the need to account for heterogeneity in a number of areas including: the cost of testing; baseline risk; quality of life and routine management costs. A range of alternative analytic approaches are likely to be required to address these different sources of heterogeneity.

The use of a single assessment of the incremental cost-effectiveness ratio (ICER), across multiple tumour sites with potentially different treatment effectiveness, comparators, costs and quality of life, may be challenging for NICE to interpret. A single ICER may conceal significant variation in the tumour specific ICERs, driven by a combination of factors, including observable variability in relative effectiveness between tumour types. Ignoring these differences could mean that a treatment which is not cost-effective for the total population (combining all subgroups) may be cost-effective in specific subgroups. Conversely, a treatment which appears cost-effective for the total population may not be cost-effective in particular subgroups. Given the amount of heterogeneity associated with a histology-independent appraisal, estimating the average cost-effectiveness for the full patient population covered by the product’s license may not provide enough information to decision-makers about whether the drug is potentially cost-effective across all subgroups.

Given the importance of exploring the impact of heterogeneity more explicitly for decision-making, explicit and transparent approaches are needed which can accommodate different sources of heterogeneity within the overall population. These assessments should allow consideration of average cost-effectiveness for the full patient population covered in marketing authorisation, as well as facilitating an assessment of whether the drug is potentially cost-effective across relevant subgroups. The use of the BHM framework provides a potential important approach which can more fully explore the potential heterogeneity in effects across tumours. The BHM approach allows assessments to be made for each tumour type, as well as a pooled assessment across all tumour types, accounting for the potential lack of uniformity of effect across tumours. An additional advantage of this type of model is the ability to predict the response probability that would be expected in a ‘new’ tumour type (i.e. a tumour that is not represented in the trial data), which will give a measure of the uncertainty in the response rates in tumour types in the target population but for which no data are available.

An exemplar case study was developed to illustrate the nature of the assessments that could be used to evaluate the cost-effectiveness of a new histology-independent treatment. The case study considered a hypothetical Trk-inhibitor (Drug X) for the treatment of solid tumours that harbour an NTRK gene fusion compared to the current standard of care (SoC). The economic model was developed using a landmark response-based structure which incorporates separate PFS and OS distributions, conditioned on response status in the overall study population. Heterogeneity in response rates across individual tumour sites was reflected using a BHM approach. By linking the BHM estimates for response rates to conditional OS and PFS estimates, the case study model explores the implications for cost-effectiveness of heterogeneity in the overall population by considering individual histology specific estimates of cost-effectiveness alongside estimates for the overall population.

In line with the NICE reference case, the model was based on a NHS and Personal Social Services perspective in terms of capturing costs and QALYs and discounts both using a 3.5% discount rate. Results are presented over a lifetime (30-year) time horizon.

The case study demonstrated the importance of understanding the frequency of histologies expected in the target population and the necessity of modelling histology specific cost and health consequences. When the expected distribution of histologies is expected to differ between the trial and the target population, failure to account for this could result in a biased estimate of the pooled ICER. The magnitude of any bias will depend on the extent of heterogeneity in relevant model inputs between tumour sites (e.g. treatment effectiveness, HRQoL, costs). Consideration will also be needed as to the potential effect in tumour histologies that are not represented in the trial data.

The case study also highlighted that even if homogeneity in all other model inputs is assumed between individual histologies (or other subgroups), the cost-effectiveness estimates will inevitably vary based on differences in the costs of identifying patients with the specific biomarker. The results demonstrate that even a low per patient testing cost can result in significant variation in the ICER estimates driven by different biomarker prevalence rates across individual histologes.

The case study was used to illustrate how heterogeneity in treatment effectiveness and testing costs could be explored using pooled ICERs and individual histology ICERs to inform decision making. However, ICERs have an important limitation; they do not give an indication of the scale of consequences for population health. Understanding the benefits and costs of treatment at a population level will help in understanding the consequences of decision making in the presence of heterogeneity and uncertainty.

We also developed a decision-framework that could be used to inform approval and research policies for histology-independent products. The framework explored the uncertainties and risks associated with different approval policies. Alternative approaches to managing risk were identified, including the role of further data collection, the use of pricing schemes and stratified decision making.

**Conclusions**

Our research found that the potential for heterogeneity in a range of model inputs, either across tumour histologies or across other characteristics, is likely to be an important issue for NICE appraisals of histology-independent technologies. Particular consideration should be given to the appropriateness of the assumptions of homogeneity of treatment effects and NICE committees should expect to see an exploration of this assumption and the potential for heterogeneity in company submissions. Where there is evidence of heterogeneity in treatment effects and estimates of cost-effectiveness, consideration should be given to optimised or ‘stratified’ recommendations. Routine presentation of the scale of the consequences of heterogeneity and decision uncertainty may provide an important additional approach to the assessments specified in the current NICE methods guide.

We identified several key areas requiring further research. Firstly, further exploration of Bayesian hierarchical methods could help inform decision makers determine whether these is sufficient evidence of homogeneity to support pooled analyses. Secondly, further research is required to determine the appropriate basis for apportioning genomic testing costs where there are multiple targets. Finally, further research is required concerning the challenges of uncontrolled Phase II studies and specifically the role and use of surrogate endpoints.

Plain English summary

In May 2017, the U.S. Food and Drug Administration (FDA) granted the first approval for a cancer treatment based on a common biomarker rather than the location in the body where the tumour originated (the tumour site), that is a site-agnostic or ‘histology-independent’ indication was granted. Research from the National Institute for Health and Care Excellence (NICE) suggests there are approximately 20 technologies currently in development for histology-independent indications with the first marketing authorisation granted in Europe in 2019.

Histology-independent treatments have the potential to have important effects in patient populations who currently have limited or no available treatment options. However, it is also important to ensure that the use of these treatments in the NHS is supported by systematic and robust assessments of clinical evidence (i.e. how well the medicine or treatment works) and economic evidence (i.e. the medicine’s value for money). These assessments are undertaken by NICE, usually for treatments targeting specific tumours sites. However, a histology-independent marketing authorisation would likely include many tumour sites and it is not possible for NICE to conduct a separate assessment for each tumour site for which the treatment could be beneficial. As a result, NICE needs to consider how these products can be appropriately assessed without creating unnecessary delays in patient access.

This research explores the extent to which NICE’s existing approaches for assessing clinical and economic value can be applied to histology-independent indications and any changes that might be required. We explore the nature and amount of evidence which is typically available at the point of initial marketing authorisation and develop recommendations to establish the evidence and analyses required to help inform NICE decisions. We use case studies to highlight possible challenges and explore ways that these might be addressed. This research will help to inform future NICE policy on how to appraise cancer drugs with histology-independent indications. It will also inform the development of guidance for those developing these treatments to help their understanding of the clinical- and cost-effectiveness assessments that will be required to inform NICE appraisals.

Table of Contents

[Abbreviations 4](#_Toc33776138)

[Abstract 7](#_Toc33776139)

[Scientific Summary 9](#_Toc33776140)

[Plain English summary 16](#_Toc33776141)

[Table of Contents 17](#_Toc33776142)

[Table of Tables 21](#_Toc33776143)

[Table of Figures 23](#_Toc33776144)

[1 Background 24](#_Toc33776145)

[1.1 Aims and objectives 24](#_Toc33776146)

[2 A review of FDA and EMA documents relating to regulatory issues and benefit-risk approaches relevant for histology-independent indications 26](#_Toc33776147)

[2.1 FDA Guidance for histology-independent products 26](#_Toc33776148)

[2.2 FDA special approval processes 29](#_Toc33776149)

[2.3 FDA review of histology-independent products 29](#_Toc33776150)

[2.3.1 FDA review of pembrolizumab 29](#_Toc33776151)

[2.3.2 FDA review of larotrectinib 31](#_Toc33776152)

[2.3.3 FDA review of entrectinib 34](#_Toc33776153)

[2.4 EMA guidance for histology-independent products 36](#_Toc33776154)

[2.5 EMA special approval processes 36](#_Toc33776155)

[2.6 EMA review of approved histology-independent indications. 37](#_Toc33776156)

[2.7 Overview of registered or completed trials for histology-independent products in development 41](#_Toc33776157)

[2.8 Summary and implications 41](#_Toc33776158)

[3 An overview of key statistical literature addressing the design and analysis of histology-independent trials 44](#_Toc33776159)

[3.1 Adaptive phase II studies 44](#_Toc33776160)

[3.2 Master protocol designs 45](#_Toc33776161)

[3.2.1 Heterogeneity of effect in basket trials 46](#_Toc33776162)

[3.3 Summary and implications 51](#_Toc33776163)

[4 A systematic review to identify published meta-analyses evaluating the use of response rates and duration of response as surrogate endpoints for progression-free and overall survival 53](#_Toc33776164)

[4.1 Introduction 53](#_Toc33776165)

[4.2 Methods 55](#_Toc33776166)

[4.2.1 Review question 55](#_Toc33776167)

[4.2.2 Inclusion and exclusion criteria 55](#_Toc33776168)

[4.2.3 Search strategy 56](#_Toc33776169)

[4.2.4 Study selection process 57](#_Toc33776170)

[4.2.5 Data extraction 57](#_Toc33776171)

[4.2.6 Data synthesis 58](#_Toc33776172)

[4.2.7 Scoring the strength of association 58](#_Toc33776173)

[4.3 Results 59](#_Toc33776174)

[4.3.1 Number of included studies 59](#_Toc33776175)

[4.3.2 Characteristics of included studies 60](#_Toc33776176)

[4.3.2.1 Surrogate relationships, cancer types and treatments 60](#_Toc33776177)

[4.3.2.2 Data types reported 61](#_Toc33776178)

[4.3.3 Results of included studies 64](#_Toc33776179)

[4.3.3.1 Absolute (individual-level) correlation and regression 64](#_Toc33776180)

[4.3.3.2 Treatment effect (trial-level) correlation and regression 69](#_Toc33776181)

[4.3.3.3 Regression equations 73](#_Toc33776182)

[4.3.3.4 Surrogate threshold effect (STE) 77](#_Toc33776183)

[4.3.4 IQWiG and BSES2 scores for strength of association 77](#_Toc33776184)

[4.4 Discussion 79](#_Toc33776185)

[4.4.1 Summary of main findings 79](#_Toc33776186)

[4.4.1.1 Types of analysis identified 79](#_Toc33776187)

[4.4.1.2 Absolute (individual-level) associations 79](#_Toc33776188)

[4.4.1.3 Treatment effect (trial-level) associations 80](#_Toc33776189)

[4.4.1.4 Regression equations and STE 80](#_Toc33776190)

[4.4.1.5 Strength of association between response and survival outcomes (IQWiG and BSES2 scoring) 80](#_Toc33776191)

[4.4.1.6 Strengths and limitations 80](#_Toc33776192)

[4.4.1.7 Summary of findings 81](#_Toc33776193)

[4.5 Implications for the economic analysis of histology-independent therapies based on ORR as a surrogate for PFS or OS 81](#_Toc33776194)

[4.6 Conclusions 83](#_Toc33776195)

[5 A targeted review of published NICE technology appraisals where initial marketing authorisation was based on response outcomes from single-arm studies 84](#_Toc33776196)

[5.1 Summary and implications 84](#_Toc33776197)

[6 Issues and challenges for exploring heterogeneity for histology-independent appraisals 86](#_Toc33776198)

[6.1 Treatment effectiveness 87](#_Toc33776199)

[6.1.1 Exploring heterogeneity in response – case study 89](#_Toc33776200)

[6.1.1.1 Methods 90](#_Toc33776201)

[6.1.1.2 Results 91](#_Toc33776202)

[6.1.2 Implications for the appraisal of histology-independent technologies 95](#_Toc33776203)

[6.2 Counterfactual 95](#_Toc33776204)

[6.2.1 Adjustment for confounding bias 97](#_Toc33776205)

[6.2.2 Alternative approaches to developing a comparator 98](#_Toc33776206)

[6.2.3 Implications for the appraisal of histology-independent technologies 99](#_Toc33776207)

[6.3 Generalisability 99](#_Toc33776208)

[6.3.1 Distribution of tumour types 99](#_Toc33776209)

[6.3.2 Unrepresented tumour types 101](#_Toc33776210)

[6.3.3 Position in treatment pathway 101](#_Toc33776211)

[6.3.4 Example 102](#_Toc33776212)

[6.3.5 Implications for the appraisal of histology-independent technologies 107](#_Toc33776213)

[6.4 Genomic testing for histology-independent drugs 107](#_Toc33776214)

[6.4.1 Overview of molecular testing in the UK 107](#_Toc33776215)

[6.4.2 Types of Genomic Test 108](#_Toc33776216)

[6.4.3 Implications of testing for appraisal of histology-independent technologies 112](#_Toc33776217)

[6.4.3.1 Costs 112](#_Toc33776218)

[6.4.3.2 Attributing testing costs 112](#_Toc33776219)

[6.4.3.3 Feasibility 113](#_Toc33776220)

[6.4.4 Identifying patients eligible for Trk-inhibitors based on the presence of an NTRK fusion a worked example 113](#_Toc33776221)

[6.4.4.1 Methods 114](#_Toc33776222)

[6.4.4.2 Results 116](#_Toc33776223)

[6.4.5 Implications for the cost-effectiveness of Trk-inhibitors 121](#_Toc33776224)

[6.5 Model structure and extrapolation 122](#_Toc33776225)

[6.6 Summary and implications 126](#_Toc33776226)

[7 A decision framework to inform approval and research policies for histology-independent technologies 127](#_Toc33776227)

[7.1 Exemplar case study 127](#_Toc33776228)

[7.2 Histology specific ICERs and overall cost-effectiveness 129](#_Toc33776229)

[7.2.1 Screening to identify eligible patients 133](#_Toc33776230)

[7.3 The value of heterogeneity and population health 136](#_Toc33776231)

[7.3.1 Histology-dependent recommendations and the value of heterogeneity 139](#_Toc33776232)

[7.3.2 No stratification: Histology-independent approval 139](#_Toc33776233)

[7.3.3 Partial stratification: Approval in a defined set of histologies 141](#_Toc33776234)

[7.3.4 Full stratification: Approval only in histologies in which cost-effectiveness is demonstrated 143](#_Toc33776235)

[7.4 Uncertainty and decision making 143](#_Toc33776236)

[7.4.1 The consequences of uncertainty 144](#_Toc33776237)

[7.4.2 Quantifying the health consequences of uncertainty 147](#_Toc33776238)

[7.4.3 Managing risk through further data collection 148](#_Toc33776239)

[7.4.4 Managing risk through pricing schemes 150](#_Toc33776240)

[7.4.5 Managing risk through stratified decision making 151](#_Toc33776241)

[7.4.6 Comparing approaches to risk management in histology-independent technologies 154](#_Toc33776242)

[7.5 Discussion 156](#_Toc33776243)

[7.5.1 Limitations of the analysis and directions for future research 157](#_Toc33776244)

[8 Recommendations for practice and further research 160](#_Toc33776245)

[8.1 Types of analyses and evidence required to inform decisions regarding histology-independent products 160](#_Toc33776246)

[8.2 Potential changes to the NICE methods guide for technology appraisals and/or additional requirements relating to histology-independent drugs. 162](#_Toc33776247)

[8.3 Priorities for future methodological research 164](#_Toc33776248)

[9 References 166](#_Toc33776249)

[10 Appendices 183](#_Toc33776250)

[Appendix 1: List of regulatory sources 183](#_Toc33776251)

[Appendix 2: Summary of trials 184](#_Toc33776252)

[Appendix 3: MEDLINE search strategy 194](#_Toc33776253)

[Appendix 4: Table of study characteristics (ordered by cancer type then author) 195](#_Toc33776254)

[Appendix 5: Absolute correlation & regression results (ordered by outcome type then cancer type then author) 201](#_Toc33776255)

[Appendix 6: Treatment effect correlation & regression results (ordered by outcome type then cancer type then author) 209](#_Toc33776256)

[Appendix 7: Studies excluded at full-text screening (n=135) 222](#_Toc33776257)

[Appendix 8: A targeted review of published NICE technology appraisals where initial marketing authorisation was based on response outcomes from single-arm studies 230](#_Toc33776258)

[Appendix 9: OpenBUGS code 252](#_Toc33776259)

[Appendix 10: Model Fit Statistics and Residual Deviance Base Case 254](#_Toc33776261)

[Appendix 11: Bayesian Hierarchical Model Sensitivity Analyses 255](#_Toc33776262)

[Appendix 12: Estimating annual eligible population 256](#_Toc33776263)

[Appendix 13 : Case Study: Economic model 266](#_Toc33776264)

Table of Tables

[Table 1: Inclusion and exclusion criteria 56](#_Toc34033515)

[Table 2: IQWiG scoring criteria71 58](#_Toc34033516)

[Table 3: BSES2 scoring criteria72 59](#_Toc34033517)

[Table 4: Study characteristics: Surrogate relationships, cancer types and treatments 62](#_Toc34033518)

[Table 5: Study characteristics: Data types 63](#_Toc34033519)

[Table 6: Summary of absolute (individual-level) correlations per study 65](#_Toc34033520)

[Table 7: Summary of treatment effect (trial-level) R2 per study 70](#_Toc34033521)

[Table 8: Regression equations for absolute associations 73](#_Toc34033522)

[Table 9: Regression equations for treatment effect (trial-level) associations 75](#_Toc34033523)

[Table 10: Surrogate threshold effect (STE) 77](#_Toc34033524)

[Table 11: FDA Results of ORR by Tumour Type 89](#_Toc34033525)

[Table 12: Overall posterior probability of response 92](#_Toc34033526)

[Table 13: Probabilities of response for all tumour types 93](#_Toc34033527)

[Table 14: Larotrectinib and Entrectinib trial tumour distributions 100](#_Toc34033528)

[Table 15: Calculation of the Annual Eligible Population for Trk-inhibitors 103](#_Toc34033529)

[Table 16: Alternative tumour distributions 105](#_Toc34033530)

[Table 17: Summary of test features 110](#_Toc34033531)

[Table 18: Sensitivity and Specificity of each Genomic Test for identifying NTRK fusions 115](#_Toc34033532)

[Table 19: Summary of NNS and testing cost under a hierarchical testing approach 117](#_Toc34033533)

[Table 20: Summary of NNS and testing cost under first-line RNA-based NGS approach 118](#_Toc34033534)

[Table 21: Summary of NNS and testing cost under exhaustive testing strategy 119](#_Toc34033535)

[Table 22: Summary of NNS and testing cost under exhaustive testing strategy when testing is done at diagnosis and treatment is provided at the appropriate line of therapy 120](#_Toc34033536)

[Table 23: Input parameters included in the economic model 129](#_Toc34033537)

[Table 24: Histology specific incremental cost-effectiveness ratios (ICERs) 131](#_Toc34033538)

[Table 25: Calculating a pooled ICER based on trial histology frequency 132](#_Toc34033539)

[Table 26: Calculating a pooled ICER based on histology frequency in the target population 133](#_Toc34033540)

[Table 27: Including testing costs into histology specific incremental cost-effectiveness ratios (ICERs) 135](#_Toc34033541)

[Table 28: Calculating per person net health effects for Drug X 137](#_Toc34033542)

[Table 29: Calculating population net health effects for Drug X 138](#_Toc34033543)

[Table 30: Incorporating unrepresented histologies into estimate of population net health effects (NHBs) 141](#_Toc34033544)

[Table 31: Decision making with partial stratification 142](#_Toc34033545)

[Table 32: Benefits of approval and further research for fully stratified decision making 153](#_Toc34033546)

[Table 33: Summary of clinical studies supporting the NICE TAs 230](#_Toc34033547)

[Table 34: Classification of approaches taken to construct a counterfactual 237](#_Toc34033548)

[Table 35: Model fit statistics for the base-case and sensitivity analyses 253](#_Toc34033549)

[Table 36: Sensitivity analyses of the Bayesian hierarchical model to alternative prior distributions 254](#_Toc34033550)

[Table 37: Illustrative example of how the diagnostic accuracy of IHC (sensitivity = 88%, specificity = 81%) is different for two tumour types with differing NTRK fusion prevalence. 255](#_Toc34033551)

[Table 38: Literature Sources of NTRK fusion prevalence, annual cancer incidence and proportion of patients diagnosed with stage 3/4 cancer in the tumour known to harbour an NTRK fusion 257](#_Toc34033552)

[Table 39: calculations of the annual eligible population for three testing strategies based on the prealence of NTRK fusion, the annual incidence of cancer in England and the proportion of patients with advanced or metastatic cancer at diagnosis 260](#_Toc34033553)

[Table 40: Testing strategy for each tumour type under the hierarchical approach 262](#_Toc34033554)

[Table 41: Testing strategy for each tumour type under the first-line RNA based NGS approach 263](#_Toc34033555)

[Table 42: Testing strategy for each tumour type under the exhaustive approach 264](#_Toc34033556)

[Table 43: Survival Estimates 265](#_Toc34033557)

[Table 44: Health state utilities 267](#_Toc34033558)

[Table 45: Drug acquisition costs and health state costs 268](#_Toc34033559)

[Table 46: Summary of NNS, Annual Eligible Population and testing costs for the tumours unrepresented in the trial 269](#_Toc34033560)

Table of Figures

[Figure 1: PRISMA flow diagram for study inclusion 60](#_Toc34033561)

[Figure 2: Correlation (r or rs) between absolute (individual-level) values of ORR and PFS\* 67](#_Toc34033562)

[Figure 3: Correlation (r or rs) between absolute (individual-level) values of ORR and OS\* 68](#_Toc34033563)

[Figure 4: Regression R2 between treatment effects (trial-level) for ORR and PFS\* 71](#_Toc34033564)

[Figure 5: Regression R2 between treatment effects (trial-level) for ORR and OS\* 72](#_Toc34033565)

[Figure 6: Summary of IQWiG scores across all 202 analyses included in the review 78](#_Toc34033566)

[Figure 7: Summary of BSES2 scores across all 202 analyses included in the review 79](#_Toc34033567)

[Figure 8: Prior and posterior distributions for the between-group heterogeneity standard deviations 91](#_Toc34033568)

[Figure 9: Posterior and predictive distributions of response probability 92](#_Toc34033569)

[Figure 10: Posterior distribution for the probabilities of response by tumours type 94](#_Toc34033570)

[Figure 11: Net health benefits per person across histologies with uncertainty 145](#_Toc34033571)

[Figure 12: Population net health effects across histologies with uncertainty 146](#_Toc34033572)

[Figure 13: Estimating the health consequences of uncertainty 147](#_Toc34033573)

[Figure 14: Population net health effects with uncertainty for partially stratified decision making 152](#_Toc34033574)

[Figure 15: Population net health effects with uncertainty for fully stratified decision making 153](#_Toc34033575)

[Figure 16: Plots showing the stylised progression free survival (a) and the overall survival (b) curves for responders and non-responders 266](#_Toc34033576)

# ****Background****

In May 2017, the U.S. Food and Drug Administration (FDA) granted accelerated approval to pembrolizumab for treating solid tumours with the microsatellite instability high (MSI-H) or mismatch repair deficient (dMMR) biomarker.1 This is the first time a cancer treatment has been approved based on a common biomarker rather than the location in the body where the tumour originated (i.e. a histology-independent approval). This represents an important paradigm shift, meaning that oncological diseases can now be classified by either tumour biomarker status or tumour histogenesis. The first histology-independent marketing authorisation was also granted by the European Medicines Agency (EMA) in 2019.2

A histology-independent marketing authorisation would likely include a large number of tumour sites. For example, larotrectinib enrolled patients across 12 tumour types.3 As it is unlikely to be feasible or desirable to conduct a separate appraisal for each tumour site contained within a histology-independent indication, NICE will need to consider how to develop a process/approach that will allow a single, biomarker-driven appraisal for histology-independent cancer drugs.

This research aims to inform future NICE policy on how to appraise cancer drugs with histology-independent indications.

## Aims and objectives

The main aim of the project is to explore the implications for the NICE technology appraisal (TA) process of appraising histology-independent products. The specific objectives are:

1. To determine the types of evidence and analyses required to support NICE appraisals of histology-independent products.
2. To identify the nature of the evidence likely to be available at the point of marketing authorisation.
3. To identify and implement a case study to highlight methods and evidence challenges and to explore alternative ways of addressing these.
4. To develop a conceptual framework to establish the evidence and analyses required to inform cost-effectiveness analyses and to guide NICE decision-making and potential Cancer Drugs Fund (CDF) data collection requirements.
5. To suggest any specific changes to the current NICE methods guide4 for technology appraisals or additional requirements relating to histology-independent drugs.
6. To make recommendations for further methodological research.

Objectives 1 and 2 are addressed in Chapters 2-5. We undertook a series of targeted reviews to determine the type of evidence that is likely to be available at the point of marketing authorisation and to consider the evidence and analyses likely to be required to support a NICE appraisal.

These reviews included:

* A review of FDA and EMA websites to identify relevant documents relating to regulatory issues and benefit-risk approaches relevant for histology-independent indications (Chapter 2).
* An overview of key statistical literature addressing the design and analysis of histology-independent trials (Chapter 3).
* A systematic review to identify published meta-analyses evaluating the use of overall response rate (ORR) and duration of response (DoR) as surrogate endpoints for progression-free (PFS) and overall survival (OS) (Chapter 4).
* A targeted review of published NICE technology appraisals where marketing authorisation was based on single-arm studies using ORR as a primary outcome (Chapter 5).

Objectives 3 and 4 are addressed in Chapters 6 and 7. Chapter 6 outlines a series of challenges for histology-independent appraisals and present alternative approaches which might be used to investigate and account for different sources of uncertainty and heterogeneity. Chapter 7 presents an exemplar economic model to illustrate the nature of the assessments that could be used to assess the cost-effectiveness of a new histology-independent treatment and to inform NICE decision-making. A framework to inform approval and research policies for histology-independent technologies is proposed to help determine the appropriateness of different policy recommendations and to identify key uncertainties which might be used to inform and prioritise the value of further data collection.

Objectives 5 and 6 are addressed in Chapter 8. Based on the findings of the research, a series of recommendations are provided concerning whether changes in the current NICE methods guide are required for the appraisal of histology-independent products. Finally, a series of recommendations are provided concerning priorities for further methodological research.

# A review of FDA and EMA documents relating to regulatory issues and benefit-risk approaches relevant for histology-independent indications

A targeted search of the FDA and EMA websites was conducted to identify relevant documents outlining regulatory approaches to the evaluation of histology-independent indications. The FDA and EMA websites were searched using the following key terms: “histology-independent”, “site-agnostic”, and “tissue-agnostic”. A narrative review of relevant documents was undertaken to summarise regulatory requirements and guidance including arrangements for post-licensing data collection. The objective was to provide insights into the current regulatory context for the benefit-risk evaluations performed by FDA and EMA and to consider their relevance for economic modelling.

It is likely that histology-independent approvals will be granted via accelerated or conditional approval processes from the FDA and EMA. Hence, the narrative review was supplemented with relevant regulatory documents related to these processes. The list of identified regulatory sources considered is reported in Appendix 1.

The targeted searches were also used to identify any completed FDA/Oncologic Drugs Advisory Committee (ODAC) and EMA/Committee for Medicinal Products for Human Use (CHMP) reviews of existing histology-independent products to provide further insights into the nature of the evidence available at the time of approval, the key issues and uncertainties raised by FDA and EMA in assessing benefits and risks, and the nature of any mandated post-licensing data collection requirements.

## FDA Guidance for histology-independent products

The website searches yielded two preliminary guidance documents issued by the FDA, addressing issues specific to histology-independent products: “*Developing Targeted Therapies in Low-frequency Molecular Subsets of a Disease Guidance For Industry*”5 and the “*Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics Guidance for Industry*”,6 both issued in 2018.

The FDA defines a therapy as ‘targeted’ if it is intended for subsets of patients within a clinically-defined disease based on either a common molecular alteration or a grouping of different underlying molecular alterations that share a common functional effect. The FDA guidance on developing targeted therapies in low-frequency molecular subsets focuses on two main issues that appear relevant to histology-independent products: (1) recommendations on how to group patients with different molecular alterations for eligibility in clinical trials; and (2) general approaches to evaluating the benefits and risks of targeted therapies, where some molecular alterations may occur at low frequencies.

The FDA guidance recognises that certain targeted therapies may be effective in multiple groups of patients who have different underlying molecular alterations because of similarities in the functional effect observed across different molecular alterations. Hence, the guidance allows grouping of patients with different molecular alterations where “*it is reasonable to expect that the grouped patients will have similar pharmacological responses based on a strong scientific rationale*”.5 Although this guidance is directed towards the grouping of molecular alterations, the same considerations might also apply to grouping of different histologies based on a common molecular alteration. The FDA guidance notes that evidence to support a grouping strategy can come from computational, experimental, or clinical sources, with the latter being considered the strongest form of evidence, but that any submitted evidence must always support a strong scientific rationale.

The FDA guidance stipulates that evidence supporting the efficacy of the drug for each molecular subset should be transparently reported, including information on the number of patients with specific molecular alterations included in the trial and the outcomes of these patients. However, the guidance also acknowledges that while targeted therapies may be effective in multiple molecular subsets, certain subsets may contain only a small number of patients (or even none) despite eligibility criteria which permit their inclusion. The FDA guidance document notes that the low numbers in this situation would preclude meaningful empirical inferences about treatment benefits or risks in patients with those particular molecular alterations. However, the FDA posits that the grouping guidance should also permit the generalisation of evidence from other, better populated, patient subgroups within the same clinical trial. Consequently, provided the company was able to support its case for molecular grouping, the FDA appears likely to approve the therapy for all patients who meet the inclusion criteria for the trial irrespective of their actual enrolment. While this issue is specifically directed towards different molecular subsets, it appears equally relevant to histology-independent products where specific histologies may include small patient numbers and some may not be represented at all.

Importantly, the FDA guidance also highlights that the indication may need to be further refined after the initial approval. If substantive data emerges indicating a lack of efficacy in certain molecular subgroups for which the drug was initially indicated, the FDA will consider narrowing the intended population as appropriate. In addition, the FDA notes that additional post-marketing studies may be required to provide additional information regarding the risks and benefits of the drug in subsets of patients with limited or no enrolment in clinical trials. Such evidence may be requested based on real-world evidence, traditional controlled trials, or data from other sources, including ongoing trials.

The FDA guidance also recognises the importance of using analytically-validated assays when enrolling patients into clinical trials. The assay should be able to identify all possible molecular alterations typical of the patient groups which is expected to respond to the developed therapy. The FDA also recommends that, if a test is necessary for the safe and effective use of the drug, an approved assay should be already commercially available at the time of drug approval. An exception to this case might be granted for conditions with high unmet need (e.g. life-threatening diseases with no suitable treatment alternatives).

An additional characteristic of histology-independent drugs is the use of novel and more efficient trial designs using master protocols. Master protocols are used to evaluate multiple drugs and/or multiple cancer subpopulations in parallel, using a single protocol. The FDA guidance document notes that a range of different terms are used to refer to the specific design of trials within a master protocol (e.g. umbrella, basket or platform – see Section 3 for further details on these designs).

The FDA guidance acknowledges the potential advantages of master protocols in terms of their flexibility and efficiency for drug development but also raises concerns regarding difficulties in attributing efficacy and assessing safety, including overinterpretation of findings. In the context of histology-independent products, the most relevant aspects of the guidance relate to the use of basket trials to evaluate a single investigational drug or drug combination in different populations (defined by disease stage, histology or treatment history) and statistical considerations for non-randomised, activity-estimating designs. The guidance document highlights that basket trials undertaken using a master protocol are usually designed as single-arm activity-estimating trials with ORR as the primary endpoint. The guidance document notes that a strong response signal seen in a sub-study may allow for subsequent expansion to generate data that could potentially support a marketing approval. The guidance document also emphasises the need for each sub-study to include specific objectives, the scientific rationale for the inclusion of each population and a detailed statistical analysis plan (SAP) which includes justification for sample size and stopping rules for futility, i.e. the inability of the study to achieve statistically significant results.

The statistical guidance also makes recommendations for studies using non-randomised protocols, where the primary endpoint is ORR, outlining that the planned sample size should be sufficient to rule out a clinically unimportant response rate based on the lower bound of the 95 percent confidence interval (CI) around the observed response rate. The guidance also recommends using designs such as Simon’s two-stage design,7 that limit exposure to an ineffective drug (see Section 3 for further details). Specific recommendations concerning the SAP include pre-specification of the timing of the final analysis, ensuring adequate data collection and follow-up on all patients for efficacy and safety, together with a description of the plan for independent review of confirmed ORR in solid tumours for each sub-study.

While the current guidance suggests that marketing approval requires subsequent expansion of a sub-study or sub-studies, the guidance on statistical considerations also notes that if preliminary results suggest a major advance over available therapy then the sponsor is encouraged to meet the FDA review division to discuss modifications to the protocol. Hence, it appears feasible for the results from master protocols using basket trials to be used to support marketing approval in specific circumstances and where the clinical protocol and SAP ensure that the data are of adequate quality.

## FDA special approval processes

The initial histology-independent cancer drugs approved by the FDA represent novel products tackling severely limiting conditions with no alternative curative options. As such, they have not been considered within the standard FDA review process, but rather have been considered under processes which make provision for special approval to facilitate and expedite development and appraisal of new drugs treating serious or life-threatening conditions.8 This has been the case for the three histology-independent approvals by the FDA for larotrectinib, pembrolizumab, and entrectinib.1, 9, 10

The Accelerated Approval pathway is intended for those drugs providing evidence of an effect on a surrogate endpoint reasonably likely to predict benefit in terms of a meaningful advantage over existing therapies. Surrogate endpoints are defined as substitutes for clinical outcomes which directly measure the effectiveness of a drug on length and quality of life, feelings, or functioning. In cases where measuring direct clinical outcomes, such as overall survival, would be impractical or unethical, surrogate endpoints can be accepted. Importantly, the surrogate outcome is not a direct measurement of clinical benefit but must predict, and at a minimum correlate, with the clinical benefit of interest. The strength of the evidence supporting the surrogate relationship is thus essential to justify the use of a specific surrogate outcome and to establish whether this can support a traditional approval route or accelerated approval.

To date, ORR has been the most commonly used surrogate endpoint supporting accelerated approvals by the FDA.11 One important reason for this is that ORR can be directly attributable to drug effect, and hence single-arm studies conducted in patients with refractory tumours where no available therapy exists are considered to provide an appropriate assessment of ORR. However, the FDA also acknowledges that the clinical benefits of interest may not always be predicted by, or correlate with, ORR. Hence the use of measures such as ORR to support an accelerated approval or traditional approval endpoint ultimately depends on the disease context and the magnitude of the effect, among other factors.

## FDA review of histology-independent products

### FDA review of pembrolizumab

The FDA approved pembrolizumab (Keytruda, Merck) on May 23rd, 2017, for the treatment of adult and paediatric patients with unresectable or metastatic, MSI-H, or dMMR solid tumours. The approval covers patients that have progressed following prior treatment and who have no satisfactory alternative treatment options, and for the treatment of unresectable or metastatic MSI-H or dMMR colorectal cancer (CRC) that has progressed following treatment with a fluoropyrimidine, oxaliplatin, and irinotecan.1

The efficacy of pembrolizumab in patients with MSI-H or dMMR solid tumours was derived from five uncontrolled, open-label, multi-cohort, multi-centre, single-arm studies. Patients received either 200 mg of pembrolizumab every 3 weeks or 10 mg/kg every 2 weeks. Treatment continued until unacceptable toxicity or disease progression (up to a maximum of 24 months treatment).

A total of 149 patients with MSI-H or dMMR cancers were included across the five clinical trials. The median age of patients was 55 years, 98% of patients had metastatic disease and 2% had locally advanced, unresectable disease. Ninety (60%) of the 149 patients had CRC with the remainder diagnosed with other tumour types. The median number of prior therapies for metastatic or unresectable disease was two.

The identification of MSI-H or dMMR tumour status was prospectively established for the majority of patients (135/149) using local laboratory-developed, polymerase chain reaction (PCR) tests for MSI-H status or immunohistochemistry (IHC) tests for dMMR. Tumours from the remaining 14 patients were retrospectively identified as MSI-H using a central laboratory-developed PCR test.

The primary endpoint used for the FDA review was ORR as assessed by blinded independent central radiologists’ (BICR) review according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1. The ORR (95% CI) was 39.6% (31.7% to 47.9%). DoR was considered as a key secondary endpoint. Although the median DoR was not reached, 78% of responding patients had a DoR of 6 months or longer. Overall, the safety profile of pembrolizumab was considered acceptable relative to durable responses observed in patients with advanced MSI-H/dMMR cancers.

A total of 16 different tumour types were included in the combined dataset. Consistent responses were reported between subjects with gastrointestinal (GI) (CRC, small bowel, gastro-esophageal junction, pancreas) and non-GI MSI-H cancer, with ORRs of 36.8% (MSI-H GI; N=125) and 41.7% (MSI-H non-GI); N=24). However, the FDA noted that some of the tumours (e.g. breast, prostate, sarcoma, renal cell) were only represented by 1 or 2 patients and that there was uncertainty whether the results apply to all disease types with MSI-H/dMMR status.

The key question considered by the FDA within their review was whether the presence of MSI-H/dMMR represents a unique biomarker that predicts a consistent response to pembrolizumab and similar clinical benefit across different primary tumours. In addressing this question, the FDA highlighted that features associated with MSI-H/dMMR that are common across primary cancers include increased lymphocytic infiltration and an increased mutational tumour burden with non-synonymous mutations. Both factors were noted to have been previously identified as correlating with an increased likelihood of response to checkpoint inhibitors, such as pembrolizumab, in tumours that have not been assessed for MSI-H or dMMR. Based on these common histologic features, the FDA concluded that there was a strong biologic rationale that MSI-H/dMMR cancer represents a specific subpopulation of patients with cancer who are likely to derive clinical benefit from pembrolizumab.

Pembrolizumab was approved by the FDA for this indication under Accelerated Approval based on ORR and DoR. Despite a common biology among MSI-H/dMMR tumours, the FDA review also highlighted other differences among patients with different types of cancer that could influence response to therapy with pembrolizumab (e.g., the degree of immunosuppression related to previous cytotoxic chemotherapy). Given the uncertainties that remain concerning the generalisability of the results to all disease types with MSI-H/dMMR status, a condition of the approval requires the sponsor to submit results of further studies to better characterise response rate and its duration. These studies are required to include 124 patients with CRC and at least 300 patients with non-CRC, including a sufficient number of patients with prostate cancer, thyroid cancer, small cell lung cancer and ovarian cancer; and 25 children.

The FDA review noted that further randomised trials will be challenging to conduct in the histology-independent setting given concerns over equipoise. The FDA also questioned whether it would be scientifically appropriate to ‘lump’ all tumour types together into a single randomised trial given the different natural histories. The FDA also noted that while response may not be entirely predictive of effects on clinical benefit, checkpoint inhibitor therapy, including pembrolizumab, has demonstrated beneficial effects on OS with similar response rates in other tumour types.

In the absence of a companion diagnostic test for the identification of MSI-H or dMMR tumour status, the FDA review noted the uncertainties regarding use of laboratory-developed tests. These uncertainties concerned the rate of false-positives in IHC tests for dMMR and false-negatives in PCR tests for MSI-H and whether performance characteristics may differ by the site of the primary tumour. Given these uncertainties, additional post-marketing studies were requested to assess and establish the performance characteristics of MSI-H and dMMR tests.

### FDA review of larotrectinib

On November 26th, 2018, the FDA granted accelerated approval to larotrectinib (Vitrakvi, Loxo Oncology Inc. and Bayer) for adult and paediatric patients with solid tumours that have a NTRK gene fusion without a known acquired resistance mutation, that are either metastatic or where surgical resection is likely to result in severe morbidity, and who have no satisfactory alternative treatments or whose cancer has progressed following treatment.9

As agreed with the FDA, the submission was supported by pooled safety and efficacy data from the first 55 patients enrolled in three multicentre, open label single-arm studies enrolling subjects with solid tumours harbouring a NTRK fusion, who met the following criteria:

1. Documented NTRK fusion as determined by local testing
2. Non-central nervous system (CNS) primary tumour with 1 or more measurable lesions at baseline as assessed by RECIST 1.1
3. Received one or more doses of larotrectinib.

ORR determined by an Independent Review Committee (IRC) was used as the primary endpoint for efficacy. DoR was a secondary endpoint, defined as the number of months from the start date of partial response (PR) or complete response (CR) to the date of disease progression or death, whichever occurred earlier.

Assuming the observed ORR was at least 50%, a sample size of 55 patients was selected to provide 80% power to achieve a lower boundary of the 2-sided 95% exact binomial CI about the estimated ORR exceeding 30%. Ruling out a lower limit of 30% for ORR was considered clinically meaningful. All patients were required to either have progressed following previous systemic therapy for their disease or that they would have required surgery with significant morbidity for locally advanced disease. The data cut-off for the primary analysis was the 17th July 2017, approximately 6 months after enrolment of the 55th patient.

The pooled sample included 12 different tumour sites, of which the most frequent were salivary gland tumours (22% of patients), soft tissue sarcoma (20%), and infantile fibrosarcoma (IFS) (13%). More common tumours like lung or colon cancer were less represented (4 patients, or 7% each), as they tend to rarely express a NTRK fusion. The sample was also heterogeneous in terms of prior cancer therapy, with patients having undergone different types of therapies (surgery, radiotherapy, systemic therapy), and different numbers of previous lines of therapy (45% having undergone 1-2 lines, and 35% three or more).

At the time of data cut-off, the estimated ORR was 75% (95% CI: 61% to 85%), including 22% of patients with a CR and 53% of patients with a PR. Although median DoR had not been reached, 30 out of 41 (73%) responders had at least 6 months’ DoR and 16 out of 41 (39%) responders had at least 12 months’ DoR.

The clinical and statistical review included exploratory subgroup analysis performed by study, demographics, and tumour type. Based on these analyses, the effectiveness of larotrectinib was reported to be reasonably similar irrespective of age, gender, and race; however, no definitive conclusions were made given the limited sample size. A numerical difference in ORR was reported among patients with different tumour types, NTRK gene fusions or status of radiotherapy. Across different tumour types, 3 tumour types had at least 7 patients: salivary gland (N=12), soft tissue (N=11) and IFS (N=7). The ORR in these tumour types was reported to be higher than 75%. Conversely it was reported that the ORR in colon cancer appeared to be lower (1 out of 4 patients). No response was reported in the 2 patients with primary CNS lymphoma.

The FDA review concluded that while the results showed that treatment with larotrectinib results in durable overall responses in patients with a variety of tumour types, there was insufficient clinical experience to conclude that the response rates achieved with larotrectinib were consistent across all NTRK-fusion cancers.

A key issue addressed in the review was the potential risk that larotrectinib could be ineffective in some tumour types even in the presence of an NTRK fusion. The FDA concluded that the risk of ineffectiveness was low due to the strong rationale presented by the company, supported by clinical and non-clinical data. The strength of the evidence was assessed against the three following criteria: the ability of the biomarker to identify a population with common features, the similarity of response across tumour types, and the ability to reliably identify the biomarker at the screening phase.

The FDA considered the totality of evidence presented by the sponsor to be sufficiently strong to consider pooling of results across trials and patients, supporting a histology-independent indication. The FDA also concluded that while there was a risk that larotrectinib may be ineffective in some tumours, the level of risk was deemed low and considered acceptable given the product is only approved for the treatment of patients who have no satisfactory alternative treatment options or whose cancer has progressed following treatment. As a result, the FDA did not consider that patients would be forgoing effective therapies when treated with larotrectinib.

The primary risks of larotrectinib were identified as hepatotoxicity and neurotoxicity. However, these adverse reactions were considered largely manageable and reversible with dose modification or discontinuation. Overall, the toxicity profile of larotrectinib was considered acceptable when considered against the durable effects across different cancer types in patients with limited or no effective treatment options.

ORR was considered a surrogate endpoint reasonably likely to predict benefit, in accordance to the requirements of the Accelerated Approval process. The clinical effect was deemed sufficiently large, and the effect was durable, providing a meaningful advantage over available therapy for patients with NTRK-fusion solid tumours. The population was also considered to have a high unmet medical need given the serious, life-threatening, and rare nature of their cancers. However, the FDA specified that the ORR evidence was not sufficiently strong to support a regular approval given the large number of histologic subtypes and the small sample size. This led to a degree of uncertainty regarding the magnitude of the treatment effect of larotrectinib in any single histologic subtype.

A key post-marketing requirement is that the company conduct further studies to provide additional data to verify and confirm the clinical benefit of larotrectinib, through more precise estimation of ORR and DoR in several specific tumour types (CRC, non-small cell lung cancer (NSCLC), CNS tumours and melanoma) which were not well represented in the existing efficacy population. A minimum of 40 patients are also required to be studied with cancers other than CRC, NSCLC, CNS tumours, melanoma, soft tissue sarcoma, thyroid cancer, IFS, and salivary cancers (e.g., breast cancer, gastrointestinal stromal tumours, cholangiocarcinoma, biliary tract cancers). ORR and DoR are required as endpoints and all responding patients are required to be followed for at least 12 months from the onset of response. In addition, a final report is requested from the first 55 patients with NTRK fusion solid tumours enrolled to further characterise the duration of response including follow-up of at least 2 years from the onset of response for responding patients.

Importantly, the FDA concluded that it would not be feasible or appropriate to conduct a randomised trial to demonstrate that larotrectinib improves OS in patients with NTRK fusion positive patients. The reasons included the extreme rarity of NTRK-fusion cancers, the lack of equipoise in settings without available therapies, and expectations for patient cross-over. Consistent with their review of pembrolizumab, the FDA again queried whether it would even be scientifically appropriate to ‘lump’ these tumour types together into a single randomised trial given differences in natural history between different tumour sites.

Identification of positive NTRK gene fusion status was determined in the clinical efficacy analysis set using next generation sequencing (NGS) for 91% of patients, and fluorescence *in situ* hybridization (FISH), for the remaining 9%. The company did not submit an application for an *in vitro* companion diagnostic device. Despite this, the clinical review team was supportive of approval, citing the availability of a reliable non-companion device and the efficacy of larotrectinib. However, the development and validation of a companion diagnostic test by the sponsor was agreed as part of a series of post-marketing commitments.

### FDA review of entrectinib

On August 15th 2019, the FDA granted accelerated approval to entrectinib (Rozlytrek, Genentech Inc.) for adults and paediatric patients 12 years of age and older with solid tumours that have an NTRK gene fusion without a known acquired resistance mutation, are metastatic or where surgical resection is likely to result in severe morbidity, and have progressed following treatment or have no satisfactory standard therapy.10

This indication was approved by the FDA under Accelerated Approval based on ORR and DoR. The submission was supported by pooled efficacy and safety results from the first 54 adult patients with unresectable or metastatic solid tumours harbouring an NTRK fusion enrolled across three‐single arm studies. All patients were required to have cancer that progressed following effective systemic therapy for their disease, if available, or would have required surgery with significant morbidity for locally advanced disease.

The median age of patients was 55 years. The most common tumours (≥ 5%) were lung cancer (56%), sarcoma (8%), and colon cancer (5%). Ninety‐six percent of patients had metastatic disease and 4% had locally-advanced, unresectable disease. All patients had received prior treatment for their cancer, including surgery, radiotherapy, or systemic anti‐neoplastic therapy.

ORR and DoR as assessed by BICR using RECIST v1.1 were the primary endpoints. PFS as assessed by BICR and OS were included as secondary endpoints. The effectiveness of entrectinib in paediatric patients 12 years of age and older was established based on extrapolation of data in adult patients with solid tumours harbouring an NTRK gene fusion and pharmacokinetic data in adolescents enrolled in the STARTRK‐NG study.

In the first 54 patients, the ORR was 57% (95% CI: 43% to 71%). This was considered to be clinically meaningful because the results excluded a lower bound of the 95% CI for ORR of 30%. At the time of data cut-off (31st May 2018), the median DoR was not reached. Among the 31 responding patients, 55% had a DoR of at least 6 months and 39% had a DoR of at least 12 months.

Exploratory ORR results for subgroups defined by tumour type and by NTRK gene fusion partner were presented. Although there was no formal discussion of these results, a general disclaimer was provided noting that the subgroup results should be treated with caution due to the small sample sizes and the single-arm design.

Only limited details were reported for secondary endpoints. The estimated median PFS was reported to be 11.2 months (95% CI: 8.0 to 14.9). Less than 30% of deaths were observed by the clinical cut-off date (31st May 2018) which was considered too immature to be considered in the clinical review.

The most serious adverse events reported with entrectinib were: congestive heart failure (CHF); CNS adverse reactions; skeletal fractures; hyperuricemia; hepatotoxicity; QT prolongation; and vision disorders. Although serious in nature, these events were also reported to be manageable and reversible with dose modification or discontinuation of entrectinib.

The FDA drew similar conclusions for entrectinib to their earlier review of larotrectinib (Section 2.3.2). While acknowledging that there was uncertainty regarding the magnitude and durability of the treatment effect of entrectinib in any specific histologic subtype of solid tumours, they concluded that the risk of treatment was low, using a similar rationale to that previously described for larotrectinib.

Similar post-marketing requirements for entrectinib as for larotrectinib were also reported. This requires the company to conduct additional single-arm studies to obtain data to verify and further characterise the clinical benefit of entrectinib, in an adequate number of patients with common histologic tumour types, including colon cancer and melanoma. Additional post-marketing requirements also include the conduct of additional studies to further characterise the risks of CHF and skeletal fractures with entrectinib.

## EMA guidance for histology-independent products

The EMA has not published any guidance specific to the evaluation of histology-independent products. However, the proceedings of two workshops were identified in the searches, one specifically addressing histology-independent indications,12 and a second discussing the use of single-arm studies in oncology.13

A revision to the current “Guideline on the evaluation of anticancer medicinal products in man”14 is currently under consultation.15 The concept paper underlying the revision explicitly states the need to address the use of biomarkers in oncology, which were not covered by the previous guideline. This development recognises the increasingly important role biomarkers have in both defining disease and developing treatment strategies. Biomarker-based treatments also have the possibility to span across tumour sites and are likely to be assessed using innovative study designs such as basket and umbrella trials. These study designs were not considered in the current guideline,14 therefore an update was recommended by the Oncology Working Party. The update will focus on better identifying the role of biomarkers in the development pathway, developing evidence standards in the context of rare cancers and outlining the main aspects and principles of innovative study design including the use of basket trials.

## EMA special approval processes

Like the FDA, the EMA provides alternative marketing authorisation pathways to cover situations where the nature or quality of the evidence would not be sufficient to support traditional approval. Conditional Approval from EMA is a form of conditional marketing authorisation for those medicines, targeting unmet medical needs for serious conditions, with a positive benefit-risk balance but which do not have comprehensive data available. In order to grant conditional approval, agreement is needed on additional post-marketing studies to confirm the initial assessment of the benefit-risk balance. This marketing authorisation is valid for 1 year and can be renewed annually following a rolling review, provided the benefit-risk assessment is still considered to be positive.

## EMA review of approved histology-independent indications.

To date, only one histology-independent product has received marketing authorisation in the EU. Larotrectinib (Vitrakvi, Loxo Oncology Inc. and Bayer) received a conditional marketing authorisation on the 19th September 2019.2 The authorisation recommends larotrectinib as monotherapy for the treatment of adult and paediatric patients with solid tumours that display an NTRK gene fusion, who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and who have no satisfactory treatment options.

The EMA review was supported using several different analysis sets. The Primary Analysis Set (PAS) was based on the same 55 patients considered in the earlier FDA review of larotrectinib. The PAS analysis was based on a pooled analysis of patients consecutively enrolled from three single-arm studies.

The EMA review identified several concerns regarding the PAS. Firstly, the restriction to the first 55 patients was considered to have been arbitrarily chosen. Secondly, the exclusion of CNS-tumours was considered to introduce a bias in the efficacy estimates. Finally, restricting the analysis to patients who received one or more doses was not considered to accord with the intention-to-treat (ITT) principle.

Following requests from the CHMP, further analysis sets (ePAS and ePAS2) were submitted which included additional data from an extended follow-up and a larger pooled analysis population. The ePAS (n=73) included all patients who met all PAS eligibility criteria as of 19th February 2018 and had central review of tumour response by the IRC. This included an additional 18 patients compared to the PAS (n=55). The ePAS2 (n=93) included all patients who met all PAS eligibility criteria and had either discontinued the study or had at least 6 months follow-up by 30th July 2018. This included an additional 38 patients compared to the PAS (n=55). The ePAS2 was the main efficacy analysis set considered in the EMA review.

A further cohort including paediatric and adult patients with primary CNS tumours (n=9) was reported separately. This cohort represented a pre-specified exclusion criterion from the original PAS analysis. This cohort was considered to have a potentially lower likelihood of response compared to the other cohorts given results from earlier animal studies, indicating low penetration of larotrectinib into CNS tissues. However, the review also acknowledged that CNS penetration in cancer patients taking larotrectinib may be more substantial than suggested by prior evidence.

The primary endpoint considered was ORR by IRC assessment, defined as the proportion of patients with best overall response of CR or PR. Secondary endpoints included: time to response (TTR), DoR, PFS (including PFS rate at 6 and 12 months) and OS (including survival rate at 12 months).

In the ePAS2 analysis, the ORR by IRC was 72% (n=67/93, 95% CI: 62% to 81%). The ORR results were considered by the EMA review to be outstanding. The median TTR was 1.8 months by IRC (25th, 75th percentiles: 1.71, 1.94 months). The median DoR was not estimable. However, 72% of responding patients were reported to have had a DoR of 6 months or more, and 42% had a DoR of 12 months or more. The review also noted that the percentage of patients with durable responses appeared higher in previously submitted data with shorter follow-up. Concerns were expressed that the difference in results between alternative follow-up times indicated that limited early data might overestimate the true treatment effect.

The EMA review noted that there was substantial heterogeneity across the three separate studies and that the primary endpoint was based on a crude proportion of responses. The review also highlighted that sensitivity analyses provided by the sponsor utilising tumour type as a random factor provided slightly lower estimates than the crude proportions. Further reanalysis by the EMA involved investigating alternative selections of cohorts from the three studies. These analyses indicated that the crude ORR appeared in the upper end (the 90th percentile) of the distribution of possible estimates, suggesting a possible selection bias. However, the review also noted that a large majority of all possible ORR estimates were above 50% indicating a true effect of a relevant magnitude.

The median PFS was 27.4 months (95% CI: 13.8, not estimable [NE]) by IRC. The PFS rate at 6 months was 77%, and the PFS rate at 12 months was 64% (95% CI: 51% to 76%). The median OS was not reached in the ePAS2 due to the low event rate of 15% (14/93 dead), at a median follow-up time of 16.7 months. The OS rate at 12 months was 88% (95% CI: 81% to 95%). All 9 patients in the CNS group were noted to still be alive at the time of the final data cut-off.

The EMA review highlighted the immaturity of the OS and PFS data. In addition, while the PFS and OS data were considered important for contextualising the ORR and DoR results, the pooling of many different types of primary malignancies with inherently different prognosis, led to a conclusion that the data should be interpreted with caution.

The subgroup analysis reported in the EMA review included an analysis of ORR by tumour type. The ORR was reported to be highly variable across the studied tumour types, ranging from 0% ORR in individual patients with breast cancer, cholangiocarcinoma and pancreatic cancer to 100% in 4 patients with gastrointestinal stromal tumour (GIST). The review indicated that tumour types where NTRK gene fusions are characteristic (or even considered pathognomonic) of the disease, such as IFS (n=13), salivary gland/mammary analogue secretory carcinoma (MASC) (n=10), and congenital mesoblastic nephroma (n=1), tended to have higher ORRs (92%, 80%, and 100%, respectively). However, the review also concluded that the tumour-specific estimates were not robust due to the small sample sizes of individual subgroups. Among the 9 patients with primary CNS tumours, 1 had an objective response (PR); the remaining 8 had stable disease as best response. Six patients were reported to be without progression at last follow-up. The CHMP considered that there was no scientific rationale to exclude previously treated CNS patients with no satisfactory treatment options available, and that the indication should also cover these patients.

A key question considered in the EMA review was whether the available data supported the assumption that NTRK gene mutations are oncogenic driver mutations and that the mechanism of action is independent of tumour histology. This assumption was considered necessary to conclude that larotrectinib would result in clinically relevant activity in tumours expressing NTRK fusion proteins regardless of the tissue of tumour origin. Additional advice was sought to address this question from the Scientific Advisory Group (SAG) in Oncology, and the EMA Biostatistics Working Party (BSWP).

The consensus view of the SAG was that the available data did not support the hypothesis that NTRK gene fusions are universally oncogenic drivers, independent of tumour type/histology and other disease characteristics. The SAG also concluded that the mechanism of action may differ according to histology and other characteristics and the existing data were insufficient to establish activity regardless of tumour type and other characteristics. However, the SAG also recognised that in some paediatric malignancies, preclinical and clinical data supported NTRK as an oncogenic driver. In addition, fusion genes affecting NTRK 1/2/3 were reported to be highly recurrent in certain rare malignancies. ETV6-NTRK3 was noted to be present in >95% of secretory carcinomas of the breast, MASC of the salivary glands, congenital fibrosarcoma and cellular mesoblastic nephromas. This led one expert to suggest the possibility to have a histology-independent approval for cancers with proven NTRK fusions as oncogenic ‘drivers’, provided that NGS could exclude other alterations being significant drivers for tumour progression. However, it was also noted that data do not exist currently to establish the efficacy of such a strategy.

The SAG acknowledged the strong rationale and the available clinical data for several specific tumour types (IFS, salivary gland/MASC and congenital mesoblastic nephroma) for which NTRK fusions have been established as oncogenic drivers independent of other characteristics. The SAG also noted that larotrectinib has shown important activity in GIST with NTRK after resistance/relapse with imatinib (ORR=5/5) reflecting a likely similar role for NTRK fusions. For these selected conditions, given the strong rationale and the available clinical data, the SAG concluded that efficacy has been established in the absence of available treatments of proven efficacy in terms of convincing clinical efficacy endpoints. However, for other conditions, the review concluded that the role of NTRK fusions had not been properly studied and could not be appropriately established with existing data given the lack of comprehensive sequencing of tumour tissue prior to treatment initiation. Concerns were also expressed from the SAG regarding the small sample sizes in different tumour types, the significant heterogeneity observed in terms of response rates and very low ORR observed in different tumour types (ORR=0% to 33%). The low ORRs were also noted to be reported in common tumour types where occurrence of NTRK gene fusion is rare (e.g. lung, colon, and breast).

The SAG concluded that neither the available evidence nor reasonable extrapolations supported the proposed indication to include all solid tumours independently of tumour type. The SAG considered that clinical decisions to use larotrectinib were justified for the rare conditions where existing evidence more clearly supported the role of NTRK fusions as oncogenic drivers. For other conditions, the acceptable safety profile supported use in situations where established alternatives are lacking or, where available alternatives are associated with high morbidity and mortality.

Further to the SAG comments, the CHMP highlighted that a certain degree of heterogeneity in response is unavoidable in the same way as there will be important effect modifiers within any indication. Thus, the critical issue considered by the CHMP was whether the studies were likely to be representative of the treated population once the product is authorised and whether the uncertainties are acceptable given available data and the intended use as a last-line treatment in patients without satisfactory treatment options.

The clinical review concluded that while the efficacy results were outstanding for a late stage disease setting, significant uncertainties remained concerning the robustness and generalisability of these estimates. The review also acknowledged that the results may change in a negative direction as further evidence is generated. However, the magnitude of the current effect estimates was considered of sufficient size to support a likely large treatment benefit observed in practice. The review also noted that the interactions between treatment and tumour type required further exploration.

The available data were considered non-comprehensive and a conditional approval was considered appropriate by the EMA. The conditional approval was granted based on a positive benefit-risk balance and the requirement that the company provide additional comprehensive data. As part of this requirement, the company is required to submit a prospective cohort of 75 patients as part of study LOXO-TRK-15002 (NAVIGATE), for which at least 1 year of follow-up is available and to perform an overall pooled analysis including the ePAS2/CNS cohort to give increased precision for the estimates of ORR and DoR. In addition, the company plans to enrol 200 additional patients in NAVIGATE (LOXO-TRK-15002) and as part of the SCOUT study (LOXO-TRK-15003) within a 36-month period post approval. Eighty patients are planned to be recruited in 4 common tumour types (lung cancer, colorectal cancer, melanoma and non-secretory breast cancer) and 120 in other tumour types. At least 9 (and up to 20) patients will be recruited in each of the 4 common tumour types, permitting a more precise estimate of efficacy in common cancers where NTRK fusions are rare.

## Overview of registered or completed trials for histology-independent products in development

Research from NICE suggests that there are approximately 20 technologies currently in development for histology-independent indications. We undertook searches of the clinicaltrials.gov website using the list of histology-independent products provided by NICE. Information was extracted for those trials which are more likely to be vehicles for regulatory approval, that is, combined Phase Ib/II, Phase II, and Phase III trials. The aim of this review was to clarify whether the level of evidence available during the FDA/EMA appraisals of the initial histology-independent products is likely to be representative of that of future products in other indications.

Appendix 2 provides a summary of registered or completed Phase Ib/II, Phase II, and Phase III trials identified using searches of the clinical trials.gov website. Of the 20 products considered, 3 specific products (pembrolizumab, larotrectinib and entrectinib) were excluded as more detailed evaluations of the regulatory submissions have been summarised in previous sections of this chapter. Of the remaining 17 products, only 13 had registered trials which were considered potentially suitable for regulatory purposes. A total of 36 relevant trials were identified for these 13 products. 13 of the trials were for one specific drug (olaparib). The products identified included drugs already approved for specific indications (e.g. olaparib) seeking to expand their existing marketing authorisation and novel products which may be seeking an initial approval in a histology-independent context (e.g. LOXO-295).

Over 90% (n=33) of the 36 registered trials were single-arm studies. ORR was the most common primary endpoint (n=27), although PFS was reported as a primary endpoint in four studies. DoR (n=18), PFS (n=28) and OS (n=24) were commonly included as secondary endpoints.

Of the 36 trials, only 3 were formally referred to as basket trials. 19 of the remaining 33 studies (58%) included separate treatment or population cohorts, suggesting that the analyses may explore differences between the separate cohorts. The remaining studies reported no details on specific cohorts or subgroups that might be considered.

## Summary and implications

The study design and evidence considered by the FDA and EMA for the initial approvals of histology-independent products appear consistent with the type of evidence that may be expected for future approvals (e.g. single-arm studies with ORR as the primary endpoint). Although the FDA has now issued specific guidance concerning the conduct and reporting of basket trials to evaluate a single investigational drug or drug combination in different populations, the design of many ongoing or recently completed studies clearly pre-date this guidance. Only a small number of the trials were formally referred to as a basket trial and there was a lack of clarity in the design of many studies concerning whether separate cohorts would be formally considered or not. As a result, it appears likely that the current case-by-case approach employed by the regulators in determining the appropriateness and quality of the underpinning evidence to support a histology-independent approval will continue for the foreseeable future.

The central question considered by both the FDA and EMA concerns the biologic rationale and strength of existing clinical evidence to support the assumption that a biomarker-defined population (e.g. MSI-H/dMMR or NTRK) is sufficient to establish clinically relevant activity independent of tumour histology. Neither the FDA nor the EMA considered that the current evidence base for any of the three products was sufficiently robust to establish this. Indeed, both agencies raised important uncertainties regarding the generalisability of the results across all individual histology sites. However, the magnitude of the effect in the overall population was considered clinically important and the risk associated with approving the treatment in specific tumours was considered to be low due to the strong biologic rationale and the intended approval as a last-line treatment in patients without satisfactory treatment options.

It is evident from the FDA and EMA reviews for larotrectinib that the evidence base is rapidly developing over time, such that the later EMA review included an additional 38 patients (n=93) compared to the FDA review (n=55). It is also notable that the advice of the SAG to the EMA, based on this larger dataset, appeared to differentiate the strength of the biological rationale and the available clinical evidence for several specific tumour types. For a few specific tumour types (IFS, salivary gland/MASC and congenital mesoblastic nephroma), the SAG concluded that NTRK fusions had been established as oncogenic drivers, independent of other characteristics. The SAG also concluded that evidence for GIST was sufficiently strong to support a similar role of NTRK fusions as an oncogenic driver. For these specific tumour types, the SAG concluded that efficacy has been established in the absence of available treatments of proven efficacy in terms of convincing clinical efficacy endpoints and that clinical decisions to use larotrectinib were justified. For other conditions, the acceptable safety profile supported use in situations where established alternatives are lacking or, where available alternatives are associated with high morbidity and mortality.

Both the FDA and EMA reviews ultimately concluded that the evidence for these existing products was not sufficient to support a routine approval for a histology-independent label. The further evidential requirements focus on three specific aspects: (i) increasing the precision for the estimates of ORR and DoR and extending the length of follow-up in the overall population; (ii) generation of new evidence to increase the precision of efficacy in more common cancers where NTRK fusions are rare (e.g. lung, colorectal, melanoma and non-secretory) and where current evidence is sparse; and (iii) the development and validation of a companion diagnostic test. As a result, important new evidence will emerge over time to address some of the key uncertainties identified by the EMA and FDA.

The reviews also highlighted two important challenges that need further consideration. Firstly, the design and conduct of trials to support histology-independent products is likely to differ from more conventional products. The use of novel and efficient basket trial designs using master protocols will present additional challenges to NICE in terms of HTA assessment. Hence, the rationale and statistical basis for the design of these studies warrants further consideration. Secondly, the initial evidence supporting the basket trials is likely to be focused on surrogate endpoints such as ORR and DoR. Our reviews show that while data on more policy-relevant outcomes such as PFS and OS are being collected, there are likely to be a number of potential challenges regarding their interpretation in the absence of a comparator arm, possible bias due to confounding (e.g. receipt of subsequent therapies) and the likely immaturity of these endpoints at the time of initial marketing authorisation.

It is notable that neither the FDA nor the EMA reviews considered that the evidence on PFS or OS was sufficiently robust to draw any meaningful conclusions in relation to these endpoints. Instead, both agencies relied on the magnitude of the ORR and DoR as providing evidence to support a potentially meaningful difference in more policy-relevant intermediate (e.g. PFS) and final clinical outcomes (e.g. OS) drawing on existing surrogate relationships. Hence, the surrogate relationships between response-based outcomes (ORR and DoR) are likely to be central to HTA and economic modelling in helping to inform and/or validate longer term extrapolations of PFS and OS due to the likely immaturity of these endpoints.

The following chapters attempt to address these challenges by considering in more detail the nature and design of the trials (Chapter 3) and existing evidence evaluating the use of response-based outcomes as surrogate endpoints for PFS and OS (Chapter 4).

# An overview of key statistical literature addressing the design and analysis of histology-independent trials

The literature on adaptive designs and complex innovative trial designs was reviewed, focussing on trial design and analysis methods proposed for oncology studies and in particular ‘master protocol’ designs proposed to assess histology-independent drugs.

## Adaptive phase II studies

The first step in evaluating a novel treatment is to conduct a Phase II study to determine whether the drug has a sufficient level of disease activity to warrant further investigation. In order to minimise exposure of patients to ineffective drugs, adaptive two-stage designs have been proposed where the second stage of the study is not activated if the first stage shows that the treatment is not effective. The first such design was proposed by Gehan16 in 1961, where the first stage enrols 14 patients and if no responses are observed the trial is terminated. If at least one response is observed in stage 1, the second stage of accrual is activated in order to obtain an estimate of the response probability with a pre-specified standard error. Patients from both stages are used in the estimation of the response rate and an implicit 20% threshold for response rates is considered promising for further study. Fleming17 also studied multi-stage designs with acceptance (i.e. proceed with study) or rejection (stop) possible at each stage based on pre-specified probabilities: *p*0, the largest response probability which, if true, would imply that the drug is not sufficiently effective to warrant further investigation; and *p*1, the smallest probability which would imply that the treatment has a therapeutic effect worthy of further investigation. The acceptable probabilities of making incorrect decisions (type I and type II errors) are also required. In Fleming's design, early rejection only occurs when interim results are quite extreme which permits final analysis to be unaffected by interim monitoring, but is not always desirable for Phase II trials of agents that are likely to be inactive. Although these designs were popular for many years, they did not optimise sample size or allow for early termination when the drug has low tumour activity, a key ethical concern. This led to the development of Simon’s two-stage design,7 which minimises expected sample size when the true response is less than some pre-determined level. Similar to Fleming’s approach,17 investigators pre-specify *p*0, *p*1 and acceptable type I and type II error bounds. This is currently one of the most commonly used adaptive designs, and although it can be extended to multiple stages, in practice only 2 stages are usually used. Extensions of Simon’s two-stage design have been proposed to address uncertainty in the expected response *p*1; if this is too optimistic, then Simon’s design would reject a potentially promising treatment, whereas if it was too pessimistic, it would require more patients to be recruited than necessary.18

Bayesian approaches to adaptive Phase II trials have also been proposed,19-22 which terminate the trial early if the predictive probability that the treatment is not sufficiently effective at the maximum sample size is below a pre-specified level; provide a posterior distribution for the true response probability; and allow the calculation that the true probability of response is above a certain value, or calculate an interval that has a 95% probability of containing the true response proportion (note that this is not provided by confidence intervals obtained using a frequentist approach).

## Master protocol designs

Typically, adaptive Phase II oncology studies are conducted separately for each patient subgroup, based on histology or biomarker activity. However, concerns have been raised about the ability of traditional clinical trial designs to facilitate timely access to innovative technologies, due to the increasingly small populations being targeted in oncology trials, where a traditional Phase III study would never be expected to recruit enough individuals to achieve statistical significance on the primary outcome. The use of complex innovative trial designs with ‘master protocols’, and basket trials in particular, has been proposed to accelerate access to innovative targeted technologies and precision medicine. A consensus statement on their design, conduct and interpretation has recently been published.23 Master protocol trials use a centralised screening platform to identify eligible patients and a common protocol for different sub-studies, which may each focus on patients with specific markers or histologies. The main advantages of master protocols are enhanced patient participation, as more patients are eligible to enter the trial, and a simplification of the trial process since a single protocol is approved for use on multiple sub-studies. Basket trials typically include patients with diverse conditions who share a particular feature or biomarker which can be treated with a single therapy. The key underlying assumption of a basket trial is that the condition depends on the target pathway and that the proposed therapy inhibits this target.24

In oncology, basket trials use a master protocol to define patient eligibility by the presence of a particular biomarker or molecular alteration, regardless of histology. The sub-studies, or baskets, are then defined by a particular histology or other disease-specific characteristics, for example, mutation type. Because individual patients are recruited independently of tumour location or subtype, they are more likely to be eligible for enrolment.24-26 However, a critical consideration is the heterogeneity in prognosis across the different histologies, therefore standardised response rates, reflecting tumour shrinkage, are typically used instead of survival outcomes such as PFS or OS.27 In addition, as the majority of basket trials do not have a control arm, stable disease or survival outcomes would be difficult to interpret unless they were clearly better than what is expected under standard therapy for all tumours.28 Therefore, a further crucial assumption in these designs is that response is a sufficient measure of clinical benefit.

Although designed to improve recruitment, basket trials can still fail to recruit sufficient patients to some or all baskets. For example, the CUSTOM trial29 failed to recruit enough patients for some baskets covering rare mutations. In addition, because they rely on the assumption that molecular profiling is a good predictor of response, they may fail in situations where histological tumour type predicts response better than the biomarkers or mutations defining the baskets.25, 27, 30

Although advocated as ideal, randomisation to a control arm is rare in basket trials,31 due to the differences in standard of care across the different tumour types defining the baskets.23, 25, 28 Adaptive designs for confirmatory basket trials with concurrent (non-randomised) control groups have been proposed and their challenges and limitations discussed.32 However, the lack of a concurrent, randomised, control arm remains a key limitation of these trial designs, and in particular, for the interpretation of such trials in HTA processes.23

Non-randomised basket trials are typically exploratory and use similar two-stage designs to traditional Phase II clinical trials, with each sub-study (basket) analysed separately. Tumour types expected to have a sufficient frequency of the targeted genomic alteration are enrolled into their own basket while others are enrolled into a combined basket. Typically, these studies are designed so that each basket will recruit a certain number of patients and if a certain pre-specified proportion of these patients respond, then the basket is considered “promising” or successful, and either accrual is expanded, or a separate confirmatory study is planned. If not enough responses are observed, the basket is “pruned” due to low promise of efficacy and recruitment to that basket is stopped. Different designs can be used with varying thresholds for response rates selected depending on the indication and prior expectations of efficacy, and with suitable corrections for false-positive rates.27, 33

### Heterogeneity of effect in basket trials

Heterogeneity of effect across different baskets is a key concern. One way to account for this is to analyse each basket separately as if it was an independent study. For example, a basket study of vemurafenib in multiple non-melanoma cancers with BRAF V600 mutations used an adaptive Simon two-stage design18 with stopping rules defined independently for each basket, and considered a response rate at week 8 of 15% to be low, a response rate of 45% to be high, and a response rate of 35% to be low but still indicative of efficacy.30 They found that not all tumour types responded homogeneously to treatment with some tumour types not meeting the pre-specified criteria for response. Similarly, the CUSTOM trial29 used Simon’s optimal two-stage design defining *p*0=0.3 and *p*1=0.6 based on previous literature. The trial aimed to identify targets for molecular biomarkers in NSCLC, small-cell lung cancer (SCLC) and thymic malignancies and to simultaneously evaluate five different targeted therapies in each of the three histologies, which resulted in a total of 15 study arms. A high response rate to erlotinib was identified from only 15 NSCLC patients with an epidermal growth factor receptor (EGFR) mutation but another therapy, selumetinib, failed to achieve a promising response in patients with Kirsten rat sarcoma viral oncogene homolog (KRAS) mutations.29

However, a separate analysis of each basket does not allow for the possibility that some subgroups may react similarly to the drug, particularly if they share a common biomarker which the novel therapy is targeting. By analysing each basket separately, efficiency may be lost by not allowing information gathered from one basket to inform the next, thus increasing the required sample sizes in each basket. In practice, many standard Phase II designs will ignore potential heterogeneity and pool all patients for analysis, which in effect, ignores the specific basket-defining tumour characteristics (e.g. histology) and assumes equal efficacy across all baskets.33 If this approach is taken, then trial planning and analysis are similar to a standard Phase II trial, and for example, Simon’s two-stage design can be used. Although allowing analysis with a much smaller number of included patients, pooling all patients ignores the potential for heterogeneity across baskets, effectively assuming that it is zero, which can miss treatments that are only active in some baskets34 and can lead to large biases in overall estimated effects. In addition, if the drug is truly active or inactive in all baskets, this will be an inefficient design.35, 36

Frequentist adaptive designs for basket studies which try to acknowledge this potential for heterogeneity across baskets have been proposed. In the context of Phase II studies with heterogeneous populations, a design that tests global response across the whole population, whilst allowing a different response for each subgroup, was proposed by London *et al*.37 Simon’s two-stage design was extended to use a more flexible strategy that both tests each subgroup and the combined population, allowing stopping of the trial if either a subgroup or the combined population show futility, i.e. the inability of the study to achieve statistically significant results, at pre-specified thresholds (which are not necessarily the same).38 Negative results in one subgroup would lead to stopping recruitment in that basket alone, unless the combined response for the whole population was below the acceptable threshold. This design leads to smaller sample sizes than separate analyses of each basket when the drug is inactive across all subgroups, and to more power when there is activity in all subgroups. It also retains the individual tests for each subgroup allowing identification of promising baskets. This design requires pre-specification of the expected response rates and prevalence in each subgroup, in order to specify the expected response rate in the overall population. Although the average prevalence in the clinical population may be known, due to the often small samples recruited, the observed prevalence as the trial enrols patients may be quite different. A design which allows adjusting the rejection values depending on the observed prevalence in the trial was proposed by Jung *et al*.39

Cunanan *et al*40 later proposed an efficient study design for the specific scenario of the typical basket trial in oncology, which assesses the homogeneity of the baskets’ response rates at an interim analysis, aggregating the baskets in the second stage (i.e. full borrowing of information) if results suggest effectiveness in all or most baskets, or treating each basket separately (i.e. no borrowing) otherwise. Their basic premise is that the design can be made more efficient by aggregating information from separate baskets in which it can be assumed that the drug has similar efficacy, based on an interim analysis. Thus, the second stage of the design could have a much smaller sample size for the same power to demonstrate clinical efficacy. The first stage of the design is based on the parallel, independent two-stage Simon’s design. When each basket has recruited a small number of patients, the heterogeneity in response across baskets is evaluated. If results support the assumption that the drug’s effects are similar across baskets, then either the trial is terminated for futility (if response is low) or a decision is made to continue to the second stage where all baskets will be pooled for analysis. If there is evidence of heterogeneity across baskets, then the trial will continue only for those baskets showing a promising level of response and these will be analysed separately at the end of the trial. This type of design answers the overall question of efficacy in the whole population more efficiently when there is evidence of homogeneity at an interim stage whilst also shortening trial duration.41 However, this is at the expense of loss of accuracy at assessing efficacy within each separate basket.40 A different approach to testing, which replaces the question of whether there is response to therapy with the question of whether there are differences by tumour type (i.e. across baskets) has also been proposed.42

Although acknowledging the potential for heterogeneity, once a decision has been made on whether heterogeneity is present or not, the analysis proceeds as either separate, independent studies for each basket, or as a single, aggregate, study combining all the baskets. Thus, either complete homogeneity or completely unrelated effects is assumed. A less restrictive assumption is that efficacy is similar (rather than equal or completely different) across baskets, with the different histologies not determining a particular ordering of effectiveness *a priori*, i.e. the baskets are exchangeable. Bayesian hierarchical models43, 44 (BHM) are particularly suited for this situation as they estimate the heterogeneity and allow borrowing of information on the effects of the treatment across baskets, increasing precision of estimates compared to analysing all baskets separately, whilst reducing the chances of obtaining extreme estimates in baskets with few patients. Thall *et al*44 proposed a BHM which produces estimates of efficacy (e.g. probability of response) for each basket that are shrunken towards the mean efficacy (e.g. pooled probability of response) across all baskets. The model is an extension of a Bayesian Phase II design where the trial is stopped if the posterior probability that the response rate is at least falls below a pre-specified cut-off and can be applied to both binary and time-to-event data.45 Each basket is assumed to have a different treatment effect (event probability or event rate) *θj* and these are assumed exchangeable (i.e. similar) and correlated *a priori*. Specifically it is assumed that the *θj* follow a BHM, while allowing a separate stopping rule for each basket. Thus, the model will identify subgroups where results are not promising which can be dropped at a subsequent stage. Because the effects are assumed to be correlated across baskets, data from each individual basket will provide information on the effects in all the other baskets, so that for example, a longer survival time for a patient in a given basket will increase the posterior distributions of all *θj*, on average. In other words, information is borrowed across baskets, which shrinks the observed effects towards the pooled mean effect. Outputs from the resulting analysis include the posterior distributions for the effect (e.g. response or event rate) in each basket, posterior distributions for the pooled effect across all baskets, as well as the posterior distribution for the heterogeneity across baskets. In addition, a predictive distribution for the effect in a new study sampling baskets from the same overall population, can be calculated to reflect the full degree of uncertainty both due to the sample size and the observed heterogeneity in effects across the observed baskets. A Phase II trial of imatinib in 10 histologic subtypes of sarcoma used this design: accrual within a sarcoma subtype would stop if it was unlikely that its response rate was at least 30%.34, 44, 46

The BHM was shown to be a better design for a single-arm, non-randomised trial with a tumour response endpoint, when there is a possibility of different effects in different subgroups of patients, when compared to Simon’s optimal two-stage design and to the Bayesian adaptive design with no borrowing.47 However, the hierarchical borrowing can make it more difficult to find a single basket where the treatment is promising, although it is more likely than the other designs to correctly conclude futility or efficacy.

Any borrowing and precision gains from a BHM are only advantageous if the exchangeability assumption is reasonable. An approach for assessing homogeneity at an interim analysis and proceeding with a BHM in the second stage only if efficacy is deemed reasonably homogeneous, has been proposed.48 This approach avoids problems caused by implementing a complete pooling model at the second stage40 or proceeding with a fully exchangeable BHM when there is evidence of outlying baskets.

Hierarchical designs have been criticised when there is insufficient information in the outcome data to determine whether borrowing across subgroups is appropriate.34, 49 In addition, unknown between-subgroup heterogeneity, which drives the amount of borrowing, poses a major problem when the number of baskets is small (less than 10, as a rule of thumb)34, because it cannot be well inferred from the data and results will be sensitive to model specification, in particular to the specification of the prior distribution for the borrowing parameter.34, 50 Alternatives to complete pooling or borrowing across all baskets have been proposed, which extend the BHM to allow borrowing of information across similar baskets while avoiding too optimistic borrowing for extreme baskets.49, 51-55

A model which allows non-exchangeable prior distributions to be specified was proposed for the scenario where it is not expected *a priori* that all subgroups will be exchangeable. For example, some tumour types may be associated with a better or worse prognosis and their response to treatment is expected to differ. Different models can be used to implement this assumption: we can accept that a particular tumour characteristic (e.g. prognosis) defines exchangeability so that different categories are formed and exchangeability is only allowed between tumours in the same category (e.g. poor, intermediate, good prognosis), or we can treat the appropriate grouping as a random quantity to be estimated from the data, indexed by a categorical covariate of interest (e.g. prognosis).51 Thus, the estimation of the treatment effect for a particular subgroup borrows more strength from other subgroups that, according to the prior beliefs, are more likely to be exchangeable, but the models allows the data to correct any prior beliefs which are not supported by the available data. When there is no *a priori* information on which subgroups might be exchangeable or not, an exchangeable-nonexchangeable model49 allows for selected special exchangeability patterns specified in the model to be determined by the treatment response data. This model extends the BHM to allow *θj* to be either exchangeable with some of the other subgroups, or nonexchangeable with any of them, in which case the effect will be estimated independently of all other subgroups. Prior weights for the exchangeable probability of each subgroup are specified to reflect an *a priori* belief that a subgroup behaves systematically differently to the others. Essentially, the model determines whether some borrowing or no borrowing of information should be carried out across subgroups. Outputs include a global heterogeneity parameter across subgroups and mixture weights which describe the similarity of subgroups in the exchangeable component of the model whilst also identifying subgroups which behave differently (i.e. show a low probability of being exchangeable). Although a pooled mean effect for the exchangeable component of the model can be obtained, focus is on the effects for each individual subgroup, which incorporate different levels of borrowing according to the model. The prior distributions specified for the heterogeneity parameter and for the exchangeability weights can influence the results and need to be specified carefully. The use of this model for trial design requires careful consideration of the specification of the prior distributions and mixture weights, but has been found to perform well in various scenarios.49 Extensions of these ideas to incorporate more information and thus improve performance of the trial design or simplify computation have been proposed. For example, the Bayesian latent subgroup trial design52 defines different latent subgroups within which more borrowing is allowed by jointly modelling biomarker measurements and treatment responses. This allows grouping of different cancers according to biomarker measurements routinely collected during a trial, effectively using internal trial information to inform the adaptive borrowing which determines the decision to proceed to the next stage. Fujikawa *et al*54 proposed a Bayesian basket design which borrows information across the subgroups which have the most similar posterior distributions based on a pre-specified threshold of similarity, which is simple to compute. Decisions can be made at the interim stage to stop or continue with the trial and this design can also determine which subgroups show efficacy in the final analysis, based on pre-defined criteria. Unlike the fully exchangeable BHM, in these models, obtaining and interpreting predictive distributions of effects is not meaningful, since we can no longer reasonably assume that a new tumour type (subgroup) would have been sampled from the same distributions as the observed subgroups, i.e. we cannot assume all subgroups are exchangeable.

Due to the increased number of parameters being estimated, the hierarchical approach may increase uncertainty unnecessarily if response to treatment is indeed homogeneous across all subgroups. Therefore, when there is a strong rationale for expecting a uniform level of response it may be preferable to use a simple pooling of information across subgroups.34 However, *a priori* assumptions of homogeneity in trial design or analysis need to be carefully justified as in most cases, basket trials include patients with very clinically heterogeneous tumour types. In addition, the available empirical evidence does not generally support the assumption of homogeneity of activity of drugs across different histologies.

Previous basket trials have shown heterogeneity in the effectiveness of agents across tumour types, which lends support to the *a priori* assumption that effects may be heterogeneous. A recent trial of vemurafenib in 122 patients with *BRAF* V600–mutated cancers across multiple tumour types (including CRC, NSCLC, Erdheim–Chester disease and Langerhans’-cell histiocytosis, primary brain tumours, cholangiocarcinoma, anaplastic thyroid cancer) found evidence of response in some tumour types including NSCLC and Erdheim–Chester disease and Langerhans’-cell histiocytosis, but not in CRC.30 This heterogeneity in response was also observed in previous separate, independent, studies which showed a positive response to vemurafenib of patients with BRAF-positive metastatic melanoma,56 but not in BRAF-positive colon cancer patients.57 A trial of imatinib, a tyrosine kinase inhibitor, that included 196 patients across 40 different subtypes, found evidence of activity of imatinib in only five malignancies.58 Another basket trial of imatinib in 10 histologic subtypes of advanced sarcoma concluded that although rare dramatic responses were seen, imatinib was not an active agent in these subtypes, although it had previously shown effectiveness in another subtype of soft tissue sarcoma.46 Similarly trastuzumab which is known to be effective in the treatment of women with HER2-positive breast cancer59 was not shown to be effective in HER2-positive recurrent endometrial cancer60 or HER2-positive non-small-cell lung cancer.61 This evidence suggests that the treatment effects in different cancer types may not in fact be exchangeable. Therefore, the design of basket trials should allow for the possibility of heterogeneity in treatment effects across tumour types, only opting for a design that assumes homogeneity in very special cases or where data from previous stages clearly supports it.

## Summary and implications

Complex innovative study designs are being used to address multiple clinical questions in an attempt to speed up regulatory approval and access to patients of drugs with new mechanisms of action. Adaptive basket trials are particularly suited to assess efficacy of histology-independent drugs, although their reliance on surrogate outcomes, small sample sizes and mostly uncontrolled designs pose challenges for HTA.

A recent consensus statement has provided recommendations for the planning, design and statistical analysis of complex study designs including considerations on ensuring their relevance for HTA.23 These include encouraging comparative randomised studies, ensuring that the primary outcome, typically a surrogate of the clinical outcome of interest in HTA, is likely to adequately predict the clinical outcomes of interest, and using analysis methods that allow borrowing of information across baskets.23

Although it is challenging to determine the correct level of borrowing of information (exchangeability) across baskets,23 the approaches described in Section 3.2.1 allow the treatment effect in any basket to be informed by the effects in all other baskets therefore maximising the information available. Their interpretation and potential use in NICE TAs is described in Chapters 6 and 7.

# A systematic review to identify published meta-analyses evaluating the use of response rates and duration of response as surrogate endpoints for progression-free and overall survival

## Introduction

It is generally accepted that decisions about the use of new and existing health technologies should ideally be informed by estimates of treatment effects derived from high quality RCTs which measure patient-relevant endpoints over a clinically appropriate timeframe. Such “final” endpoints typically involve the measurement of health benefits and adverse events which reflect aspects of the disease and its treatment which are important to patients (and potentially also their carers) and which relate to “*how the patient feels, functions or survives*.”62 In the context of evaluating treatments for advanced/metastatic cancer, the key matter of concern is often whether the use of a given heath technology leads to improvements in OS (a final endpoint) compared to existing standard treatments. However, the estimation of treatment effects on OS may be subject to numerous problems, including: potential confounding resulting from the use of post-progression treatments, insufficient study follow-up resulting in data immaturity, or simply that data on OS have not been collected. In such instances, determining the impact of health technologies becomes more challenging and may rely on the use of other surrogate or intermediate endpoints to estimate treatment effects on final endpoints. These surrogate endpoints are intended to substitute for, and predict, a final patient-relevant clinical outcome.63 In terms of advanced/metastatic cancer, potentially relevant surrogate endpoints may vary according to the tumour type and site, but commonly include PFS, time to progression (TTP), and response-based outcomes, which may include ORR, CR, PR, very good partial response (VGPR) and DoR. These surrogate endpoints are often considered attractive as they typically require smaller sample sizes, occur faster and are less expensive to collect in clinical trials compared with final outcomes, thereby reducing costs associated with data collection and expediting the time required for bringing new technologies to market.

It has long been recognised that the reliance on surrogates may lead to invalid conclusions regarding the net health effects of technologies, which in turn, have the potential to lead to patient harm.64 Much of the published literature around the use of surrogate endpoints has focussed on the development and application of frameworks for their validation.65, 66 In his seminal paper, Prentice65 put forward stringent criteria for the validation of surrogate endpoints in Phase III trials. In general terms, these criteria require that the surrogate endpoint must be a correlate of the net effect of treatment on the final clinical outcome – in other words, there must be a single pathway from the treatment to the true endpoint which is mediated exclusively by the surrogate endpoint.67 Applied surrogate validation studies commonly adopt a meta-analytic (meta-regression) approach based on multiple studies in order to assess whether the apparent relationship between the surrogate and the final endpoint remains constant in the presence of various sources of heterogeneity, such as differences in patient population, study design and treatments received.66

Based on the National Institutes of Health (NIH) Biomarkers Definition Working Group’s preferred terms and definitions68 and the 2001 JAMA User’s Guide,69 Taylor and Elston70 proposed a hierarchy of levels of surrogate validation. Level 3 of the hierarchy relates to biological plausibility – this is the weakest form of validation and is typically based on pathophysiological studies and/or an understanding of the disease process. Level 2 requires the presence of a consistent association between the surrogate outcome and the final endpoint; this may be assessed using observational studies or arm-based analyses of trials which have measured both the surrogate and the final outcome. This level of validation requires an assessment of the individual-level (absolute) association between endpoints, and is usually undertaken using correlation analysis. Level 1 of the hierarchy represents the strongest level of surrogate validation: in order to achieve this level of validation, the treatment effects on the surrogate outcome must correspond to a commensurate treatment effect on the final outcome. Demonstrating this level of validity requires an analysis of correlation in terms of treatment effects between arms based on data from RCTs (sometimes referred to as trial-level association). Other validation frameworks have been proposed to assess the strength of association between surrogate and final endpoints. These include the criteria proposed by the German Institute of Quality and Efficiency in Health Care71 (IQWiG; based on the treatment effect association only) and the Biomarker-Surrogate Evaluation Schema criteria72 (BSES2; based on both absolute and treatment effect associations). These frameworks differ in terms of the types of analyses and the strength of the relationship required to determine the reliability of the surrogate.

The means by which health economic models use information on relationships between surrogate and final endpoints differs between appraisals, but may be broadly categorised into two general situations: (i) data are available on both the surrogate and final endpoints from one or more studies relating to the technology under consideration, and the relationship between surrogate and final endpoints is not informed by external data (and in some instances may not be quantified at all); and (ii) data are available on the impact of the technology on the surrogate endpoint, but information relating to the final endpoint from the same study is not available or is not used to inform the model – in this case, external data (e.g. meta-regressions and/or other forms of predictive model) may be required to quantify the relationship between the surrogate and final endpoint. This review is more relevant to the latter situation, whereby the degree of confidence which can be placed in the results of the model may be influenced by judgements about whether the surrogate can be considered valid.

In the context of histology-independent treatments, data on OS, and potentially also other time-to-event outcomes such as PFS, are likely to be immature. Consequently, there may be a need to rely on surrogate outcomes, such as response rate, using data from external sources, in order to estimate other more clinically meaningful final outcomes. This section presents a systematic review of response-based outcomes as surrogates for PFS, TTP and OS in advanced or metastatic cancer, across any tumour site. The review focuses on meta-analyses and meta-regressions. Analyses are presented both for absolute associations and treatment effect associations between response-based outcomes and PFS, TTP and/or OS. In addition, the IQWiG and BSES2 criteria are used to assess the strength of association between surrogate and final endpoints. Where data permit, the review also explores the surrogate threshold effect (STE) associated with response-based outcomes: this corresponds to the smallest treatment effect on the surrogate that predicts a non-zero treatment effect on the true endpoint.73

Where available, the results of published regression models are also reported; if ORR was deemed to be valid in one or more tumour types, one option would be to use the coefficients from models to quantify the relationship between ORR and PFS/OS. Other approaches for incorporating surrogate outcomes in health economic models are discussed at the end of the chapter.

## Methods

### Review question

This systematic review sought to address the following research question: *“What is the strength of the association between response outcomes and PFS, TTP or OS across different types of cancer (primarily advanced or metastatic), based on meta-analyses or meta-regression studies assessing the statistical relationship between these outcomes?”*

### Inclusion and exclusion criteria

The inclusion and exclusion criteria for the review are shown in Table 1. Inclusion was restricted to articles reporting meta-analyses and meta-regressions across multiple studies, and reporting the strength of association between response outcomes (ORR, CR, PR, VGPR or DoR) and either PFS, TTP or OS. The included meta-regressions could themselves include RCTs and/or single-arm studies. However, individual reports analysing single trials or single cohorts were excluded. Included meta-analyses could report absolute associations and/or treatment effect associations. These associations had to be reported as a correlation coefficient (e.g. Pearson r or Spearman rs) and/or a coefficient of determination (R2) between relevant outcomes.

Studies of any cancer and any treatment were included. The review focussed mainly on studies of advanced or metastatic cancers (and/or treatment with palliative intent), as these studies were more likely to report PFS and OS. However, studies reporting relevant outcomes were included even where the stage was not specifically restricted to advanced/metastatic disease for all patients or where this was unclear (this applied particularly to haematological cancers). Studies were excluded if they explicitly referred to adjuvant or neo-adjuvant treatment, or treatments which are given with curative intent.

Table 1: Inclusion and exclusion criteria

| **Field** | **Inclusion** | **Exclusion** |
| --- | --- | --- |
| Disease area | * Any cancer * Mainly advanced or metastatic cancer, and/or where treatment intent was palliative * Also studies reporting relevant outcomes where stage was not restricted to advanced/metastatic or where this was unclear (particularly haematological cancers) | * Treatment with curative intent * Neo-adjuvant treatment * Adjuvant treatment |
| Surrogate endpoints | Response endpoints:   * Overall response rate (ORR = CR+PR) * Complete response (CR) * Partial response (PR) * Very good partial response (VGPR) * Duration of response (DoR) | * Other endpoints |
| Final endpoints | * Progression-free survival (PFS) * Time to progression (TTP) * Overall survival (OS) | * Other endpoints |
| Study and data type | * Meta-analyses and meta-regressions across multiple studies * Included meta-analyses could include RCTs and/or single-arm studies * Included meta-analyses could use aggregate data (e.g. medians per study arm) and/or individual patient data (IPD) | * Analyses of single trials or single cohorts |
| Type of analysis reported | * Studies must report absolute associations and/or treatment effect associations between relevant endpoints (see above) * Associations must be expressed as a correlation coefficient (e.g. Pearson r or Spearman rs) and/or as a coefficient of determination (R2) | * No correlation coefficient or regression R2 reported |
| Language | * English language * Other language if sufficient detail in English abstract | * Non-English with insufficient detail |
| CR, complete response; DoR, duration of response; IPD, individual patient data; OS, overall survival; PFS, progression-free survival; PR, partial response; RCT, randomised controlled trial; ORR, overall response rate; TTP, time to progression; VGPR, very good partial response. | | |

### Search strategy

Five databases (MEDLINE, EMBASE, Web of Science, the Cochrane Database of Systematic Reviews and CINAHL) were searched from inception to March 2019. Search terms included: cancer terms AND response terms AND terms for PFS, TTP and/or OS AND terms for regression, correlation, prediction, association or relationship AND terms for endpoint and/or surrogate. Search results were limited to the English language and to studies undertaken in humans. The MEDLINE search strategy is provided in Appendix 3.

In addition, a citation search was undertaken based on two existing meta-reviews of surrogate relationships; this identified studies which have cited any of the 48 articles included in the review by Fischer *et al*. (2016)74 and/or any of the 19 articles included in the review by Davis *et al*. (2012).75 In addition, relevant existing meta-reviews, including Fischer *et al*. (2016),74 Davis *et al*. (2012),75 Savina *et al*. (2018),76 Haslam *et al*. (2019),77 and any further reviews identified during searching, were checked for relevant studies.

### Study selection process

Titles and abstracts of articles retrieved by the search were examined by one reviewer and a subset were checked by a second reviewer early in the process, followed by a discussion to ensure consistency in the selection decisions. Full texts were examined by one reviewer and a subset were checked by a second reviewer, with any discrepancies resolved through discussion.

### Data extraction

Data were extracted by one reviewer and all data were checked by a second reviewer. The following data were extracted:

* Author and date
* Cancer type, stage, number of patients, number of included studies, design of included studies (RCT or single-arm; publication dates)
* Treatment type, treatment line, other subgroups as reported
* Data type (aggregate-level data or individual patient data [IPD])
* Surrogate and final endpoints analysed (e.g. ORR to OS)
* Response criteria used, if reported (e.g. RECIST)
* Measures of outcomes (e.g. hazard ratio [HR], odds ratio [OR], relative risk (rr) or difference between medians)
* Statistical methods for correlation and regression, whether weighted, whether adjusted, coefficient reported, e.g. Pearson or Spearman correlation coefficient (r or rs), regression coefficient of determination (R2).
* Absolute association results (i.e. between absolute values of the surrogate and final endpoints, based on data from individual arms of RCTs or single-arm studies): correlation coefficient, regression R2, regression equation
* Treatment effect association results (i.e. between treatment effects for surrogate and final endpoints, based on between-group differences from RCTs): correlation coefficient, regression R2, regression equation
* Data as above for subgroups
* STE,73 i.e. the smallest treatment effect on the surrogate that predicts a non-zero treatment effect on the true endpoint.

### Data synthesis

Data were tabulated and described in a narrative synthesis. Plots were constructed to illustrate the reported associations. Some of the included meta-regression studies reported multiple subgroup analyses with differing results. Therefore, for associations between absolute values of endpoints, the plots show the range of correlation coefficients per study, across all subgroup analyses. Where an included meta-regression study reported on more than one cancer type, these are shown separately on the plots. All types of correlation coefficient were included, e.g. Pearson r and Spearman rs. If no correlation coefficient was reported, then Pearson r was calculated as the square-root of R2, if available.

For associations between treatment effects, the plots show the range of regression coefficients of determination (R2) per study, across all subgroup analyses. The plots include both adjusted and unadjusted R2 values, as well as values from weighted and unweighted regressions. For studies in which R2 was not reported, this was calculated as the square of the Pearson (r) correlation coefficient, if available. R2 was not calculated from other correlation coefficients such as Spearman, or where the method of correlation was unclear.

### Scoring the strength of association

Two separate sets of criteria have been developed to assess the strength of association between endpoints. These include the criteria proposed by IQWiG71 (based on the treatment effect association) and the BSES2 criteria72 (based on both absolute and treatment effect associations). In this review, both the IQWiG and BSES2 criteria were used to assess the strength of association between surrogate and final endpoints.

The IQWiG criteria71 (Table 2) are based on the correlation coefficient (r) for the treatment effect association. Where r was not reported, it was calculated as the square-root of R2, if available. Some slight modifications were made to the IQWiG scoring criteria, because the medium score bracket was not clearly defined (see Table 2); these modifications were based on the approach used in the previous review by Savina *et al*.76 The IQWiG score was generated based on the magnitude of r, irrespective of its sign (i.e. a negative correlation could generate a high score).

Table 2: IQWiG scoring criteria71

|  |  |
| --- | --- |
| **IQWiG Score** | **Criteria (based on r for treatment-effect association)\*** |
| High | Lower confidence interval of r is ≥ 0.85 |
| Medium+ | r ≥ 0.85 with no reported confidence interval **or** r ≥ 0.85 with wide confidence intervals (lower limit <0.85) |
| Medium | 0.85 > r ≥ 0.7 and upper confidence interval of r is ≥ 0.7 and lower confidence interval of r is < 0.85, **or** 0.85 > r ≥ 0.7 with no reported confidence interval |
| Low | Upper confidence interval of r is < 0.7 **or** r < 0.7 with no reported confidence interval |
| Notes:  \*r is defined as any correlation parameter for the treatment-effect association, e.g. Pearson, Spearman, Kendall's Tau. Where no correlation parameter was reported, if a univariate regression was performed and an R2 value attained, then r (Pearson correlation coefficient) was calculated as the square-root of R2. The reported r could be for any treatment effect estimate (hazard ratio, difference in medians, etc.); where more than one was reported, relative estimates (e.g. hazard ratio, odds ratio) were used in preference to difference in medians. The Medium+ category was based on the approach used in Savina *et al*.76 | |

The BSES2 criteria72 (Table 3) require R2 values for both the individual and treatment effect associations. Where R2 was not reported, it was calculated as the square of r, if available. BSES2 criteria were used as an adaptation from the original BSES criteria, as described in Savina *et al*. (2018).76 The original BSES criteria require R2 for both individual and treatment effect associations and a value for the STE. Since so few articles report STE, this review used BSES2, which does not require the STE.

Table 3: BSES2 scoring criteria72

|  |  |
| --- | --- |
| **BSES2 score** | **Criteria (based on R2 for both treatment effect and individual-level associations)\*** |
| Excellent | R2 (treatment effect) ≥ 0.6 and R2 (absolute) ≥ 0.6 |
| Good | R2 (treatment effect) ≥ 0.4 and R2 (absolute) ≥ 0.4 |
| Fair | R2 (treatment effect) ≥ 0.2 and R2 (absolute) ≥ 0.2 |
| Poor | R2 (treatment effect) < 0.2 and/or R2 (absolute) < 0.2 |
| Notes:  \* R2 is the coefficient of determination for a regression analysis. Where R2 was not reported, it was calculated as the square of the Pearson correlation coefficient (r), if available. The reported R2 could be for any treatment effect estimate (hazard ratio, difference in medians, etc.); where more than one was reported, relative estimates (e.g. hazard ratio, odds ratio) were used in preference to difference in medians. | |

## Results

### Number of included studies

The literature search generated 2,829 citations (Figure 1), of which 2,630 were excluded during the review of titles and abstracts. In total, 64 references to 63 studies were included in the review.78-141 Study characteristics for the 63 included studies are shown in Appendix 4. Detailed results of included studies are shown in Appendices 5 and 6. Studies excluded at the full-text stage, with reasons for exclusion, are listed in Appendix 7.

Figure 1: PRISMA flow diagram for study inclusion

Total references identified  
(n = 2,829)

References excluded at title/abstract stage  
(n = 2,630)

Full-text references excluded (n = 135):

Not clinical study (n = 4)

Not meta-analysis of multiple studies (n = 28)

Neo-adjuvant or adjuvant (n = 10)

No relevant outcomes (n = 67)

No correlation coefficient or R2 (n = 13)

Secondary publication, no additional data (n = 10)

Insufficient data reported (n = 1)

Non-English and insufficient detail (n = 1)

Not available (n = 1)

Included in systematic review

(n = 64 references to 63 studies)

References identified from database searches  
(n = 2,822)

- Keyword search (n = 1102)

- Citation search (n = 1759)

- Both (n = 39)

References identified from other sources e.g. other reviews (n = 7)

- From previous reviews (n = 5)

- Chance find (n = 2)

Full-text references screened  
(n = 199)

### Characteristics of included studies

Full details of study characteristics for the 63 included studies are shown in Appendix 4 (note that 8 references78, 106, 108, 125, 127, 137, 138, 141 appear on more than one row as they report on more than one cancer type).

#### Surrogate relationships, cancer types and treatments

A summary of surrogate relationships, cancer types and treatments is provided in Table 4. The most commonly reported surrogate relationships were ORR to OS (57 studies), ORR to PFS (22 studies), CR to OS (8 studies) and CR to PFS (7 studies). Other response outcomes (DoR, PR, VGPR/CR) were only reported in 1-2 studies each.

Twenty different cancer types were analysed (Table 4), the most common being NSCLC (16 studies), CRC (10 studies), various solid tumours (8 studies) and breast cancer (5 studies). Disease stage was advanced/metastatic in 43 studies and unclear in 9 studies (Table 4), while the remainder (11 studies) gave other descriptions mostly indicating advanced, extensive or recurrent disease. Treatment was first-line in 23 studies, later lines or combinations of lines in 32 studies, and not reported in 8 studies. Treatment type was chemotherapy in 21 studies, immune checkpoint inhibitors in 9 studies, targeted therapy in 8 studies, and various other treatment combinations in the remainder.

#### Data types reported

Table 5 summarises the data types reported in the included meta-regressions. The various meta-regressions included between 4 and 191 primary studies and between 407 and 44,125 patients each. The majority of meta-regressions (N=44) included only RCTs, while 17 included both RCTs and single-arm studies and 2 included only single-arm studies. Most meta-regressions (N=58) analysed aggregate data (e.g. medians or other summary measure per study arm), whilst 5 analysed IPD. Across all meta-regressions, 32 reported absolute (individual-level) associations, 38 reported treatment effect (trial-level) associations, and only 4 reported the STE.

Table 4: Study characteristics: Surrogate relationships, cancer types and treatments

| **Surrogate relationship** | **N** | **Cancer type** | **N** | **Disease stage** | **N** | **Line of treatment** | **N** | **Treatment type** | **N** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ORR to OS  ORR to PFS  CR to OS  CR to PFS  DoR to OS  ORR to TTP  PR to PFS  PR to OS  VGPR/CR to PFS  DoR to PFS | 57  22  8  7  2  1  1  1  1  1 | Lung (NSCLC)  Colorectal  Various solid  Breast  NHL  Lung (SCLC)  Ovarian  Pancreatic  Renal cell  Gastric  Neuroendocrine  Soft tissue sarcoma  Urothelial  AML  Biliary tract  Gastroesophageal  Glioblastoma  Multiple myeloma  Prostate  Unknown primary | 16  10  8  5  4  3  3  3  3  2  2  2  2  1  1  1  1  1  1  1 | Advanced/metastatic  Unclear  Advanced, locally advanced, unresectable or metastatic  Extensive disease  Limited or extensive disease  Advanced or recurrent  Advanced, locally advanced or recurrent  Relapsed / refractory  Most stage III/IV  Recurrent / platinum-resistant  Various | 43  9  2  2  1  1  1  1  1  1  1 | 1st  All / various  NR  1st + 2nd  2nd  2nd + subsequent  2nd + 3rd | 23  18  8  5  4  3  2 | Chemo  Immune checkpoint inhibitors  Targeted  Various  Systemic  Chemo or targeted  Chemo, immune or targeted  NR  Chemo + targeted  Chemo or immune  Chemo, hormonal + targeted  Chemo or biologic  Cytokine or targeted  Gemcitabine + chemo or targeted  Bevacizumab + chemo | 21  9  8  7  5  3  2  1  1  1  1  1  1  1  1 |
| Note: Ns may sum to more than total number of studies (N=63) as some studies reported more than one surrogate relationship or cancer type.  AML, acute myeloid leukaemia; chemo, chemotherapy; CR, complete response; DoR, duration of response; immune, immunotherapy; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; SCLC, small cell lung cancer; TTP, time to progression; VGPR, very good partial response. | | | | | | | | | |

Table 5: Study characteristics: Data types

| **N primary studies per meta-regression (range)** | **N patients per meta-regression (range)** | **Included study types per meta-regression** | **Data types** | **Absolute association reported?** | **Treatment effect association reported?** | **STE reported?** |
| --- | --- | --- | --- | --- | --- | --- |
| 4 to 191 | 407 to 44,125 | RCT only (N=44)  RCT+SA (N=17)  SA only (N=2) | AD (N=58)  IPD (N=5) | N=32 | N=38 | N=4 |
| AD, aggregate data; IPD, individual patient data; RCT, randomised controlled trials; SA, single-arm studies; STE, surrogate threshold effect. | | | | | | |

### Results of included studies

#### Absolute (individual-level) correlation and regression

The range of absolute (individual-level) correlation coefficients reported in each meta-regression is summarised in Table 6 and illustrated in Figure 2 (for the association between ORR and PFS) and Figure 3 (for the association between ORR and OS). Each horizontal row in the plots illustrates the range of correlation coefficients across all subgroup analyses within a single meta-regression study. Where an included meta-regression reported on more than one cancer type, these are shown separately on the plots. It is worth noting that the meta-regressions varied both in terms of the number of included primary studies (shown as N on the plots) and in terms of the treatment type, line of treatment and precise clinical population; all these details are provided in Appendix 5, together with correlation coefficients for all individual subgroup analyses.

**ORR and PFS (or TTP):** The reported correlation coefficients (Pearson r or Spearman rs) between absolute ORR and PFS ranged from -0.72 to 0.96, based on multiple analyses within 12 studies across 10 cancer types104, 105, 112, 114, 115, 119, 122, 123, 125, 126, 132, 138 (Figure 2 and Table 6; full details in Appendix 5). Across those studies which report only a single analysis, the correlation coefficient was generally above 0.60; however, some estimates were lower. Confidence intervals around the correlation coefficients were rarely reported (not shown in the plot, see Appendix 5). Few separate meta-regressions reported on the same tumour site, hence it is difficult to assess whether ORR may be a more reliable surrogate in certain cancer types than others. One study reported on ORR and TTP (gastric cancer; correlation rs = 0.41 to 0.56 across subgroup analyses, not shown on the plot).102

**ORR and OS:** The reported correlation coefficients between absolute ORR and OS ranged from ‑0.40 to 1.00, based on 27 studies across 15 cancer types78, 103, 79, 80, 95, 97, 98, 102, 105, 109-112, 119-126, 128, 130-132, 135, 138 (Figure 3 and Table 6; full details in Appendix 5). Confidence intervals around the correlation coefficients, where reported, were generally fairly wide (not shown in the plot). The majority of correlation coefficients were above 0.40; however, several estimates were lower. The correlation coefficients reported from multiple analyses within the same study, and those reported across separate studies, did not suggest a clear pattern by cancer type.

**CR and PFS or OS:** The correlation coefficients between absolute CR and PFS in two studies of small-cell lung cancer (SCLC)119 and non-Hodgkin’s lymphoma (NHL)141 ranged from 0.22 to 0.83, while the correlation coefficients between absolute CR and OS ranged from ‑0.04 to 0.62, based on 3 studies of NSCLC,109 SCLC119 and gastroesophageal cancer121 (Table 6; full details in Appendix 5).

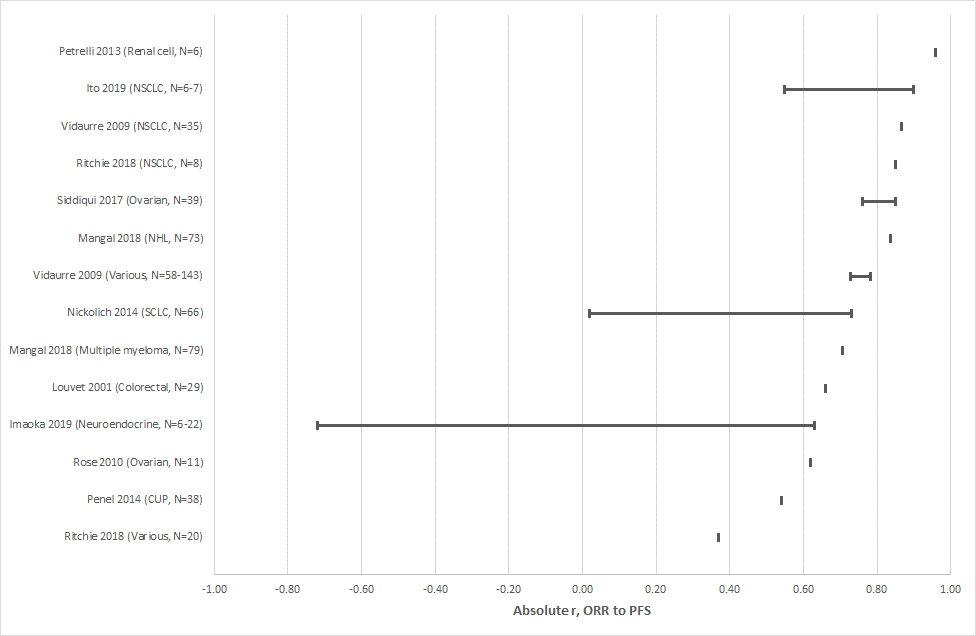
**PR and PFS or OS:** The correlation coefficient between absolute PR and PFS ranged from 0.35 to 0.70 across subgroup analyses within one study of SCLC,119 while the highest correlation coefficient between absolute PR and OS ranged from 0.29 to 0.66 in the same study119 (Table 6; full details in Appendix 5).

**DoR and PFS or OS:** No studies reported on the absolute association between DoR and PFS or OS.

Table 6: Summary of absolute (individual-level) correlations per study

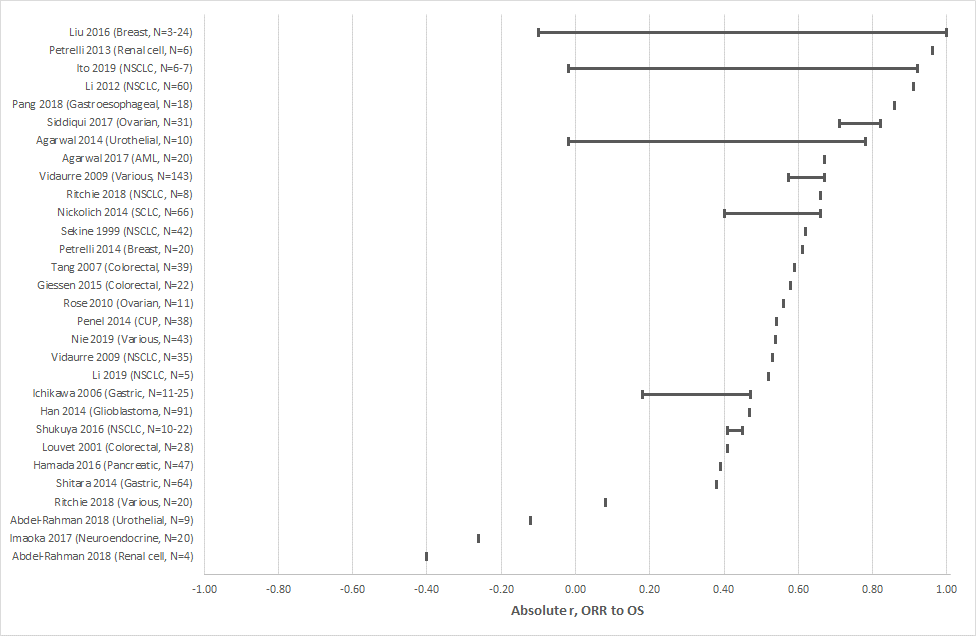
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Surrogate relationship** | **N studies** | **Cancer types and refs** | **Range of r or rs across studies and subgroup analyses** | **Further detail** |
| **ORR to PFS** | 12 | NSCLC,105, 125, 138 ovarian,126, 132 RCC,123 NHL,114 SCLC,119 MM,115 CRC,112 CUP,122 NET,104 various138 | NSCLC,105, 125, 138 ovarian,126, 132 RCC,123 NHL,114 SCLC,119 MM,115 CRC,112 CUP,122 NET104 | NSCLC,105, 125, 138 ovarian,126, 132 RCC,123 NHL,114 SCLC,119 MM,115 CRC,112 CUP,122 NET104 |
| **ORR to TTP** | 1 | Gastric102 | Gastric102 | Gastric102 |
| **ORR to OS** | 27 | NSCLC,105, 109, 110, 125, 128, 131,138 CRC,95, 112, 135 ovarian,126, 132 breast,111, 124gastric,102, 130 various,125,120, 138 pancreatic,97 RCC,78,123 gastroesophageal,121 urothelial,78,79 AML,80 SCLC,119 glioblastoma,98 CUP,122 NET103 | **Pos:** NSCLC,105, 109, 110, 125, 128, 131,138 CRC,95, 112, 135 ovarian,126, 132 breast,111, 124gastric,102, 130 various,125,120, 138 pancreatic,97 RCC,78,123 gastroesophageal,121 urothelial,78,79 AML,80 SCLC,119 glioblastoma,98 CUP,122 NET103 | **Pos:** NSCLC,105, 109, 110, 125, 128, 131,138 CRC,95, 112, 135 ovarian,126, 132 breast,111, 124gastric,102, 130 various,125,120, 138 pancreatic,97 RCC,78,123 gastroesophageal,121 urothelial,78,79 AML,80 SCLC,119 glioblastoma,98 CUP,122 NET103 |
| **CR to PFS** | 2 | SCLC,119 NHL141 | SCLC,119 NHL141 | SCLC,119 NHL141 |
| **CR to OS** | 3 | NSCLC,109 SCLC,119 gastroesophageal121 | NSCLC,109 SCLC,119 gastroesophageal121 | NSCLC,109 SCLC,119 gastroesophageal121 |
| **PR to PFS** | 1 | SCLC119 | SCLC119 | SCLC119 |
| **PR to OS** | 1 | SCLC119 | SCLC119 | SCLC119 |
| **VGPR/CR to PFS** | 0 |  | (see footnote)\* | Appendix 5 |
| **DoR to PFS** | 0 |  | - | - |
| **DoR to OS** | 0 |  | - | - |
| Notes: Further detail on all studies and outcomes is shown in Appendix 5. N studies per outcome may vary from Table 4 since not all data in correct format. \*One study of MM reported VGPR/CR to PFS as adjusted R2=0.64 but this could not be converted to r because it was adjusted.115  AML, acute myeloid leukaemia; CR, complete response; CRC, colorectal cancer; CUP, cancer of unknown primary; DoR, duration of response; MM, multiple myeloma; NET, neuroendocrine tumour; NHL, non-Hodgkin’s lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; RCC, renal cell carcinoma; SCLC, small cell lung cancer; TTP, time to progression; VGPR, very good partial response. | | | | |

Figure 2: Correlation (r or rs) between absolute (individual-level) values of ORR and PFS\*



\*For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. CUP, cancer of unknown primary; NHL, non-Hodgkin’s lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; PFS, progression-free survival; SCLC, small cell lung cancer.

Figure 3: Correlation (r or rs) between absolute (individual-level) values of ORR and OS\*



\*For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. AML, acute myeloid leukaemia; CUP, cancer of unknown primary; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; SCLC, small cell lung cancer.

#### Treatment effect (trial-level) correlation and regression

The range of treatment effect (trial-level) R2 values reported in each meta-regression is summarised in Table 7 and illustrated in Figure 4 (for the association between ORR and PFS) and Figure 5 (for the association between ORR and OS). Each horizontal row in the plots illustrates the range of R2 values across all subgroup analyses within a single meta-regression study. Where an included meta-regression reported on more than one cancer type, these are shown separately on the plots. It is worth noting that the meta-regressions varied both in terms of the number of included primary studies (shown as N on the plots) and in terms of the treatment type, line of treatment and precise clinical population; all these details are provided in Appendix 6, together with R2 values for all individual subgroup analyses.

**ORR and PFS:** The regression R2 values for the treatment effect association between ORR and PFS ranged from 0.18 to 0.94, based on 9 studies across 4 cancer types: NSCLC,81, 82, 105, 137 ovarian cancer,87, 132 colorectal cancer86 and various solid tumours127, 139 (Figure 4 and Table 7; full details in Appendix 6). The majority of R2 values were above 0.40. The R2 values reported from multiple analyses within the same study, and those reported across separate studies, did not suggest a clear pattern by cancer type. Confidence intervals around the R2 values, where reported, were generally fairly wide (not shown in the plot, see Appendix 6).

**ORR and OS:** The regression R2 values for the treatment effect association between ORR and OS ranged from -0.08 to 0.84, based on 30 studies across 11 cancer types81-83, 85-92, 94, 96, 97, 100, 101, 105-107, 113, 116-118, 120, 123, 127, 133, 134, 137, 139 (Figure 5 and Table 7; full details in Appendix 6). With the exception of one analysis, all R2 values were below 0.60. The R2 values reported from multiple analyses within the same study, and those reported across separate studies, did not suggest a clear pattern by cancer type. Confidence intervals around the R2 values, where reported, were generally wide (not shown in the plot).

**CR and PFS or OS:** The regression R2 for the treatment effect association between CR and PFS ranged from 0.45 to 0.93 in one study of NHL,129 while the regression R2 for the treatment effect association between CR and OS within two studies of breast cancer96 and SCLC94 ranged from 0.05 to 0.48 (Table 7; full details in Appendix 6).

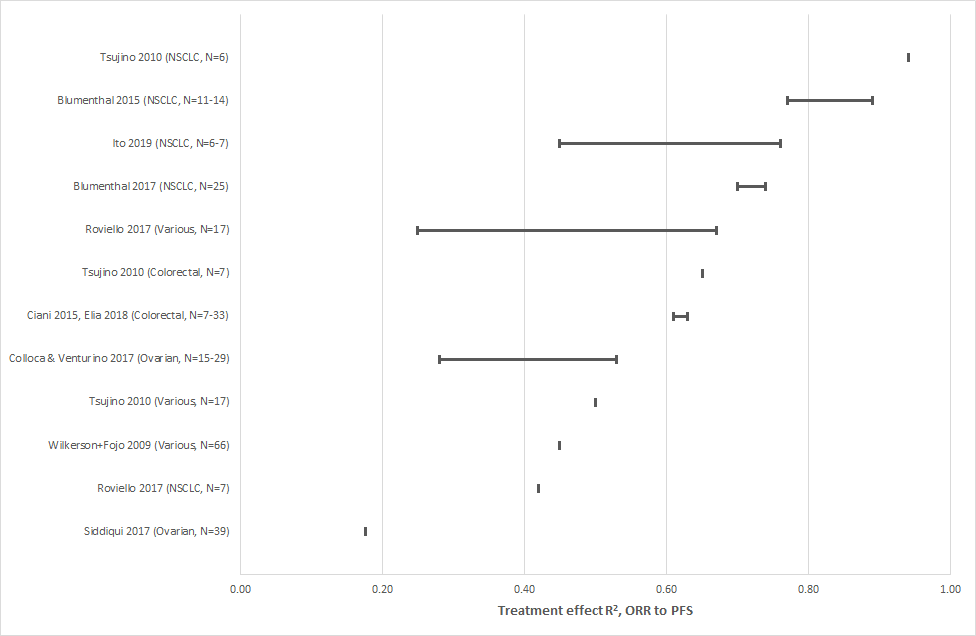
**PR and PFS or OS:** No studies reported the treatment effect association between PR and PFS or OS.

**DoR and PFS or OS:** No studies reported R2 between DoR and OS or PFS. Two studies in colorectal cancer89 and pancreatic cancer88 reported Spearman correlation coefficients between DoR and OS ranging from 0.40 to 0.76 (Table 7; full details in Appendix 6).

Table 7: Summary of treatment effect (trial-level) R2 per study

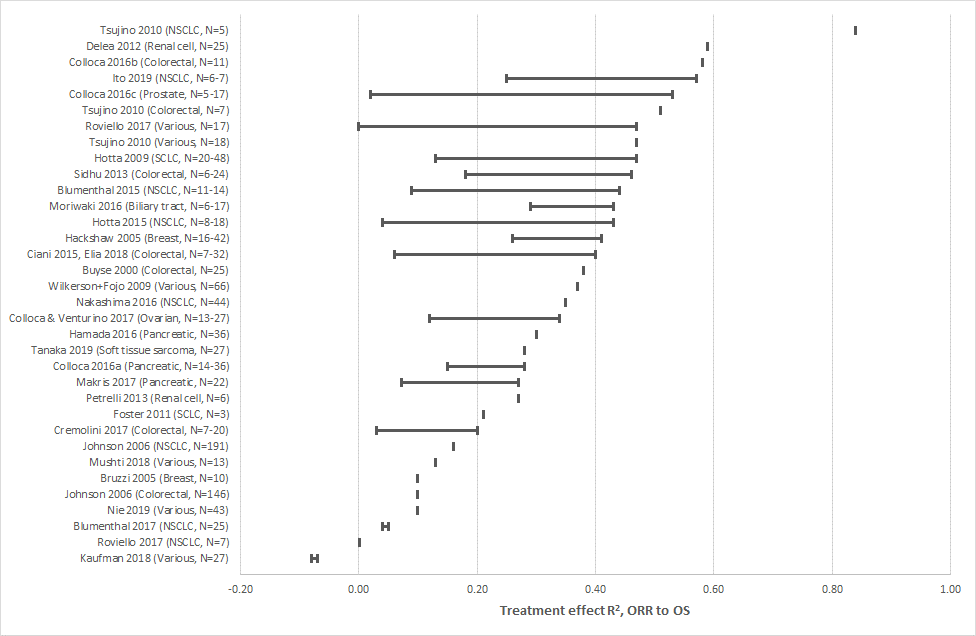
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Surrogate relationship** | **N studies** | **Cancer types and refs** | **Range of R2 across studies and subgroup analyses** | **Further detail** |
| **ORR to PFS** | 9 | NSCLC,81, 82, 105, 137 ovarian,87, 132 various,127, 139 CRC86 | 0.18 to 0.94 | Appendix 6  Figure 4 |
| **ORR to TTP** | 0 |  | - | - |
| **ORR to OS** | 30 | NSCLC,81, 82, 100, 105, 106, 118, 137 CRC,85, 86, 89, 91, 133 various,107, 117, 120, 127, 139 pancreatic,88, 97, 113 SCLC,94, 101 RCC,92, 123 breast,83, 96 ovarian,87 prostate,90 BTC,116 STC134 | -0.08 to 0.84 | Appendix 6  Figure 5 |
| **CR to PFS** | 1 | NHL129 | 0.45 to 0.93 | Appendix 6 |
| **CR to OS** | 2 | Breast,96 SCLC94 | 0.05 to 0.48 | Appendix 6 |
| **PR to PFS** | 0 |  | - | - |
| **PR to OS** | 0 |  | - | - |
| **DoR to PFS** | 0 |  | - | Appendix 6 |
| **DoR to OS** | 0 |  | (see footnote)\* | Appendix 6 |
| Notes: Further detail on all studies and outcomes is shown in Appendix 6. N studies per outcome may vary from Table 4 since not all data in correct format. \* Two studies in CRC89 and pancreatic cancer88 reported Spearman correlation coefficients between DoR and OS ranging from 0.40 to 0.76.  BTC, biliary tract cancer; CR, complete response; CRC, colorectal cancer; DoR, duration of response; NHL, non-Hodgkin’s lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; PR, partial response; RCC, renal cell carcinoma; SCLC, small cell lung cancer; STC, soft tissue sarcoma; TTP, time to progression; VGPR, very good partial response. | | | | | |

Figure 4: Regression R2 between treatment effects (trial-level) for ORR and PFS\*



\*For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. NSCLC, non-small cell lung cancer; ORR, overall response rate; PFS, progression-free survival.

Figure 5: Regression R2 between treatment effects (trial-level) for ORR and OS\*



\*For each study, the plot illustrates the range of correlation coefficients across all subgroup analyses. N represents the number of studies included in each meta-regression. NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; SCLC, small cell lung cancer.

#### Regression equations

##### Regression equations for absolute (individual-level) relationships

Regression equations for absolute (individual-level) associations were reported in six studies102, 112, 114, 132, 136, 141 and are summarised in Table 8.

**ORR to PFS/TTP:** For the relationship between ORR and median PFS/TTP, five studies across five cancer types102, 112, 114, 132, 136 reported regression equations (one study used log odds ORR),114 with intercepts ranging from 1.73 to 3.20 and slopes ranging from 0.07 to 0.41.

**ORR to OS:** For the relationship between ORR and median OS, four studies across four cancer types102, 112, 132, 136 reported regression equations, with intercepts ranging from 5.89 to 10.45 and slopes ranging from 0.08 to 0.28.

**CR to PFS:** For the relationship between CR and median PFS, two studies in NHL114, 141 reported regression equations (one study used log odds CR),114 with intercepts ranging from 0.83 to 2.38 and slopes ranging from 0.34 to 0.46.

Table 8: Regression equations for absolute associations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Surrogate relationship** | **Cancer types and references** | **Surrogate** | **Final** | **Intercept** | **Slope** |
| **ORR to PFS** | Colorectal112 | ORR | Median PFS | 3.20 | 0.10 |
| Lung (NSCLC)136 | ORR | Median PFS | NR | 0.07 |
| Ovarian132 | ORR | Median PFS | 2.59 | 0.12 |
| NHL114 | log odds ORR | log median PFS | 1.97 | 0.41 |
| **ORR to TTP** | Gastric102 | ORR | Median TTP | 1.73 | 0.09 |
| **ORR to OS** | Colorectal112 | ORR | Median OS | 10.45 | 0.09 |
| Lung (NSCLC)136 | ORR | Median OS | NR | 0.26 |
| Ovarian132 | ORR | Median OS | 9.48 | 0.28 |
| Gastric102 | ORR | Median OS | 5.89 | 0.08 |
| **CR to PFS** | NHL141 | CR | Median PFS | 0.83 | 0.46 |
| NHL114 | log odds CR | log median PFS | 2.38 | 0.34 |
| CR, complete response; NHL, non-Hodgkin’s lymphoma; NSCLC, non-small cell lung cancer; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; TTP, time to progression; VGPR, very good partial response. | | | | | |

##### Regression equations for treatment effect (trial-level) relationships

Regression equations for treatment effect (trial-level) associations were reported in thirteen studies84, 86, 91-93, 96, 101, 106, 116, 118, 127, 129, 137 and are summarised in Table 9. These are presented separately for regressions based on difference in response and regressions based on relative risk or OR for response. There was substantial variation in effect measures for both the surrogate and final outcomes (e.g. difference in medians, HR, OR).

**ORR to PFS:** For the relationship between ORR and PFS, one study of three cancer types137 reported regression equations for difference in ORR vs. HR for PFS, with slopes ranging from -0.02 to -0.04 (intercepts were not reported). Three studies across three cancer types84, 86, 93, 127 reported regression equations for difference in ORR vs. difference in median OS, with intercepts ranging from -0.05 to 0.34 and slopes ranging from 0.07 to 0.14. One study of three cancer types137 reported regression equations for difference in ORR vs. HR for OS, with slopes ranging from -0.01 to -0.03 (intercepts were not reported). Seven studies across six cancer types86, 91-93, 96, 116, 118, 127 reported regression equations for ratio measures of ORR (OR or relative risk) vs. ratio measures of OS (generally HR), with intercepts ranging from -0.13 to 0.12 and slopes ranging from -0.26 to 0.30. One study in SCLC101 reported a regression equation for relative risk of ORR vs. difference in median OS, with intercept of 0 and slopes ranging from 0.04 to 0.09.

**CR to PFS:** For the relationship between CR and PFS, one study in NHL129 reported regression equations for logOR CR vs. logHR PFS, with intercepts ranging from -0.09 to 0.04 and slopes ranging from -0.73 to -0.64.

**CR to OS:** For the relationship between CR and OS, one study in breast cancer96 reported regression equations for logOR CR vs. logHR PFS, with intercept -0.01 (where reported) and slopes ranging from 0.09 to 0.16.

Table 9: Regression equations for treatment effect (trial-level) associations

| **Surrogate relationship** | **Cancer types and refs** | **Subgroup** | **Based on difference in response** | | | | | **Based on relative risk or odds ratio for response** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Surrogate** | **Final** | **Intercept** | **Slope** | **Surrogate** | | **Final** | **Intercept** | **Slope** |
| **ORR to PFS** | Lung (NSCLC)137 |  | Diff ORR | HR PFS | NR | -0.02 |  | |  |  |  |
|  | Colorectal137 |  | Diff ORR | HR PFS | NR | -0.04 |  | |  |  |  |
|  | Various137 |  | Diff ORR | HR PFS | NR | -0.02 |  | |  |  |  |
|  | Colorectal86, 93 |  |  |  |  |  | logOR ORR | | logHR PFS | -0.05 | -0.32 |
|  | Breast84 |  |  |  |  |  | logOR ORR | | logHR PFS | 0.10 | 0.50 |
|  | Various (immuno)127 |  |  |  |  |  | logOR ORR | | logHR PFS | -0.13 | -0.24 |
| **ORR to OS** | Colorectal91 | - All  - Anti-angio  - Non-anti-angio | Diff ORR | Diff median OS | NR | 0.07 0.13 0.14 |  | |  |  |  |
|  | Colorectal106 |  | Diff ORR | Diff median OS | 0.34 | 0.10 |  | |  |  |  |
|  | Lung (NSCLC)106 |  | Diff ORR | Diff median OS | -0.05 | 0.09 |  | |  |  |  |
|  | Colorectal137 |  | Diff ORR | HR OS | NR | -0.03 |  | |  |  |  |
|  | Lung (NSCLC)137 |  | Diff ORR | HR OS | NR | -0.01 |  | |  |  |  |
|  | Various137 |  | Diff ORR | HR OS | NR | -0.02 |  | |  |  |  |
|  | Colorectal86, 93 | - All  - No crossover |  |  |  |  | logOR ORR | | logHR OS | -0.03  -0.04 | -0.05  -0.10 |
|  | Breast96 | - All  - Recr. pre-1990  - Recr. 1990 or after |  |  |  |  | logOR ORR | | logHR OS | -0.01  NR  NR | 0.28  0.28  0.24 |
|  | Lung (NSCLC)118 |  |  |  |  |  | lnOR ORR | | lnHR OS | -0.02 | -0.13 |
|  | Various (immuno)127 |  |  |  |  |  | logOR ORR | | logHR OS | -0.13 | -0.26 |
|  | Colorectal91 | - All  - Anti-angio  - Non-anti-angio |  |  |  |  | rr ORR | | HR OS | NR | -0.03  -0.11  -0.06 |
|  | Renal cell92 |  |  |  |  |  | ln rr ORR | | -lnHR OS | -0.11 | 0.30 |
|  | Biliary tract116 | - Chemo  - Gemcitabine  - Targeted |  |  |  |  | Ratio of ORR | | log ratio of median OS | 0.01  0.02  0.12 | 0.28  0.27  0.16 |
|  | Lung (SCLC)101 | - All  - Pub. 1990-1996  - Pub. 1997-2008 |  |  |  |  | rr ORR | | Diff median OS | 0.00  0.00  0.00 | 0.06  0.04  0.09 |
| **CR to PFS** | NHL129 |  |  |  |  |  | logOR CR 30mo | | logHR PFS | -0.09 | -0.64 |
|  | NHL129 |  |  |  |  |  | logOR CR 24mo | | logHR PFS | 0.04 | -0.73 |
| **CR to OS** | Breast96 | - All  - Recr. pre-1990  - Recr. 1990 or after |  |  |  |  | logOR CR | | logHR OS | -0.01  NR  NR | 0.13  0.09  0.16 |
| Anti-angio, anti-angiogenic; CR, complete response; diff, difference; HR, hazard ratio; ln, natural logarithm; log, logarithm; NHL, non-Hodgkin’s lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; pub, published; recr, recruited; rr, relative risk; SCLC, small cell lung cancer. | | | | | | | | | | | |

#### Surrogate threshold effect (STE)

The STE (the smallest treatment effect on the surrogate that predicts a non-zero treatment effect on the true endpoint)73 was reported in only four studies (Table 10).86, 99, 129, 137 For the relationship between ORR and PFS, one study in various solid tumours137 reported that a difference in ORR of 15% would be required to predict a non-zero treatment effect on the HR for PFS. For the relationship between ORR and OS, two studies in various solid tumours137 and NSCLC99 reported that a difference in ORR of 21% and 55% respectively would be required to predict a non-zero treatment effect on the HR for OS, while one study99 also reported that a difference in ORR of 41% would be required to predict a non-zero treatment effect on the difference in median OS, and a further study in colorectal cancer86 reported that an OR for ORR of 0.28 would be required to predict a non-zero treatment effect on the OR for OS. Finally, for the relationship between CR and PFS, one study in NHL129 reported that an OR for CR (at 30 months) of 1.56 would be required to predict a non-zero treatment effect on the HR for PFS.

Table 10: Surrogate threshold effect (STE)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Surrogate relationship** | **Cancer types and refs** | **Based on difference in response** | | | **Based on odds ratio for response** | | |
| **Surrogate** | **Final** | **STE** | **Surrogate** | **Final** | **STE** |
| **ORR to PFS** | Various137 | Diff ORR | HR PFS | 15% |  |  |  |
| **ORR to OS** | Colorectal86 |  |  |  | OR ORR | OR OS | 0.28 |
| NSCLC99 | Diff ORR  Diff ORR | HR OS  Diff median OS | 55%  41% |  |  |  |
| Various137 | Diff ORR | HR OS | 21% |  |  |  |
| **CR to PFS** | NHL129 |  |  |  | OR CR 30mo | HR PFS | 1.56 |
| CR, complete response; diff, difference; HR, hazard ratio; NHL, non-Hodgkin’s lymphoma; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate; OS, overall survival; PFS, progression-free survival; STE, surrogate threshold effect. | | | | | | | |

### IQWiG and BSES2 scores for strength of association

This section reports results from the IQWiG and BSES2 scoring for strength of association between surrogate and final endpoints. As described in Section 4.2.7, IQWiG scoring requires a correlation coefficient (r) for the treatment effect association, while BSES2 scoring requires R2 values for both the individual and treatment effect associations. IQWiG and BSES2 scores were calculated for all subgroup analyses with sufficient data; therefore, studies reporting more subgroups were more strongly represented in this analysis.

In terms of IQWiG scores (Figure 6), of 202 analyses (across 63 studies), 0 (0%) scored high, 15 (7%) scored medium+, 26 (13%) scored medium, 76 (38%) scored low and 85 (42%) were not evaluable.

In terms of BSES2 scores (Figure 7), of 202 analyses (across 63 studies), 0 (0%) scored excellent, 3 (1%) scored good, 3 (1%) scored fair, 7 (3%) scored poor and 189 (94%) were not evaluable.

Figure 6: Summary of IQWiG scores across all 202 analyses included in the review

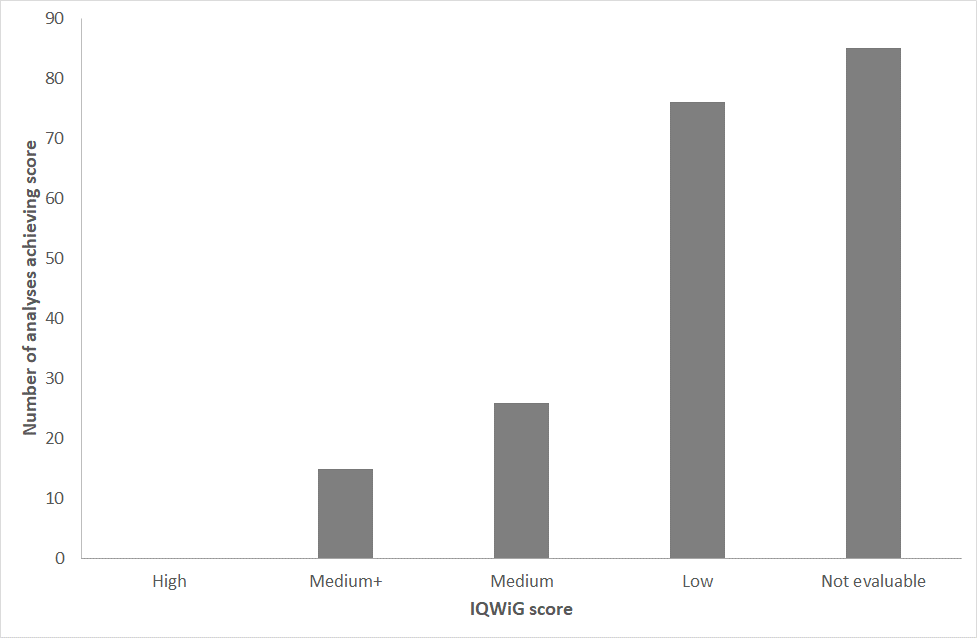
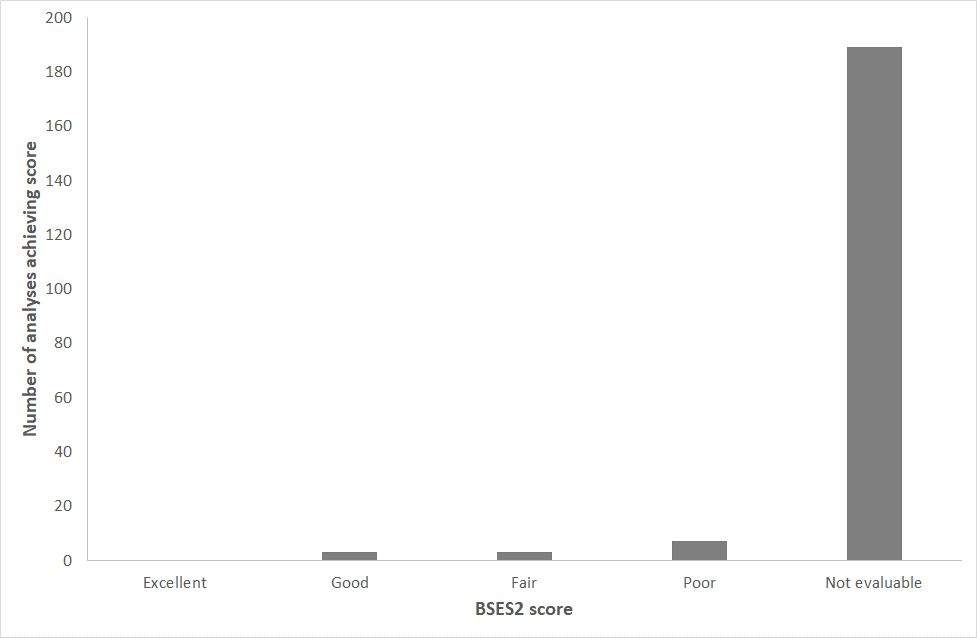


Figure 7: Summary of BSES2 scores across all 202 analyses included in the review



## Discussion

### Summary of main findings

#### Types of analysis identified

This systematic review summarises correlation and regression analyses for the strength of the association between response outcomes and PFS, TTP or OS across different types of cancer (primarily advanced or metastatic), based on included meta-analyses and meta-regression studies. In total, the review included 63 studies across 20 cancer types, most commonly NSCLC, CRC, breast cancer and analyses of various solid tumours. The most commonly analysed relationships were between ORR and either PFS or OS, with other response outcomes (such as CR, DoR and PR) reported in fewer analyses. The majority of studies (44) included only RCTs, while the remainder also included single-arm studies.

#### Absolute (individual-level) associations

For absolute (individual-level) association, the reported correlation coefficients between ORR and PFS ranged from -0.72 to 0.96 based on multiple analyses within 12 studies across 10 cancer types, while correlations between ORR and OS ranged from -0.40 to 1.00 based on 27 studies across 15 cancer types. Confidence intervals were generally fairly wide and were often not reported. The correlation coefficients reported from multiple analyses within the same study, and those reported across separate studies, did not suggest a clear pattern by cancer type. For analyses of CR, the correlation coefficients between CR and PFS in two studies ranged from 0.22 to 0.83, while those between CR and OS ranged from -0.04 to 0.62 based on 3 studies.

#### Treatment effect (trial-level) associations

For treatment effect (trial-level) association, the regression R2 between ORR and PFS ranged from 0.18 to 0.94 based on 9 studies across four cancer types, while the R2 values between ORR and OS ranged from -0.08 to 0.84 based on 30 studies across 11 cancer types. Again, there was no clear pattern between cancer types. For analyses of CR, the highest R2 between CR and PFS ranged from 0.45 to 0.93 in one study, while that between CR and OS ranged from 0.05 to 0.48 within two studies.

#### Regression equations and STE

Regression equations were reported in fourteen studies for the relationship between ORR and OS, and in eight studies for the relationship between ORR and PFS. There was substantial variation in effect measures for both the surrogate and final outcomes (e.g. difference in medians, HR, OR). The STE, the smallest treatment effect on the surrogate that predicts a non-zero treatment effect on the true endpoint,73 was reported in only four studies.

#### Strength of association between response and survival outcomes (IQWiG and BSES2 scoring)

Strength of association across all studies and all subgroup analyses was assessed using the IQWiG and BSES2 scoring systems. In general, scores were relatively low, indicating poor association between response and survival outcomes overall. Of 202 analyses using IQWiG scoring, 42% were not evaluable and 38% scored low, with 13% scoring medium, 7% medium+ and 0% high. When using BSES2 scores, the majority of analyses (94%) were not evaluable as they did not report R2 for both individual and treatment effect associations, with 3% scoring poor, 1% fair, 1% good and 0% excellent.

#### Strengths and limitations

In this review, a comprehensive search was undertaken to identify relevant studies. The reported data were highly heterogeneous in terms of effect measure and method of analysis. Therefore, some simplifying assumptions had to be made to allow the data to be summarised. Correlation coefficients were summarised regardless of method (Pearson, Spearman or other). R2 values were summarised irrespective of whether or not the regression was weighted and whether or not the R2 was adjusted. For treatment effect associations, R2 values were summarised regardless of effect measure (e.g. HR, OR, difference in medians, etc.).

#### Summary of findings

Based on this review, the association between response outcomes and PFS/TTP/OS varies widely between studies and generally scores low to medium on IQWiG and BSES2 scoring systems; however, a large number of analyses were not evaluable. There is no clear pattern for strength of association by cancer type. Previous reviews assessing multiple surrogate endpoints have also concluded that response-based endpoints were poor surrogates for OS.76, 77

## Implications for the economic analysis of histology-independent therapies based on ORR as a surrogate for PFS or OS

The review presented in this chapter provides information which could be used to inform judgements about whether response-based outcomes might be considered as a valid surrogate for PFS and OS. If the surrogate endpoint is considered valid, or potentially even if it is not, one may consider using that surrogate as the basis for estimating health gains within a health economic model. There are four main options relating to the use of response-based outcomes as a surrogate for OS or PFS within the economic analysis of histology-independent therapies.

###### 1. Use meta-analyses to predict the relationship between the surrogate and final outcome.

As shown in Table 8 and Table 9, fourteen studies report regression equations for ORR to OS and eight studies report equations for ORR to PFS. These equations could be used together with the observed ORR in the studies of histology-independent therapies in order to estimate absolute PFS/OS, or to estimate incremental gains in PFS/OS. However, the patient populations included in these studies may not correspond to the populations in the studies of histology-independent therapies in terms of tumour sites or types, and none specifically relate to patients with NTRK fusion-positive cancers (or other relevant biomarkers). From a practical point of view, a number of decisions would be required in order to apply these analyses within a model: (a) which regression equation to use in instances whereby multiple analyses exist for an individual histology site; (b) the form of regression analysis used to estimate the relationship (i.e. “absolute” regressions which estimate final outcomes for an individual treatment group, or “trial-level” equations which predict the treatment effects between groups), and; (c) how to model the surrogate relationship where no studies exist for an individual histology site. In addition, concerns regarding the strength of the relationship between ORR and PFS/OS within the tumour sites under consideration should be borne in mind.

It has been suggested that the stringent application of criteria for surrogate validation based on correlations may not be important, and that predictions may still be made even where the association is weak, provided they reflect all uncertainty surrounding the treatment effects.142 In addition, NICE TSD 20 notes that the meta-regression approaches included in this review are limited in that they ignore the uncertainty associated with the treatment effect on the surrogate endpoint (which is treated as a fixed covariate in the analysis), the consequence being that predictions based on these regression analyses will fail to fully reflect that uncertainty. Recently developed methods, such as the bivariate random effects meta-analysis (BRMA) model and its extensions,142, 143provide an approach for both the validation and prediction of surrogate endpoints within a Bayesian framework. In principle, this approach could be used to generate predictions of treatment effects on final outcomes in a way which allows for borrowing of information across studies and which fully accounts for all uncertainty surrounding the surrogate relationship. In instances whereby the surrogate association is weak, this would manifest as wider interval around the prediction and increased uncertainty surrounding modelled outcomes and costs. This approach is intuitively appealing; it would however render the published meta-regressions redundant as it would require re-analyses of the input data and the implementation of new meta-analyses for each histology site.

###### 2. Land-marking analysis

This review included only meta-analytic studies, and by design, excluded individual studies which did not include multiple cohorts of patients. Some of the studies which were excluded from the review during the sifting stage adopted a land-marking approach (see Section 7.5 for more details of this approach) within individual patient cohorts to explore the impact of response-based outcomes on OS, with differences between responders and non-responders reported in terms of an HR. Given an underlying baseline model of OS for non-responders, it may be possible to estimate the incremental impact on OS by combining the ORRs observed in the histology-independent studies with the HR derived from the land-marking analyses. However, the published land-marking studies generally related to a single tumour type and the study populations do not specifically relate to patients with NTRK fusion-positive tumours (or other relevant biomarkers).

###### 3. Risk-prediction models

During sifting, the review authors identified a small number of risk prediction studies. These studies reported multivariable statistical models to estimate the final outcome (OS/PFS) as a function of some response-based variable (e.g. ORR) together with other clinical parameters (e.g. age, sex, clinical characteristics). These studies may also provide a source of HRs for the impact of response on OS/PFS, but again these typically relate to a single tumour type and do not specifically relate to patients with NTRK fusion-positive tumours (or other relevant biomarkers).

###### 4. Do not use response as a surrogate for PFS/OS

The systematic review suggests that, taken generally, ORR may not be a reliable surrogate for PFS or OS based on current frameworks for surrogate validation. The review did not indicate any particular pattern whereby ORR performs better or worse according to tumour type or site. Even where there exists a means of predicting PFS/OS on the basis of ORR for a given tumour site (e.g. using conventional meta-regressions or BRMA), in the absence of a strong relationship between the surrogate and final endpoints, the resulting estimates may be highly uncertain and difficult to interpret. It should be noted however that the alternative may involve extrapolating highly immature PFS and OS data which are also subject to substantial uncertainty, hence this may not represent a sufficiently robust solution either.

## Conclusions

This systematic review suggests that response endpoints such as ORR and CR may not be reliable surrogates for PFS or OS. Strength of association varied widely between studies and subgroups, and in general, there was no clear pattern by cancer type.

Despite the potentially weak validity of response as a surrogate for PFS and OS, it may still be considered preferable to adopt a surrogate-based modelling approach informed by predictions from meta-analyses which capture all relevant uncertainty, than to ignore potential surrogate relationships and extrapolate heavily censored PFS and OS data. The recently developed BRMA approach outlined in DSU TSD 20142 may serve an important role in ensuring that all uncertainty around the surrogate relationship is reflected in the predictions used in the model. Ultimately, the most appropriate modelling approach will depend on the characteristics of the evidence available from the histology-independent study.

# A targeted review of published NICE technology appraisals where initial marketing authorisation was based on response outcomes from single-arm studies

We undertook a targeted review of 10 published NICE technology appraisals (TAs) where marketing authorisation was based on response rates from single-arm studies. The aim of the review was to highlight alternative analytic and structural approaches which have been proposed in previous appraisals to inform the extrapolation of surrogate endpoints based on ORR and DoR and/or to handle uncertainties due to immaturity in PFS and OS data. The case studies also served to identify a broader range of issues that are likely to be relevant for the appraisal of histology-independent products.

A thematic-based review is used to summarise key issues and uncertainties raised by the evidence review groups (ERGs) and NICE committees. The review is presented in Appendix 8.

## Summary and implications

The challenges of using a partitioned survival approach and relying on independent extrapolations of PFS and OS based on immature data are particularly evident in those appraisals where median OS had not been reached. In these specific appraisals, a range of alternative approaches were used including conventional parametric extrapolation approaches, the use of expert judgement and evidence from a proxy population with more mature evidence. In each of these appraisals, the committee highlighted significant concerns regarding the uncertainty and robustness of the ICER estimates, leading to recommendations within the CDF rather than routine NHS commissioning.

One important finding was that none of the 10 TAs explored the use of surrogate relationships to help inform the PFS and OS extrapolations. This could be considered surprising since the primary endpoint in the underpinning studies is a surrogate endpoint for clinical benefit and concerns noted by EMA and FDA regarding the challenges of interpretation and potential bias in assessing time to event endpoints based on single arm studies using ORR as the primary endpoint. However, it might also reflect the concerns regarding the reliability of ORR and CR as surrogates for PFS or OS (see Chapter 4).

Due to the nature of basket trials, significant heterogeneity may be present in the study populations enrolled in the trials (see Chapter 3). The potential importance of accounting for heterogeneity and exploring the cost-effectiveness in subgroups of the target population is acknowledged in the current NICE methods guide. Differences in the cost-effectiveness and decision uncertainty across these separate subgroups may lead to an optimised recommendation that is more restrictive than the marketing authorisation.

The review also demonstrated that heterogeneity within an overall target population is often a critical aspect of the appraisal. The committee acknowledged the importance of accounting for heterogeneity in a variety of sources in addition to relative effectiveness, including prognosis, health-related quality of life (HRQoL) and cost of comparator therapies which were likely to differ, impacting the cost-effectiveness estimates. The majority of TAs included only a small number of subgroups, most commonly based on alternative positions of a new treatment in an existing pathway. It is notable that in most of these appraisals, either separate studies were available for different subgroups or it was more feasible to undertake subgroup analyses than in histology-independent appraisals given the larger sample sizes. Although examples were identified which appeared more relevant to histology-independent appraisals, these were also limited to relatively small numbers of subgroups informed by separate studies or with sufficient numbers to present stratified results. However, it was evident from the appraisals of interventions with a broad marketing authorisation that the committee preferred to be explicit about the different sources of heterogeneity leading to specific recommendations for subgroups within the broader population.

Committees have routinely considered the diagnostic accuracy of available testing and the appropriateness of proposed testing strategies. The feasibility of introducing new testing pathways was also the subject of committee discussions. The predictive validity of target genetic mutations was well established, with company submissions providing an overview of the clinical basis for the predictive validity of target mutation. The prognostic validity of target mutations was in contrast poorly understood in all three appraisals reviewed, meaning only limited conclusions could be drawn regarding the prognosis of patients when receiving standard care.

Although the review of TA appraisals identified a number of important themes that are likely to be relevant to histology-independent appraisals, there are also important differences due to the nature of the study designs and greater levels of heterogeneity within the target population. Chapter 6 provides a more detailed consideration of some of the potential challenges that are envisaged and considers a range of alternative analytic approaches that might be required.

# Issues and challenges for exploring heterogeneity for histology-independent appraisals

The clinical and cost-effectiveness of a treatment will often depend on the characteristics of patients and the circumstances under which they receive treatment. The fact that different patient groups have different characteristics and so will derive different benefit from treatments is called ‘heterogeneity’.144, 145 146

Heterogeneity matters for two main reasons: first, if benefits differ by patient characteristics then estimates of treatment benefit must match the patient population that is expected to receive the treatment (the target population) in routine clinical practice. The second reason is that there can be health benefits from making tailored decisions for particular groups of patients. This gain from recognising differences between subgroups of patients and potentially ‘optimising’ recommendations within a product’s licence is called the value of heterogeneity (VoH).145-147

Exploration of sources of heterogeneity and the use of subgroup analysis is recommended within the NICE reference case analysis.4 Ignoring these differences could mean that a treatment which is not cost-effective for the total population (combining all subgroups) may be cost-effective in specific subgroups. Making a ‘one size fits all’ recommendation would then result in a potentially cost-effective treatment being withheld from a subset of patients for whom the treatment would represent an appropriate use of NHS resources. Conversely, a treatment which appears cost-effective for the total population may not be cost-effective in particular subgroups. In this case a ‘one size fits all’ approach could result in the treatment being recommended in identifiable subgroups in which the value of providing the new treatment is lower than the opportunity cost. That is, the health gain for these specific subgroups is not sufficient to offset the potential health lost from a reduction in the provision of services elsewhere in the NHS that is necessary to fund the new treatment.

In the case of histology-independent treatments, heterogeneity is particularly important to consider. This is because an important source of heterogeneity is differences in tumour histology. Though a treatment may be clinically effective across a range of tumour sites, there are theoretical and empirical reasons to expect that cost and health consequences could vary significantly across tumour sites. This is in addition to the usual sources of heterogeneity (e.g. age, gender etc.) which are present in conventional treatments.

There are a number of sources of heterogeneity which are relevant to histology-independent decision-making. The main focus of this chapter is heterogeneity between subgroups as defined by histology. However, it must be stressed that heterogeneity due to other characteristics are also relevant.

The following sections identify a number of particular challenges for histology-independent appraisals and present alternative approaches which might be used to investigate and account for different sources of heterogeneity. A formal framework is presented in Chapter 7.

## Treatment effectiveness

The available evidence is likely to consist of response and immature PFS and OS data for patients with different tumours, included in one or more single-arm studies with a basket design. Methods to test for heterogeneity in response by tumour type, or other relevant characteristics (typically related to the target mutation) can be used during trial conduct and can inform stopping rules in an adaptive trial design framework (Section 3). If heterogeneity is explored within the trial and it is concluded that a pooled analysis is justified (i.e. the treatment effect is sufficiently homogeneous across the tumour types), these results can be re-evaluated during the appraisal process and a decision made on whether or not it is appropriate to accept the company’s proposal of a homogeneous treatment effect across the tumours. A decision made within the trial to discontinue recruitment to one or more baskets due to an unsuitable response, should caution against a completely histology-independent recommendation. In such cases there will need to be a case made for which tumour types can be considered to have sufficient evidence of effect for a recommendation and which should be excluded, given the trial evidence of insufficient response.

Regardless of how the trial was originally designed and analysed, if outcomes are available for each tumour type, some of the frameworks described in Chapter 3 can be useful to explore the potential for heterogeneity in effects across tumours. The adaptive phase can be ignored and the methods can be used to estimate mean outcomes for each histology, with appropriate uncertainty, as well as pooled posterior and predictive mean outcomes which account for the potential lack of uniformity of effect across tumours. The BHM44 is simple to implement and particularly suited to this framework, since it starts from the assumption that treatment effects are exchangeable (rather than identical) across tumours – a more reasonable assumption in the absence of evidence to the contrary – and produces estimates of the level of heterogeneity across tumours and of the pooled treatment effects for each tumour, which can be used to judge whether the assumption of homogeneity is reasonable. In addition, this model allows prediction of the effect in unrepresented tumour types as long as they can also be assumed to have exchangeable effects (i.e. drawn from the same distribution of effects) as the included tumour types.

The BHM works by assuming that for each tumour type *j*, the measures of effect are exchangeable and follow a Normal distribution

where *σ* is the standard deviation quantifying the between-tumour heterogeneity and *μ* is the pooled mean effect across all tumour types. Prior distributions must be selected for *μ* and *σ*, and are likely to have some influence on the posterior estimates,44, 50 particularly when a small number of tumour types and patients per tumour are included. A Uniform(0,5) prior distribution was found to be robust in a simulation study.50 Sensitivity of results to the prior distributions should be assessed. When the outcome is binary, e.g. response, represents the log-odds of response in tumour site *j* and the probability of response in each site, , is recovered as

The probabilities that the response rates for each tumour type are at least of a certain magnitude can also be calculated and heterogeneity in these probabilities can guide conclusions on the plausibility of a homogeneous response. Typically, a value of 30% is used to define a meaningful response, but any other value can be used, depending on context.

In the case where the tumour types included in the trial are not reflective of the entire licensed indication, the predictive distribution of effect (e.g. the probability response) in a new histology, , can be obtained as

This will reflect the full degree of uncertainty both due to the sample size and the observed heterogeneity in effects across observed tumour sites. The resulting distribution represents the predictive probability of response in a ‘new’, i.e. unrepresented tumour type.

Although in theory the BHM can be applied to dichotomous (e.g. tumour response) or time-to-event outcomes (e.g. PFS and OS),44 the assumption of exchangeability of the effects of treatment on survival outcomes across tumours outcomes is harder to justify than the equivalent assumption made for the effects of treatment on response. As noted in Section 3, a critical consideration in designing basket trials is the heterogeneity in survival prognosis across the different histologies, and this is the motivation for evaluating measures such as standardised response rates which reflect tumour shrinkage, rather than survival outcomes.27, 28 In addition, the nature of the survival data available, which tends to be immature and based on only a few patients per tumour type, will make estimation of a hierarchical model challenging, unless informative prior distributions are used on key parameters.

The model proposed by Leon-Novelo *et al*.51 can be used in the scenario where it is not expected that all subgroups will be *a priori* exchangeable (Section 3) and there is a particular tumour characteristic (e.g. prognosis or type of NTRK fusion) that defines exchangeability so that different categories can be pre-defined (e.g. poor, intermediate, good prognosis). If these *a priori* exchangeable categories can be pre-defined, then the approach is similar to the BHM and a prediction for unrepresented tumour types in each category can be made. However, this would no longer generate a truly histology-independent recommendation, as results might differ across tumours in different categories.

Hybrid exchangeable/non-exchangeable models,49, 51 where exchangeability is determined by the data, are less relevant as exploratory models for HTA as their results would be harder to interpret. This is because exchangeability cannot be assumed to apply across all tumour types and predictive distributions of effects can no longer be assumed to represent the expected effect in unrepresented tumour types, since by definition these are not necessarily fully exchangeable with the included tumour types. In addition, in an exchangeable/non-exchangeable scenario, a histology-independent recommendation would be hard to justify as the assumption would be that effects differ across tumour types, which could impact on effectiveness and cost-effectiveness estimates.

### Exploring heterogeneity in response – case study

To demonstrate the impact of allowing for heterogeneity in response and to explore the potential heterogeneity in effects across tumours, response data were analysed using a BHM framework.44

For the purpose of this analysis, the response data used was the published efficacy evidence available for the Trk-inhibitor, larotrectinib.3, 148 The results, presented as a *post hoc* pooling of 55 patients covering 12 tumour types from three non-randomised single-arm Phase I/II basket studies, including the number of patients and responses by tumour type can be seen in Table 11.

Table 11: FDA Results of ORR by Tumour Type

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tumour  ID | Tumour Type | Patients (n) | ORR | |
| Responders (n) | Observed response (%) |
| 1 | Soft tissue sarcoma | 11 | 10 | 91% |
| 2 | Salivary gland | 12 | 10 | 83% |
| 3 | IFS | 7 | 7 | 100% |
| 4 | Thyroid | 5 | 5 | 100% |
| 5 | Lung | 4 | 3 | 75% |
| 6 | Melanoma | 4 | 2 | 50% |
| 7 | Colon | 4 | 1 | 25% |
| 8 | GIST | 3 | 3 | 100% |
| 9 | Cholangiocarcinoma | 2 | 0 | 0% |
| 10 | Appendix | 1 | 0 | 0% |
| 11 | Breast | 1 | 0 | 0% |
| 12 | Pancreas | 1 | 0 | 0% |
|  | Total | 55 | 41 | 74.5% |
| IFS, infantile fibrosarcoma; GIST, gastrointestinal stromal tumours | | | | |

We can consider each of the tumour types as a ‘basket’ or group and analyse the response data using a BHM framework to explore the potential heterogeneity in effects across tumours.

#### Methods

For the response outcome, data available for each of the tumour types in the published literature are the number of responders, , out of the total number of patients, for tumour type *j*, which are assumed to follow a binomial likelihood

where is the probability of response for tumour type *j*, with , and is the total number of tumour types. The log-odds of response in tumour type *j*, , was modelled on the log-odds scale: . The BHM assumes that for each of the tumour types, the log-odds of response, , are exchangeable and follow a Normal distribution (equation 1).

We used a relatively conservative normal prior distribution for *μ*, centred around a probability of response of 0.3 (a log-odds of -0.8473) which is often considered as a promising response rate, with a variance of 10 across all tumour types. Sensitivity of the results to a more favourable prior distribution where the prior probability of response across all tumour types is centred around a mean of 0.5 (a log-odds of 0) with the same variance was assessed.

The prior for the between-tumour heterogeneity standard deviation is specified as Uniform(0,5) which was found to be robust in a simulation study.44, 50 An Inverse Gamma(2, 20) prior distribution for the between-tumour variance had previously been proposed,44 meaning the between-tumour precision has prior mean 0.10 and variance 0.005. Inverse-gamma prior distributions were found to lead to posterior distributions which are highly sensitive to the chosen parameters and are therefore not recommended in most cases.50 The sensitivity of the results to the inverse-gamma prior distribution, the between-tumour heterogeneity variance, and to using different half-normal prior distributions for the between-study standard deviation was assessed.50 Half-Normal prior distributions with precision from 0.01 to 0.1 and 1 were also assessed.

As the tumour types included in the analysis population are not reflective of the full licensed indication (i.e. a truly histology-independent marketing authorisation will encompass all tumour types, not just those represented in the trial), the predictive distribution for the response rate in a new tumour type is calculated to reflect the full degree of uncertainty both due to the sample size and the observed heterogeneity in effects across the observed tumours. The resulting distribution is the probability of response in a ‘new’, i.e. unrepresented tumour type.

The model was adapted from Thall et al44 and estimated using Markov chain Monte Carlo in OpenBUGS,149 implemented in R150 (version 3.6.0) using R2OpenBUGS151 (version 3.2.3.2). The BUGS code used is presented in Appendix 9.

Model fit was assessed by plotting individual tumour 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 tumour types, . Convergence was assessed by visual inspection of the Brooks-Gelman-Rubin plots and assessment of the statistic.152, 153

#### Results

For all analyses 55,000 iterations were run on two parallel chains and the first 5,000 iterations discarded as ‘burn-in’. Model fit statistics are presented in Appendix 10.

The prior distributions used for the base-case analysis are

The BHM estimates substantial between-group heterogeneity, posterior median 2.86 on the log-odds scale, although there is considerable uncertainty (95% credible interval (CrI) of 0.92 to 4.83) (Figure 8). This suggests there is considerable variability across tumour types.

Figure 8: Prior and posterior distributions for the between-group heterogeneity standard deviations

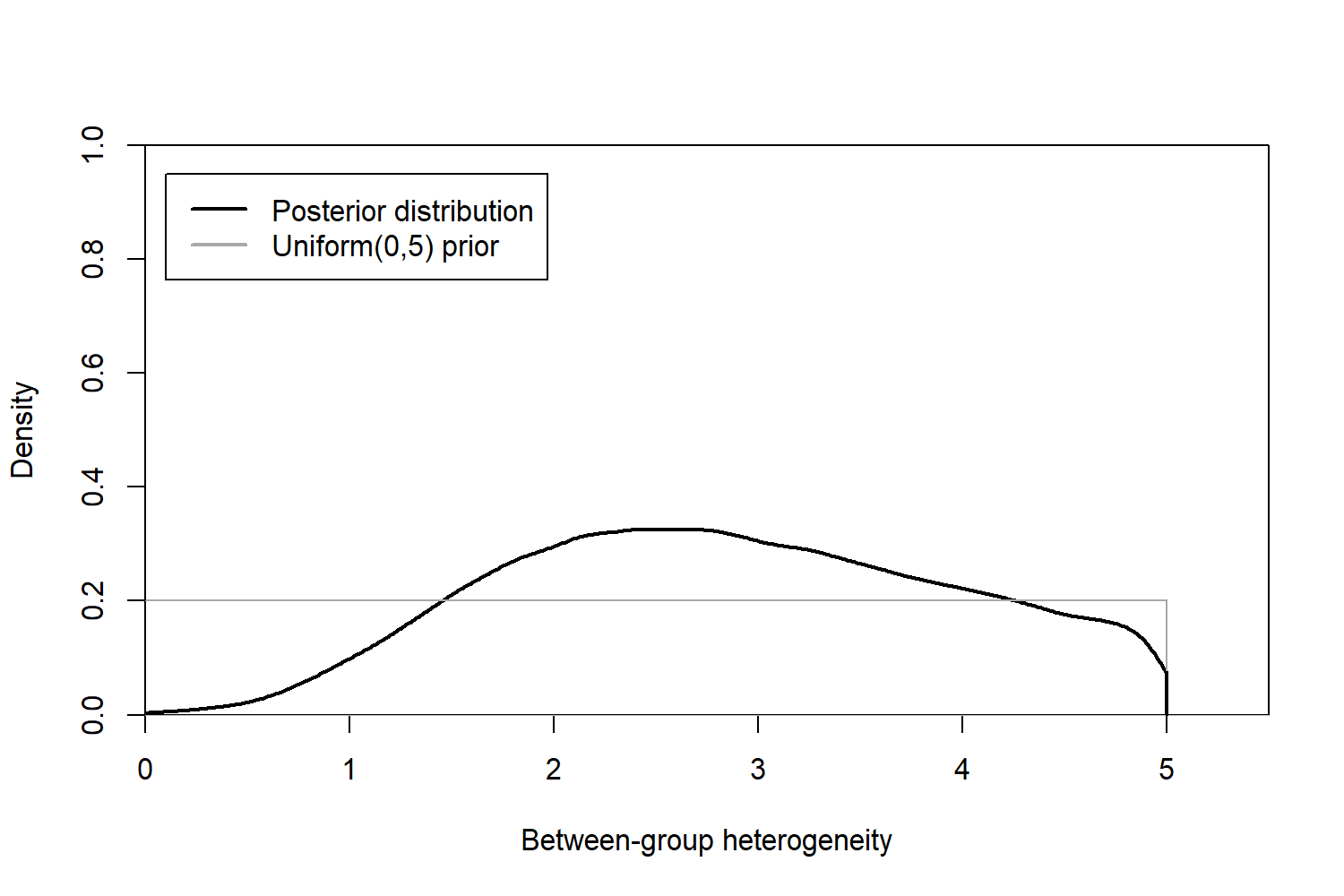
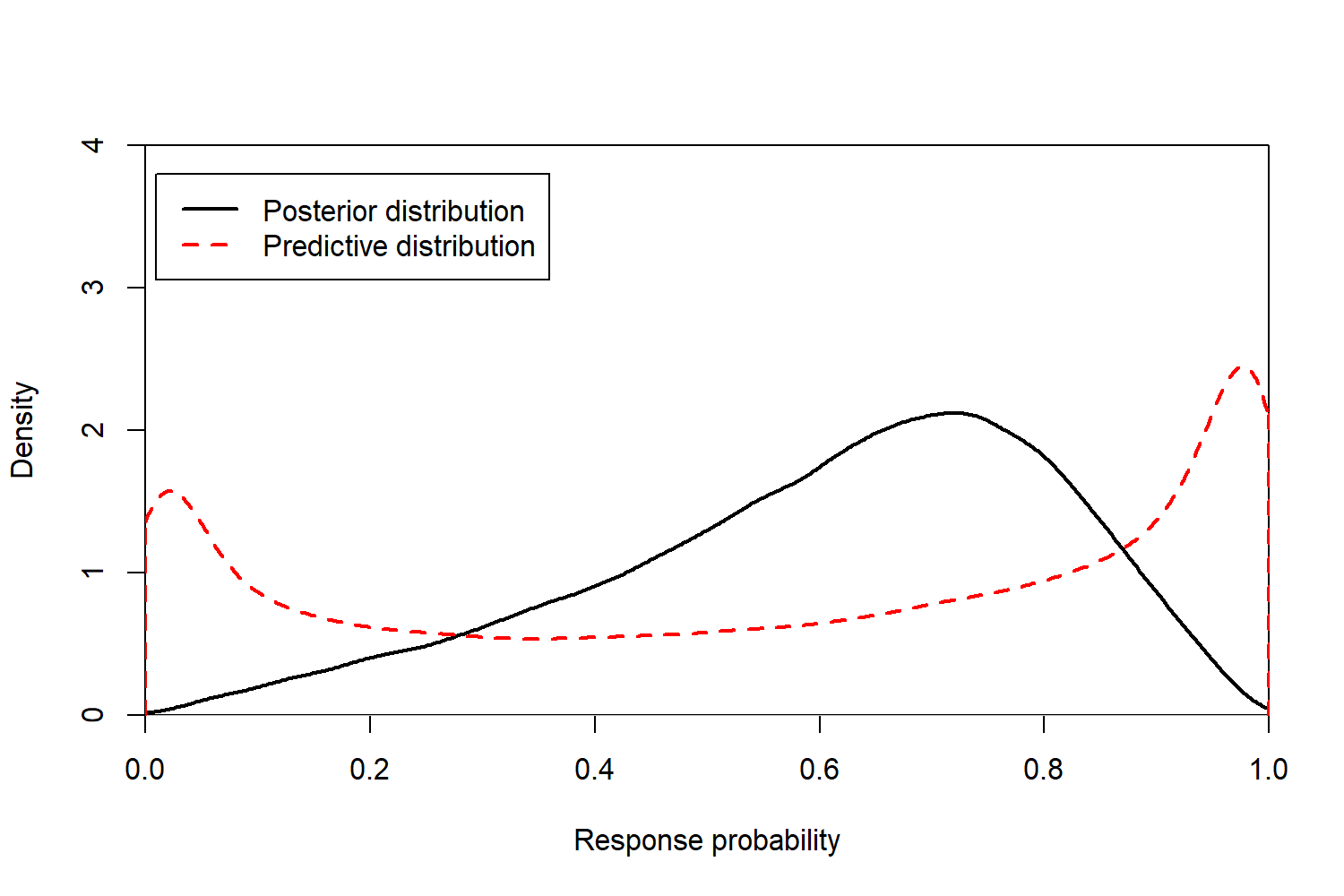


Table 12: Overall posterior probability of response

|  |  |  |  |
| --- | --- | --- | --- |
|  | Overall posterior probability of response | | |
|  | mean | median | 95% CrI |
| Posterior probability of response | 0.609 | 0.641 | (0.160,0.918) |
| Predictive probability of response | 0.569 | 0.649 | (0.002,0.999) |

The estimated mean response rate across all tumour types is 0.609 with 95% CrI (0.160, 0.918). This is lower than the mean response rate of 0.745 observed in the efficacy evaluable data set. The response probability predicted for an unrepresented tumour type is 0.569; however, the 95% CrI is wide meaning this probability could be as low as 0.2% or as high as 99.9% (Table 12, Figure 9).

Figure 9: Posterior and predictive distributions of response probability



The estimated probabilities of response for each tumour type are shown in Table 13. The effect of allowing borrowing of information across the tumour types is to shrink the observed response probabilities towards the pooled mean response probability. Tumour types with few patients borrow more information than tumour types with more patients and therefore have values closer to the pooled mean.

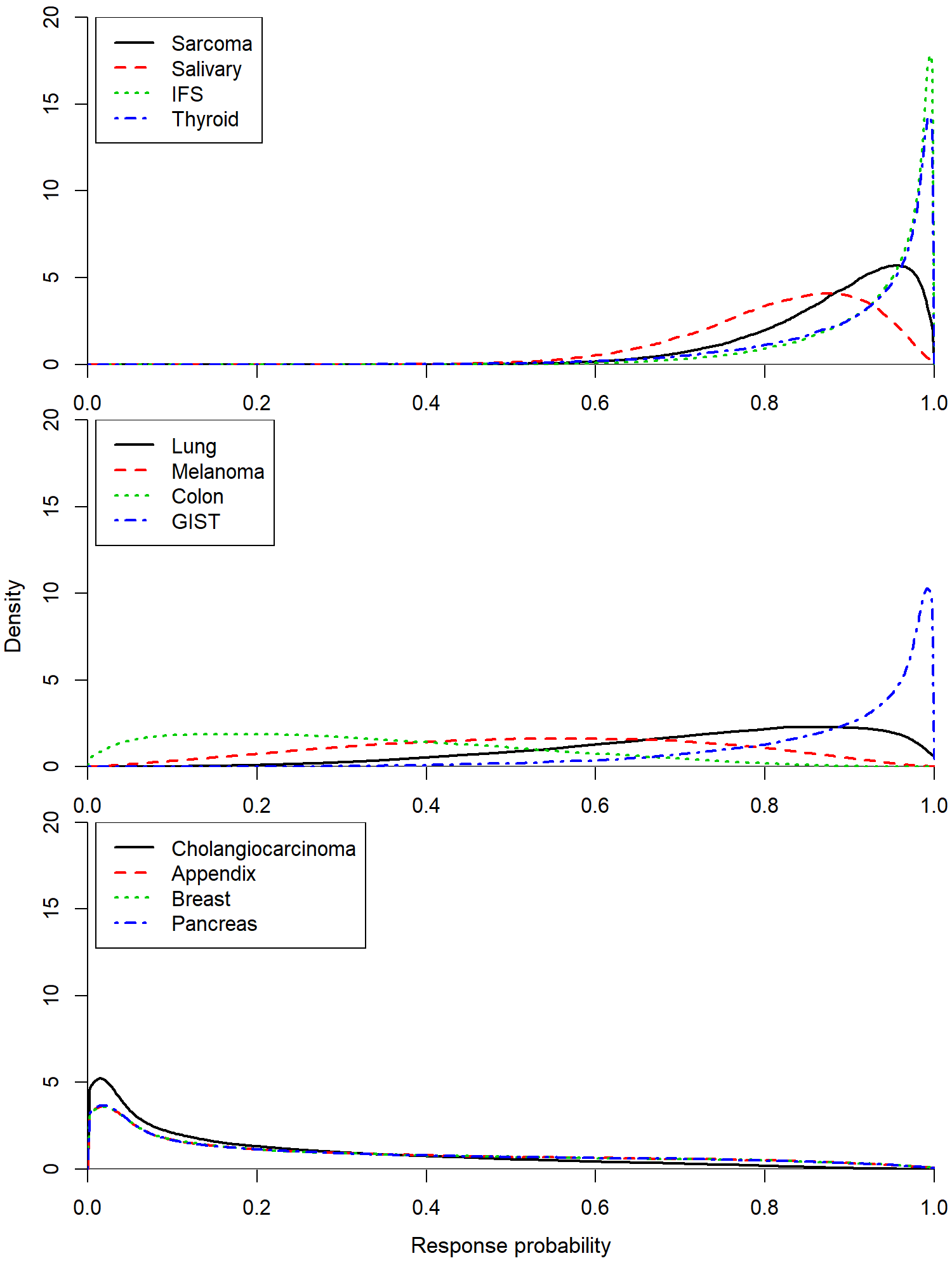
Table 13: Probabilities of response for all tumour types

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tumour Type | Observed Response | Estimated mean response based on BHM (%) | 95% CrI | Probability of response rate at least 30% | Probability of response rate at least 10% |
| Sarcoma | 91% | 88% | 66 - 99% | 1.000 | 1.000 |
| Salivary | 83% | 82% | 58 - 97% | 1.000 | 1.000 |
| IFS | 100% | 93% | 70 - 100% | 1.000 | 1.000 |
| Thyroid | 100% | 92% | 63 - 100% | 1.000 | 1.000 |
| Lung | 75% | 73% | 30 - 98% | 0.976 | 0.999 |
| Melanoma | 50% | 52% | 12 - 89% | 0.835 | 0.984 |
| Colon | 25% | 32% | 3 - 75% | 0.484 | 0.854 |
| GIST | 100% | 88% | 49 - 100% | 0.996 | 1.000 |
| Cholangiocarcinoma | 0% | 21% | 0 - 76% | 0.281 | 0.555 |
| Appendix | 0% | 30% | 0 - 90% | 0.416 | 0.650 |
| Breast | 0% | 30% | 0 - 90% | 0.415 | 0.653 |
| Pancreas | 0% | 30% | 0 - 90% | 0.413 | 0.648 |
| Prior distribution for log-odds of response centred on a probability of 0.3; Uniform prior distribution for the between-tumour standard deviation.  IFS, infantile fibrosarcoma; GIST, gastrointestinal stromal tumour. | | | | | |

Figure 10 shows the posterior distributions of the probabilities of response for each of the 12 tumour types included in the efficacy evaluable data set. Although the observed response suggested cholangiocarcinoma, appendix, breast and pancreas did not respond to larotrectinib, the posterior distributions of these tumour types are wide and their 95% CrI suggest response rates of 76% are plausible.

The results were insensitive to the use of the inverse gamma prior, the half-normal prior and the uniform prior centred on a log-odds of response of 0.5. The results were also insensitive to the use of a more favourable precision of the between-tumour heterogeneity standard deviation of 0.1 and 0.5. For full results of the sensitivity analysis, see Appendix 11.

Figure 10: Posterior distribution for the probabilities of response by tumours type



IFS, infantile fibrosarcoma; GIST, gastrointestinal stromal tumour.

### Implications for the appraisal of histology-independent technologies

Heterogeneity in the treatment effects are likely to be an important issue in the appraisal of histology-independent technologies. As can be seen from the results of the worked example, the BHM suggests there is substantial heterogeneity in response across tumour types. This can be seen in the estimate of the between-group heterogeneity, with the BHM estimating a posterior median standard deviation for the heterogeneity of 2.86 on the log-odds scale, which is considered large. Heterogeneity can also be seen in the predictive distribution of response, appearing on visual inspection of the posterior response in Figure 9 as a bimodal distribution with density concentrated around a probability of response of 0 and 1. This can be explained by the individual tumour response rates shown in Figure 10, which suggest that even under the assumption of exchangeability of response, there are tumour types in which the data suggest that response is likely and those in which it is not.

The results of this analysis challenge the strong assumption of homogeneity in response across such a variety of tumour types when treated with larotrectinib. The assumption of homogeneity may mask important information about empirical evidence of tumour response and the BHM provides a vehicle through which to account for the potential heterogeneity.

## Counterfactual

The feasibility of conducting RCTs in histology-independent populations is likely to be challenging and it is expected that histology-independent technologies will often seek (and receive) EMA/FDA approval with limited or no data from randomised experiments, see Sections 2.3.2 and 2.3.3. The lack of control data means that the evaluation of the cost-effectiveness of histology-independent technologies will require the generation of a comparator arm, for example, by generating a control based on a historical control.

The interpretation of relative effect estimates from single-arm studies compared with historical controls is potentially subject to bias, due to differences between patients selected as historical controls and those recruited to the single-arm studies. Differences between the patient populations can arise for a variety of reasons including differences among accrual sites or in patient characteristics (e.g., age, performance status, or other prognostic factors). For example, more recently diagnosed patients may have milder manifestations of a condition due to improved (and therefore commonly increased) diagnostic sensitivity. Treatment effect differences may also be attributable to secular trends in clinical care (e.g., changes in diagnostic methods, classification criteria, or outcome ascertainment) or other unknown confounders.

The challenges of generating an appropriate historical control are present in many appraisals wherever such comparisons are made, but may be particularly acute when considering a histology-independent technology. In the context of a histology-independent appraisal, the generation of an appropriate historical control data set is complicated by the need to cover multiple tumour types/histologies, which not only creates challenges for generating an appropriate comparator data set, but also potentially exacerbates the potential for confounding bias.

The need to cover multiple tumour types/histologies means that it is unlikely that any single data set will provide sufficient coverage to represent the whole target population. It is therefore likely that multiple data sources will need to be identified, as was implemented in the ongoing NICE appraisal of larotrectinib.154 Identification of historical control data would ideally be undertaken through an appropriate systematic review; however, this creates practical challenges as the resource required to implement this across multiple tumour types/histologies is extensive. In the larotrectinib appraisal, the focus of company searches for historical control data was limited to previous NICE technology appraisals covering the tumour types included in the company’s single-arm study. While this can be considered a reasonable pragmatic step, there is potential for alternative, and plausibly more relevant, sources of historical data to be missed.

Other challenges that result from the need to generate a data set that covers such a broad variety of tumour types, include the possibility that no relevant data exist for some relevant tumour types. For example, NTRK fusions are present in a number of rare tumour types which have not been subject to NICE guidance and in the ongoing larotrectinib appraisal,154 the company submission was forced to make arbitrary assumptions regarding the outcomes of patients for whom relevant comparator data could not be identified.

Identification of appropriate historical data will also need to address uncertainties regarding the positioning of therapy and any discrepancy between the licensed indication and the trials. Line of therapy may be an important prognostic factor, as patients in later lines of therapy will tend to have fewer treatment options and have potentially also accrued chemotherapy-related toxicity limiting their tolerability to further treatment. Attempting to match control patients characteristics to observed line of therapy in the intervention arm, however creates difficulties, as there is a tension as to whether to attempt to generate a comparator data set which is internally valid, i.e. where lines of therapy in the historical comparator data set match those of the intervention arm; or whether the historical control should attempt to reflect the eligible population and therefore maintain external validity, in which case the relative effect estimates may be sbiased. Indeed, this tension between internal and external validity may extend to other patient characteristics, particularly where the pool of patients in the intervention arm for a particular tumour is small, as recruited patients may not be fully representative of the eligible population. This tension therefore may typify a general issue of whether to match patient characteristics in the control arm to those in the intervention arm.

A further issue with using historical controls is that the target mutation may be prognostic in some or all tumours and it may be difficult to obtain relevant historical data limited to patients that harbour the target mutation, particularly where this mutation is rare. There is also the possibility that the prognostic value of the mutation may differ across tumour sites further complicating any attempt to adjust for the prognostic value of a mutation. In addition, in the context of a new target mutation, the prognostic value in different tumour types may not have been investigated sufficiently and is likely to be unknown for most, if not all, tumour types. For example, there is evidence to suggest an association between the presence of an NTRK fusion and unfavourable disease presentation155, 156 and better prognosis in patients with congenital mesoblastic nephroma (CMN), who harbour an NTRK fusion compared to those without the genetic abnormality.157 The evidence across tumour types is limited but the prognosis of patients with NTRK fusions may vary between cancer types and between NTRK fusion types.157 From the evidence available, it is also unclear whether NTRK fusions are in themselves prognostic, or whether it is their association with other specific prognostic factors, such as age and Eastern Cooperative Oncology Group (ECOG) performance status, that drives the observed differences in prognosis.

### Adjustment for confounding bias

A key factor in the reliability of estimates of effectiveness based on observational data is the statistical analysis used; a large number of studies have sought to develop and evaluate methods for adjusting and eliminating bias resulting from confounding. These include methods such as regression analysis, propensity scoring and population-adjusted indirect comparisons (MAIC and STC).158 These and other methods are frequently used in the literature and have been previously applied and accepted by NICE appraisal committees where no randomised evidence exists. In theory, these methods could be applied in the context of a histology-independent appraisal. Implementing such approaches could, however, be challenging because of the large number of source data sets involved, meaning population characteristics may not be reported across all comparator data sources and would necessarily require strong assumptions about the prognostic value of population characteristics across tumour types. Furthermore, even if a suitable adjusted comparison could be generated, the small sample sizes typically seen in the Phase II trials would only be able to account for a small number of observed characteristics. This limits the potential for these methods to fully account for confounding biases and increases the likelihood of residual confounding bias. Despite these limitations, such methods would generally be considered to be preferable to a naïve comparison which takes no account of differences across groups.

Gaps in the reporting of baseline characteristics, variability in the prognostic value of characteristics across tumour types/histologies and difficulties of matching comparator data to the likely limited available Phase II trial for histology-independent technologies will also create additional challenges of interpretation and validation of comparisons with historical data as it will be challenging to assess the comparability of patients in the historical control with those in the available single-arm trial data.

### Alternative approaches to developing a comparator

Because of these significant concerns about confounding bias and the challenges of generating a truly comparable comparator data set, other approaches to generating a comparator data set should be considered and their limitations explored. For example, two alternative methods outlined in Hatswell *et al*159 could be used in which patients in the single arm trial are used to generate a control group*.*

The first approach proposed by Hatswell *et al.*159 uses effectiveness data on non-responders as a proxy for patients not receiving an active treatment. Comparator effectiveness estimates of PFS and OS under this approach would therefore be based on observed PFS and OS amongst non-responders in the integrated efficacy analysis. The advantage of this approach is that all patients in the non-responder subgroup met the same trial inclusion and exclusion criteria, and are receiving the same line of treatment. The rationale behind this approach is that patients in whom no response is observed represent those with a lack of treatment effect (as they have no response to treatment) and therefore are representative of a counterfactual where no effective therapy exists. The patient population are therefore likely to be better matched with the intervention arm because they are drawn from the same population.

This approach, however, also requires strong assumptions, namely that there are no differences other than response status between responders and non-responders that explain the survival outcomes, and that non-responders derive equivalent benefit to that received on current standard of care. The reasonableness of these assumptions is likely to be specific to a particular appraisal, though as discussed in Chapter 4, the reliability of response as a surrogate is likely to be variable across tumour types. The assumption of no treatment benefit or harm may also not hold, as some patients may receive some benefit from treatment, even if they do not have a partial or complete response.

When considering the appropriateness of this approach, the relative advantages and disadvantages will need to be considered and it may be that this approach is only considered reasonable where there is substantive evidence of heterogeneity in treatment effects justifying the need to appropriately account for this heterogeneity in the economic analysis.

The second approach160 uses data taken from the trial patients’ previous line of treatment to derive OS and PFS curves. In this approach, the inverse of the ratio between average TTP on their previous therapy and the mean extrapolated PFS with the active therapy (also called the Growth Modulation Index (GMI) multiplier) is applied to all health outcomes (PFS and OS) for the active therapy. This crude adjustment assumes that the active therapy is more effective in terms of both PFS and OS than the comparator by the same proportion as the GMI multiplier. Therefore, the resulting GMI-adjusted total mean life years gained (LYG) and quality-adjusted life years (QALYs) are assumed to correspond to comparator outcomes and applied in the calculation of the ICER (based on LYG and QALYs). The main advantage is that effect estimates are drawn from the same population as the intervention arm and therefore better matched; however, there are also disadvantages. Firstly, this can only be implemented for patients who have received a previous line of therapy. Secondly, it also assumes that the ratio of TTP across lines of therapy is indicative of the treatment effect and it is uncertain to what degree this is likely to hold true. Finally, because this method can only estimate PFS, it requires that assumptions be made about the impact of TTP progression gains on OS. Namely that either OS increases proportionally with TTP or that PPS survival is the same across therapies which similarly may not hold true. Further research considering the reasonableness of these assumptions may be helpful. Consideration should also be given to the potential role of expert elicitation to inform these judgements.

### Implications for the appraisal of histology-independent technologies

The broad marketing authorisation, heterogeneous populations and uncertainties regarding the position of histology-independent technologies creates a number of significant challenges to creating an appropriate historical control data. Confidence in estimates of effect may increase by utilising methods of population-adjustment, but the scope of such methods may be more limited in the context of histology-independent appraisals. Assessment of the scope for residual confounding bias is also likely to be made more complicated, further reducing confidence in comparisons. It is unclear whether the use of non-randomised evidence and in particular single-arm studies will ever be considered adequate. Alternative methods of developing a comparator may therefore be of value to decision makers and should be considered as alternatives.

## Generalisability

The extent to which evidence is generalisable to the population of interest is a key consideration in the appraisal of histology-independent cancer technologies. There may be a number of uncertainties concerning the generalisability of available evidence, including the different types and distribution of histologies in the clinical studies (and the extent to which these represent the specific types and distribution of histologies that would be expected in routine clinical practice), the potential impact of unobserved histologies not represented in existing clinical studies and the position in the treatment pathway. Each of these issues is discussed in turn in the sections below.

### Distribution of tumour types

As outlined previously, the nature of the evidence likely to be available for decision making will involve a number of histologies or tumour types with limited data on each. When integrating this clinical data into a cost-effectiveness analysis one important issue is to consider the distribution of patients across the different tumour types.

One approach would be to utilise the distribution of tumour types present in the clinical evidence to generate an average cost-effectiveness estimate. Underlying this approach, however, is the assumption that cost-effectiveness of a histology-independent technology either does not vary across tumour types or that the proportions of histologies are representative of the proportions eligible to receive the intervention in the full licensed population. The former assumption is very unlikely to hold due to the potential for differences in effectiveness across tumour types, differences in prognosis, comparators, costs and HRQoL. Further, as can be seen from the comparison of the distributions in Table 14, there is a mismatch between the distributions of certain histologies in the trial populations. For instance, breast cancer patients represent 11% of the entrectinib trial but only 1.8% of the larotrectinib trial. If we imagine larotrectinib and entrectinib produce identical clinical results, then the resulting average cost-effectiveness estimates of the two distributions will be different given that different tumour types have different testing costs (see Section 6.4), different standard of care costs and outcomes and potentially different prognoses.

Table 14: Larotrectinib and Entrectinib trial tumour distributions

|  |  |  |  |
| --- | --- | --- | --- |
| **Larotrectinib Efficacy Evaluable Data Set (FDA)** | | **Entrectinib Efficacy Evaluable Data Set (FDA)** | |
| Histology | Trial proportion | Histology | Trial proportion |
| Soft tissue sarcoma | 20.00% | Sarcoma | 24.00% |
| Salivary gland | 21.80% | Salivary gland (MASC) | 13.00% |
| Thyroid | 9.10% | Thyroid | 9.00% |
| Lung | 7.30% | NSCLC | 18.00% |
| Colon | 7.30% | Colorectal | 7.00% |
| Cholangiocarcinoma | 3.60% | Cholangiocarcinoma | 2.00% |
| Breast | 1.80% | Breast | 11.00% |
| Pancreas | 1.80% | Pancreatic | 6.00% |
| IFS | 12.70% | Neuroendocrine | 6.00% |
| Melanoma | 7.30% | Gynaecological | 4.00% |
| GIST | 5.50% | - | - |
| Appendix | 1.80% | - | - |
| GIST, gastrointestinal stromal tumour; IFS, infantile fibrosarcoma; MASC, mammary analogue secretory carcinoma; NSCLC, non-small cell lung cancer | | | |

The significance of these differences for the overall assessment of cost-effectiveness will depend on the degree of heterogeneity across separate inputs relevant to economic modelling. If the trial distribution is not considered to represent the distribution expected to be seen in the population under a histology-independent license, any decision based on a single ICER estimate for the trial population will be subject to potential bias. The magnitude and direction of this bias will be difficult to determine without a more explicit assessment of heterogeneity in different sources relevant to the economic model.

Where differences between the trial population and licenced population are considered significant, approaches should be explored that allow for the re-weighting of the clinical population so that the model population better reflects the treated population.

### Unrepresented tumour types

A further issue to consider is whether the trial evidence encompasses all of the histologies covered by marketing authorisation. If histologies exist which are not represented in the trials but are covered under the marketing authorisation, decision-makers will only have evidence on effectiveness from the subset of the total population that is potentially eligible for the intervention.

For example, within the clinical evidence available for larotrectinib (Table 11), 12 histologies were included. However, it is known that upwards of a further 17 histologies have been shown to harbour NTRK fusions and will be covered by the anticipated marketing authorisation; hence, this total could be even higher.161 Any decisions made on the evidence alone, would therefore be implicitly assuming that the 12 included types are representative of the full population.

The impact of the unrepresented population is potentially significant and its importance will depend upon the number of unrepresented histologies and the proportion of the eligible population in unrepresented histologies relative to the observed histologies It is also important to consider unrepresented tumours where there is significant uncertainty regarding the homogeneity of clinical benefits or significant heterogeneity in costs across tumour types. For example, where there is limited support for the assumption of homogenous efficacy across histologies, it may be important to characterise the uncertainty in the efficacy within the unobserved population. Equally there is significant evidence of variability of testing costs across tumour types and therefore ignoring unrepresented tumour may impact significant on average testing costs.

### Position in treatment pathway

A further potential limitation regarding the generalisability of available clinical evidence relates to the position in the pathway in which patients are treated. This is complicated in part because the position in the pathway may vary substantively across tumour types according to the availability of alternative treatments, but also because of the potential for a mismatch between the trial population and the eligible population as dictated by the marketing authorisation. The latter may be a significant issue, as the recruitment of patients to a histology-independent trial is necessarily more complicated and there are potential significant challenges to identifying patients due to the relative rarity of target genetic mutations. Thus, as observed in the entrectinib and larotrectinib clinical data, patients were recruited across multiple lines of therapy even within the same tumour type.

This heterogeneity generates a number of issues, not least with respect to the external validity of the trial. Line of therapy may be a significant prognostic factor and failure to adjust for this may impact significantly on estimates of relative effectiveness, particular if the comparator population is not matched to the position patients were treated in the treatment arm. This issue may also impact in other ways, including upon the final distribution of patients eligible for treatment, because fewer patients will be eligible for treatment in second and subsequent lines of therapy. Further, it may have implications for testing, either impacting on the total costs of testing or the population that will be eligible for testing.

### Example

The example considers the Trk-inhibitor, larotrectinib, used for the treatment of solid tumours harbouring an NTRK gene fusion, and the proportion of tumour types presented in the clinical evidence, as outlined in Section 6.1.1.

First, in order to quantify the size of the population to benefit from Trk-inhibitors, the total number of patients eligible each year was calculated. This was estimated by using the tumour-specific NTRK fusion prevalence, cancer incidence and proportion of patients with advanced or metastatic disease.

The full method of calculating the annual eligible population is described in Appendix 12.

Table 15 presents the calculations of the annual eligible population for Trk-inhibitors, based on the tumour types represented in the larotrectinib trial. The prevalence of NTRK fusion varies across tumour types, ranging from 92.2% (MASC) to 0.07% (Breast Cancer). Paediatric patients with infantile fibrosarcoma, a tumour type with a high NTRK fusion prevalence make up the largest proportion of the eligible population (n = 27). Despite its low NTRK fusion prevalence, patients with colorectal cancer contribute a substantial proportion of the eligible population (n = 23).

Table 15: Calculation of the Annual Eligible Population for Trk-inhibitors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tumour Type** | **Prevalence of NTRK fusion** | **Cancer incidence (England)** | **% with stage III/IV cancer at diagnosis** | **Annual Trk-inhibitor eligible population** |
| Soft tissue sarcoma | 0.56% | 2740 | 32% | 5 |
| Appendix | 4.00% | 540 | 74% | 16 |
| Breast | 0.07% | 46102 | 15% | 5 |
| Cholangiocarcinoma | 0.10% | 556 | 60% | 0 |
| Colorectal | 0.12% | 34825 | 55% | 23 |
| IFS | 90.90% | 59 | 51% | 27 |
| MASC | 92.90% | 11 | 22% | 2 |
| Melanoma | 0.21% | 13740 | 9% | 3 |
| NSCLC | 0.09% | 32576 | 57% | 17 |
| Pancreatic | 0.26% | 8388 | 78% | 17 |
| Thyroid | 0.92% | 2195 | 31% | 6 |
| GIST | 1.28% | 734 | 40% | 4 |
| **Total** |  |  |  | **125** |
| IFS, infantile fibrosarcoma; MASC, mammary analogue secretory carcinoma; NSCLC, non-small cell lung cancer | | | | |

The resulting eligible population distribution (Table 16, column (2)) shows the proportions of tumour types in the eligible population differs substantially to the proportions in the larotrectinib trial (Table 16, column (1)). For example, soft tissue sarcoma represents 20% of the population in the larotrectinib trial yet only represents 4% of the population eligible to receive larotrectinib.

###### Unrepresented Tumour Types

In addition, NTRK fusions have been found in numerous tumour types not included in the larotrectinib trial. Following a histology-indepndent approval decision, patients with these tumour types will be eligible for treatment. In addition to the 12 tumour sites included in the larotrectinib trial, there is evidence of NTRK fusions in an additional 17 tumour types or anatomical sites.161 The annual eligible population making up the unrepresented tumour types for larotrectinib was again estimated using equation (4). The size of the unrepresented population was calculated to be 152 patients, 55% of the annual eligible population. The calculation of the size of the unrepresented population can be seen Appendix 11.

The eligible population including the unrepresented population is shown in Table 16. As can be seen from the comparison of the proportion of tumour types in eligible population with and without the inclusion of the unrepresented tumour types, the proportions differ. Individuals with soft tissue sarcoma represented 4% of the tumours in the eligible population and 1.8% when the unrepresented population is included.

###### Position in Treatment Pathway

If we assume the Trk-inhibitors will be given as a first-line therapy and 100% of patients will receive it in every tumour type, then the distribution of tumour types will be the real-world distribution. However, testing and position in the treatment pathway can impact this distribution.

The position in which genomic testing is offered to identify NTRK fusions will alter the annual population eligible for larotrectinib and other Trk-inhibitors. If testing were to be offered at the position in the treatment pathway where larotrectinib would be given, the annual eligible population will be lower than the population identified by upfront screening as some individuals who were NTRK fusion positive could have responded to alternative therapies, not be fit enough or have died before becoming eligible for Trk-inhibitor treatment (See Section 6.4).

As the position of larotrectinib is likely to differ between tumours, owing to the availability of other ‘satisfactory’ therapies, the overall distribution of individuals across tumour types in the eligible population is likely to change relative to the distribution assuming 100% of eligible patients receive the Trk-inhibitor as a first line therapy, i.e. Table 16, column (2).

To demonstrate the impact the position in the treatment pathway will have on the distribution of patients eligible for treatment with larotrectinib, an estimate of the likely position was obtained from the FDA review for larotrectinib.9 For the tumour types where there was no indication of where larotrectinib would be positioned, it was assumed that the drug would be offered as a third line therapy. Based on clinical advice, it was assumed that for the tumours where larotrectinib was offered first-line, 90% of eligible patients would be fit enough for treatment. It was assumed that, for those in which larotrectinib was offered second line and third line, 60% and 30% of the eligible population would be treated with larotrectinib, respectively.

As can be seen in Table 16, column (4), the final distribution when considering positioning in the treatment pathway again changes the distribution of tumour types. For example, infantile fibrosarcoma increased from 9.8% of the population to 22.4% of the population when accounting for position in the treatment pathway.

By comparing the trial distribution in Table 16, column (1) to the eligible population including the unrepresented tumour types and allowing for the position in the treatment in the pathway, we can see the considerable difference in the proportions of tumour types. As has been discussed previously, if the cost-effectiveness of a histology-independent technology is based on single estimates of costs and outcomes based on the average of tumour types present in the clinical evidence, the resulting uncertainty will be significant given incremental costs and outcomes are likely to differ substantially across tumour types.

Table 16: Alternative tumour distributions

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tumour Type** |  | **Trial population (n)** | **Trial proportions (1)** | **Eligible population (n)** | **Eligible population proportions (2)** | **Eligible population + unrepresented proportions (3)** | **% treated based on line of therapy** | **Treated population based on line of therapy** | **Eligible population including line of therapy** |
| **Represented** | | | | | | | | | |
| Soft tissue sarcoma |  | 11 | 20% | 5 | 4.00% | 1.80% | 90% | 4.5 | 4.20% |
| Appendix |  | 1 | 2% | 16 | 12.80% | 5.80% | 30% | 4.8 | 4.40% |
| Breast |  | 1 | 2% | 5 | 4.00% | 1.80% | 30% | 1.5 | 1.40% |
| Cholangiocarcinoma |  | 2 | 4% | 0 | 0.00% | 0.00% | 30% | 0 | 0.00% |
| Colorectal |  | 4 | 7% | 23 | 18.40% | 8.30% | 30% | 6.9 | 6.40% |
| GIST |  | 3 | 5% | 4 | 3.20% | 1.40% | 30% | 1.2 | 1.10% |
| IFS |  | 7 | 13% | 27 | 21.60% | 9.80% | 90% | 24.3 | 22.40% |
| MASC |  | 12 | 22% | 2 | 1.60% | 0.70% | 60% | 1.2 | 1.10% |
| Melanoma |  | 4 | 7% | 3 | 2.40% | 1.10% | 30% | 0.9 | 0.80% |
| Lung |  | 4 | 7% | 17 | 13.60% | 6.20% | 60% | 10.2 | 9.40% |
| Pancreatic |  | 1 | 2% | 17 | 13.60% | 6.20% | 30% | 5.1 | 4.70% |
| Thyroid |  | 5 | 9% | 6 | 4.80% | 2.20% | 30% | 1.8 | 1.70% |
| **Unrepresented** | | | | | | | | | |
| Congenital Mesoblastic Nephroma |  | - | - | 0 | - | 0.10% | 30% | 0.07 | 0.10% |
| Cervix |  | - | - | 2 | - | 0.70% | 30% | 0.62 | 0.60% |
| Gastro-oesophageal junction |  | - | - | 4 | - | 1.40% | 30% | 1.16 | 1.10% |
| HNSCC |  | - | - | 24 | - | 8.60% | 30% | 7.14 | 6.60% |
| Neuroendocrine |  | - | - | 7 | - | 2.50% | 30% | 2.08 | 1.90% |
| Ovarian |  | - | - | 4 | - | 1.40% | 30% | 1.13 | 1.00% |
| Papillary thyroid tumour |  | - | - | 44 | - | 15.80% | 30% | 13.07 | 12.10% |
| Paediatric High Grade Glioma |  | - | - | 4 | - | 1.30% | 30% | 1.06 | 1.00% |
| Paediatric melanoma |  | - | - | 2 | - | 0.80% | 30% | 0.62 | 0.60% |
| Prostate |  | - | - | 44 | - | 16.00% | 30% | 13.29 | 12.30% |
| Renal cell carcinoma |  | - | - | 8 | - | 3.00% | 30% | 2.45 | 2.30% |
| Salivary gland |  | - | - | 6 | - | 2.00% | 30% | 1.68 | 1.60% |
| Secretory Breast Carcinoma |  | - | - | 1 | - | 0.20% | 60% | 0.35 | 0.30% |
| Sinonasal adenocarcinoma |  | - | - | 0 | - | 0.00% | 30% | 0 | 0.00% |
| Uterine |  | - | - | 1 | - | 0.50% | 30% | 0.43 | 0.40% |
| High Grade Glioma |  | - | - | 1 | - | 0.50% | 60% | 0.8 | 0.70% |
| Total |  | 55 | 100% | 276 | 100% | 100% | - | 108 | 100% |
| GIST, gastrointestinal stromal tumour; HNSCC, head and neck squamous cell carcinoma; IFS, infantile fibrosarcoma; MASC, mammary analogue secretory carcinoma | | | | | | | | | |

### Implications for the appraisal of histology-independent technologies

In summary, the trial population may include only a subset of the total population potentially eligible for the intervention. Indeed, it is feasible that the majority of histologies potentially harbouring the biomarker will not be represented in the evidence. Furthermore, matching of the line of therapy in the trial population to the eligible population is important given the likely prognostic effect of line of therapy. However, matching can be difficult if there is ambiguity in the treatment position as outlined in the marketing authorisation.

## Genomic testing for histology-independent drugs

Genomic testing is likely to be integral to identify patients eligible for histology-independent therapy. The NICE approval of numerous targeted therapies has been coupled with significant investment in genomic testing services in the NHS.148 Although genomic services are currently set up to identify oncogenic mutations in over 60 tumour types,162 the provision of histology-independent testing poses new challenges that need to be considered before appraising the value of a histology-independent technologies.

### Overview of molecular testing in the UK

Substantial investment and changes to genomic testing services have been undertaken in the last five years after a demand to improve access to genomic services in the NHS, to inform the most effective treatment pathway for a patient with cancer.154 In 2018, the NHS launched the Genomic Medicine Service and a National Genomic Testing Strategy, based in seven Genomic Laboratory Hubs across England.148 While this is providing positive steps to improve availability of genomic testing across the UK, the services are still being implemented leading to limited capacity in some Genomic Laboratory Hubs.

In March 2019, the genomic test directory listed 968 genomic tests available for 64 adult and paediatric tumour types.162 While this may seem an exhaustive number of tests for a large proportion of tumour types, it is far from inclusive. Patients with some common tumours including prostate cancer, a population which contributes 15% of the annual incidence of solid tumours in England, are not eligible for any form of genomic testing, as there is currently no effective targeted therapies licensed on the NHS. Although the absence of genomic testing until the present time may be due to limited evidence of known somatic or hereditary mutation that will be of prognostic or diagnostic value, the provision of a targeted histology-independent therapy would require pan-cancer screening, regardless of current availability.

### Types of Genomic Test

Tumourigenesis, the process of cancer growth and development, is driven by genetic alterations which result in sustained cell proliferation or the inhibition of cell division and death.163 In fact, by the time a cancer is diagnosed, there are likely to be millions of genetic mutations within a single malignancy.164

Many of these alterations occur during the tumour development but do not contribute to tumour growth, commonly known as ‘passenger mutations’, and therefore play no functional role in the cancer’s development. These mutations may occur in non-coding sequences of DNA, which, during the transcription of DNA to RNA as part of gene expression, are removed. In contrast, the ‘driver’ mutations are involved in the neoplastic growth of the tumour, which directly result in the prolific growth of the cancer cells. Driver mutations can be differentiated further, with respect to whether they solely influence the initial cancer development, or whether oncogenic growth and proliferation is dependent upon the mutation regardless of its position in the disease pathway.165

Therefore, the role of genetic testing is two-fold: first, to detect the presence or absence of a specific mutation, but secondly, to determine whether the mutation is acting as an oncogenic driver or whether it is merely a passenger mutation in tumourigenesis.

There are a variety of tests that are available to identify the presence of a mutation in individuals. These include DNA and RNA based panel tests, whole genome sequencing (WGS), IHC, FISH and RT-PCR. Each of these tests determines the presence or absence of a genetic mutation in different ways, from identifying a known driver mutation using targeted tests in DNA and RNA, to sequencing the entire genome, or determining the level of expression of a particular protein. The suitability of alternative types of test will likely depend on the target mutation and the test’s diagnostic accuracy to correctly detect the respective alteration, the prevalence of the genetic mutation within each tumour type as well as current testing provision. Table 17 summarises the key characteristics of each test type noting key advantages and limitations.

Tests may be combined as part of a testing strategy, where confirmatory testing is implemented to verify that a mutation is being expressed. This allows for diagnostic accuracy to be maintained, while reducing the use of more expensive and resource intensive test types. For example, IHC may be used as a screening tool to detect protein expression, with a further confirmatory test implemented to verify that the protein expression is caused by the mutation of interest. The relevance of strategies based around IHC may, however, become more limited as panel testing using NGS is expanded within the NHS.

Because of the variable provision of testing in the NHS across tumour types, the most appropriate testing strategy will likely depend on tumour type. For example, all paediatric patients with advanced and metastatic cancer in the NHS will receive WGS at diagnosis by 2020154 and therefore any testing strategy is likely to be built upon this provision. The appropriateness of each testing strategy will also depend on the prevalence of the genetic alternations across tumour types. Diagnostic accuracy will vary depending on the prevalence of the genetic alteration within each tumour type even when the sensitivity and specificity are held constant, see example in Appendix 12.

Table 17: Summary of test features

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Methodology** | **Advantages** | **Disadvantages** |
| **DNA-based NGS** | Analyses genomic DNA from a tumour sample, and can be used to identify mutations in multiple genes concurrently. Targeted panels can be used to identify particular DNA rearrangement, known to have an oncogenic effect. | * Negligible costs to add an extra mutation target to a panel. * Simultaneous detection of more than one mutation. * Routinely used in the NHS to detect a variety of structural variants across a range of cancer types.162 | * Limited coverage of non-coding (intronic) regions of DNA potentially leading to false negatives. * DNA-based NGS relies on targeted panels, this means that mutations that have not been previously identified and hence, not available on a target panel, cannot be detected.166 |
| **WGS** | Sequences the entire genome of DNA against a comparator to identify specific genetic alterations known to play a role in tumourigenesis. | * The most comprehensive method to detect mutations, especially for novel mutations.166 * Currently available for paediatric patients and patients with soft-tissue sarcoma.162 | * Depth of coverage is much smaller due to the amount of DNA that needs to be sequenced.166 * Resource intensive as there is significant amounts of data produced.164 |
| **RNA-based NGS** | Analyses the transcriptome (the collection of all RNA sequences in a cell). RNA-seq provides a more accurate test for determining whether genetic mutations are expressed as proteins. | * RNA-based NGS provides a more accurate proxy to determine whether the DNA-level mutation has led to protein expression.164 * Simultaneous detection of more than one mutation. * Can also be used to detect novel mutations, which are likely to be missed by targeted panel DNA-based NGS. | * Requires high quality samples, which may make it unsuitable for high-throughput testing.166 * High failure rate, which means that samples will have to be re-tested. * A relatively new test, so currently not routinely available on the NHS162 |
| **IHC** | IHC detects the expression of a protein through the use of antibodies, which bind to a specific receptor (or antigen) on the protein of interest. A tag attached to the antibody will react if bound, producing a stain, signalling the expression of the protein. | * Inexpensive and high throughput167 | * Diagnostic accuracy can be highly variable depending upon target biomarker.168 * If a protein is naturally produced within a cell, IHC cannot differentiate this and the oncogenic, dysfunctional protein.167 * Assays are specific to each individual biomarker so multiple tests would be required to identify multiple mutations. |
| **FISH** | Uses a probe on a sequence of DNA that complements a particular genetic alteration.169 Each probe is labelled with a fluorescent marker, which when illuminated, will indicate the presence of the mutation. | * Inexpensive167 * Relatively high sensitivity and specificity167 | * As the probes are often specific to each mutation, identifying novel mutation or multiple targets using conventional FISH is more challenging.167 |
| **RT-PCR** | Uses a probe on a sequence of RNA that complements a particular genetic alteration.169 Each probe is labelled with a fluorescent marker, which when illuminated, will indicate the presence of the mutation. | * Inexpensive168 * Currently used to detect mutations including gene-fusions167 | * Knowledge of the mutation and target sequence is required.168 * High quality RNA is required. |

### Implications of testing for appraisal of histology-independent technologies

The need for companion diagnostic testing to implement histology-independent technologies has several consequences for cost-effectiveness. These considerations include resource implications associated with implementing testing, the impact of alternative testing strategies on the modelled population, as well as broader implications regarding the feasibility of expanding testing services. These issues are briefly discussed in turn below and followed by a worked example considering the implementation of testing for NTRK fusions.

#### Costs

The costs associated with identifying patients will be driven by a range of factors including the testing strategy adopted and current provision of testing in the NHS. Because these may vary across tumour types, incremental testing costs may also vary across tumour types. Variability in testing costs across tumour types will also be determined by variability in the frequency of a genetic mutation across specific tumour types, with increased rarity increasing the costs of identifying an eligible patient. As is illustrated in the worked example of NTRK fusions below, variability in the frequency of target genetic alterations can be significant, ranging from <0.2% to >90%. This has a significant impact on the number of patients that need to be screened (NNS) and as a consequence, the variability in the tumour type specific costs of identifying patients is similarly wide. Testing costs are a significant source of heterogeneity and, if all testing costs are attributable to a single histology-independent drug, are likely to render a technology cost-ineffective for some tumour types. In the context of NTRK fusions, which on average occur in <0.5% of all advanced cancer patients, the average costs of testing are high and are likely to represent a significant proportion of the total incremental costs associated with the implementation of Trk inhibitors.

#### Attributing testing costs

The current NICE methods guide outlines that the costs of testing should be included if they are specifically associated with the provision of the technology being appraised.4 The implementation of wide scale genomic testing is, however, likely to represent a public good which may allow for the identification of other relevant genetic alterations (e.g. where wide spread panel testing is implemented). This may be of particular relevance where there are multiple targeted therapies available or likely to become available in the near future. Accounting for such positive externalities may be important, as testing costs may not justify the implementation of a specific single technology but may be justifiable when shared across multiple technologies. Estimation of the magnitude of any positive externalities resulting from testing is however, non-trivial and methods of how to attribute testing costs across multiple technologies have not been established. How costs should be attributed across technologies is currently unclear: for example, costs could be split equally or by the size of the eligible population and would necessitate a coordinating role for either NHS England or NICE to potentially set a tariff upon which attributable testing costs could be based.9, 10

#### Feasibility

Although there is currently provision for genomic testing for several cancers within the NHS, there are significant uncertainties surrounding the practical feasibility of providing wide-scale histology-independent testing.

The feasibility of testing is also dependent on whether testing is offered at the point of diagnosis, or at the position in the treatment pathway where the drug would be given. Where testing is implemented upon eligibility for treatment, the NNS will be indicative of the number of patients who will go on to receive treatment, as it is expected that all patients that test positive receive the therapy. Given that entrectinib and larotrectinib are offered when there is no ‘acceptable’ alternative therapy,9, 10 there may be significant disparity between the NNS and the final number of patients that go on to receive therapy. This is because there is significant attrition in the number of patients that go on to receive second or later lines of therapy, either as a result of patients dying or becoming unfit for treatment.

Given the potential variety of histology-independent drugs that could be available across a range of positions in treatment pathways in each tumour type, genomic testing at diagnosis of advanced or metastatic cancer is the most plausible, even though the initial investment would be significant. Based on the annual incidence of cancer in the UK, and the average proportion of individuals with advanced or metastatic cancer, 94,595 individuals would require genomic testing each year. This figure represents a significant increase in the number of molecular and genomic tests and given the variability in the UK’s capacity to implement wide scale NGS testing, it is expected that it will take some time for appropriate infrastructure to be put in place. A phased introduction of NGS panel testing is likely over the next few years with NHS England anticipating that full implementation of pan-cancer testing will be in place by the end of 2022.154

### Identifying patients eligible for Trk-inhibitors based on the presence of an NTRK fusion a worked example

This section considers how testing for NTRK fusions might be implemented in the NHS and provides an illustrative example of how both NNS and testing costs might vary across tumour types.

A variety of testing strategies have been proposed for identifying patients with an NTRK fusion, pending the approval of two histology-independent Trk-inhibitors.168, 170 ESMO proposes that the standard testing pathway should differ depending on the frequency of NTRK fusions in each tumour type, and whether genomic sequencing is currently provided by the NHS.162 In the tumour types where there is a lower frequency of NTRKfusions and where there is no genomic testing available, it is suggested that IHC is used for initial screening; NTRKgene rearrangements are then confirmed using RNA-based NGS. IHC is high-throughput and inexpensive, making it a practical screening tool to use in a large population. Following this with more expensive and highly accurate RNA-based NGS is a plausible testing strategy for identifying tumours. Diagnostic accuracy is further taken into account by stratifying the tumour types into different testing strategies, depending on their NTRK fusion prevalence.

Conversely, it has also been suggeste that front-line NGS should be offered to all individuals to detect an NTRK fusion.170 Whilst this would require substantial investment, this testing strategy would ‘future-proof’ histology-independent testing, as additional mutations could be added to pre-existing panels. This would mean a single tumour sample could be screened to identify a number of genetic alterations. However, in the short-term, this testing strategy will require significant resources to implement nationally.

The most exhaustive approach to identify NTRK fusions utilises DNA-based NGS and RNA-seq. As DNA-based NGS is currently available for some tumours, there will be reduced incremental investment in providing RNA-seq, which would be used to confirm protein expression in the positive cases.

While there is the potential for NTRK fusions to be observed in any tumour type, current evidence only documents the occurrence of NTRK fusions in around 30 different tumour types.171-174 However, this list cannot be considered complete. It is plausible that NTRK fusions occur in common tumour types, but with such rarity that they are yet to be detected. There are also likely to be a number of rarer tumour types which express NTRK fusions, but which are not included in any current database. To align with the available evidence on the prevalence of NTRK fusion, we only distinguish between tumour types within a single anatomical site when there is supporting evidence on the prevalence of NTRK fusions to do so.

#### Methods

For each tumour type where there is evidence to support the prevalence of NTRK fusions, we estimated the following:

* The number of individuals who would require genomic testing each year to identify one individual with an NTRK fusion (NNS)
* The average cost of testing associated with identifying one NTRK-fusion patient
* An illustration of the cost-effectiveness of NTRK testing for each tumour type.

This was implemented for three testing strategies that could be adopted to identify patients with NTRK fusions.

The first testing strategy is based on recent recommendations for identification of NTRK fusions published by the ESMO.170 IHC followed by confirmatory RNA-based NGS would be recommended for the tumour types where NTRK fusions are rare. In the tumours where NTRK fusions are highly prevalent, first-line FISH should be utilised. For the tumour types where WGS is currently available and reimbursed by the NHS, RNA-based NGS would be required to confirm the presence of an oncogenic NTRK fusion.

To complement the substantial investment in genomic testing services in the NHS, the second strategy was assumed to be based on using RNA-based NGS as a first-line test for all patients. For tumour types where whole genome sequencing is currently available, it was assumed that RNA-based NGS would be used to confirm the presence of an oncogenic NTRK fusion.

Finally, an alternative testing strategy was considered based on an exhaustive approach outlined by ESMO, which seeks to maximise current testing availability of DNA-based NGS in each tumour type. Under this approach DNA-based NGS is used as a first-line screening tool, followed by confirmatory RNA-based NGS. This was suggested by ESMO to be the most exhaustive approach to identify NTRK fusions.170

##### Number Needed to Screen

The NNS in order to identify one patient eligible for Trk-inhibitors is based on NTRK fusion prevalence and the diagnostic accuracy of the respective tests, see Appendix 11 for details of calculations. To our knowledge, there is no literature concerning the diagnostic accuracy of WGS in detecting NTRK fusions. As a result, the diagnostic accuracy of WGS for detecting NTRK fusions was based on sensitivity and specificity estimates of DNA-based NGS. Table 18 presents the diagnostic accuracy for each test.

Table 18: Sensitivity and Specificity of each Genomic Test for identifying NTRK fusions

|  |  |  |
| --- | --- | --- |
|  | **Sensitivity** | **Specificity** |
| RNA-seq 175 | 100.00% | 100.00% |
| Whole Genome Sequencing and DNA-based NGS167 | 81.10% | 99.86% |
| Immunohistochemistry 176 | 87.90% | 81.10% |
| Fluorescent in-situ Hybridisation (ETV6-NTRK3) 177 | 80.00% | 100.00% |

The NNS with a first-line (FL) test was estimated using the tumour type specific prevalence of NTRK fusions and the respective first-line test sensitivity (Sn) using the following equation.

Confirmatory RNA-testing is required for patients who require first-line IHC, WGS or DNA-based NGS. The NNS with a confirmatory (C) test was estimated using the sensitivity and specificity of the respective test and the tumour-specific NTRK fusion prevalence.

##### Cost of Testing to Identify One Eligible Patient for Trk-inhibitors

The incremental cost of testing in order to identify one eligible patient was estimated for each tumour. Genomic testing, in the form of DNA-based NGS, WGS and FISH, are currently reimbursed by the NHS for some tumours. The prices for each test were acquired from a UK Genomic Centre; IHC and FISH were costed at £150, DNA and RNA-based NGS were priced at £250 and £350, respectively. The cost of WGS was assumed to be £800.178

The incremental cost to identify one eligible patient in each tumour type was calculated by:

An average of the tumour-specific testing costs was used to calculate the cost of identifying one individual with an NTRK fusion for each testing strategy. The average cost was weighted in accordance with the annual eligible population for Trk-inhibitors in each tumour.

##### Value of implementing NTRK testing

In order to illustrate how cost-effectiveness may vary across tumour types because of the variation in testing costs, a hypothetical scenario was considered in which incremental cost-effectiveness ratios (ICERs) were calculated for each tumour type. This analysis was based on the incremental costs of testing only and excludes other costs associated with the treatment (e.g. drug costs). The ICERs estimated the difference in testing costs between testing for NTRK and current testing provision, relative to the benefits (quantified by QALYs). Incremental benefits were based on the Canadian Agency for Drugs and Technologies in Health (CADTH)179 assessment of the cost-effectiveness of larotrectinib, which estimated that larotrectinib produced an additional 0.833 QALYs per patient compared to standard care.

#### Results

Table 19, Table 20 and Table 21 respectively present, for each testing strategy, the number of individuals who would require testing in order to identify one eligible patient. The incremental cost associated with testing and hypothetical estimates of cost-effectiveness of testing are also presented.

##### Hierarchical Approach: treating at diagnosis

For the tumour types represented in the trial, the tumour specific costs to identify one eligible patient ranged from £0 (MASC) to £351,567 (Breast). Assuming the eligible population is distributed in line with the trial distribution, the average incremental cost of testing is £62,072. Re-weighting the tumour specific costs so that the tumour types included in the trial aligns with the prevalence of these tumours expected in the real world increases this £108,963. However, this does not account for the testing costs associated with tumours not represented in the larotrectinib trial. When these are included and appropriately weighted in line with the prevalence of these tumour types, the average incremental cost to identify one individual eligible for treatment is £85,253. Based on this cost estimate andassuming that larotrectinib generates an additional 0.833 incremental QALYs, the ICER is estimated to be £102,345 per QALY gained, with tumour-specific ICERs ranging from less than £500 per QALY gained to over £500,000 per QALY gained.

Table 19: Summary of NNS and testing cost under a hierarchical testing approach

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tumour Type** | **Number Needed to Screen** | | **Incremental cost to identify one patient** | **Incremental Cost-effectiveness Ratio**  **(NTRK fusion testing vs current testing provision)** |
| **First Line** | **Confirmatory** |
| **Tumours in trial** | | | | |
| Appendix | 30.83 | 1.04 | £4,989 | £5,989 |
| Breast | 1625.22 | 307.95 | £351,567 | £422,049 |
| Cholangiocarcinoma | 1137.66 | 215.80 | £246,179 | £295,533 |
| Colorectal | 948.05 | 179.97 | £205,195 | £246,333 |
| GIST | 88.88 | 17.58 | £19,486 | £23,393 |
| Infantile fibrosarcoma | 1.36 | 1.00 | £350 | £420 |
| MASC | 1.35 | 0.00 | £0 | £0 |
| Melanoma | 541.74 | 103.17 | £117,372 | £140,903 |
| NSCLC | 1264.06 | 239.69 | £273,502 | £328,334 |
| Pancreatic | 437.56 | 83.48 | £94,853 | £113,870 |
| Soft tissue sarcoma | 220.19 | 1.31 | £457 | £549 |
| Thyroid | 123.66 | 24.16 | £27,003 | £32,417 |
| **Tumours not represented in trial** | | | | |
| Cervix | 344.74 | 65.94 | £74,791 | £89,785 |
| Congenital Mesoblastic Nephroma | 2.03 | 1.00 | £350 | £421 |
| Gastro-esophageal junction | 1137.66 | 215.80 | £246,179 | £295,533 |
| HNSCC | 299.38 | 57.37 | £64,986 | £78,015 |
| High grade glioma | 2275.31 | 430.82 | £492,084 | £590,737 |
| Neuroendocrine | 379.22 | 72.46 | £82,243 | £98,731 |
| Ovarian | 455.06 | 86.79 | £98,637 | £118,411 |
| Papillary thyroid tumour | 8.55 | 2.40 | £2,124 | £2,549 |
| Paediatric high grade glioma | 23.27 | 1.03 | £361 | £433 |
| Paediatric melanoma | 11.10 | 1.01 | £355 | £426 |
| Prostate | 455.06 | 86.79 | £98,637 | £118,411 |
| Renal cell carcinoma | 455.06 | 86.79 | £98,637 | £118,411 |
| Salivary gland | 66.14 | 13.29 | £14,572 | £17,493 |
| Secretory breast carcinoma | 1.36 | 0.00 | £0 | £0 |
| Sinonasal adenocarcinoma | 455.06 | 86.79 | £98,637 | £118,411 |
| Uterine | 1137.66 | 215.80 | £246,179 | £295,533 |

##### First line RNA-based NGS: treating at diagnosis

The incremental cost of testing to identify one individual eligible for Trk-inhibitors using first-line RNA-based NGS is higher across tumour types compared to the hierarchical approach (with the exception of the tumours where WGS is available). For the tumours included in the larotrectinib trial, the incremental testing costs range from £215 (MASC) to £500,000 (breast cancer). Based on the trial distribution, the average incremental cost to identify one eligible patient is £95,324. Using the real-world distribution, this increases substantially to £159,960 per patient identified. When including the unrepresented tumour types, the average incremental costs fall slightly to £123,023 per patient identified. If larotrectinib were to provide an incremental clinical benefit of 0.833 QALYs, and assuming zero treatment costs, these testing costs would imply an ICER of £147,687 per QALY gained.

Table 20: Summary of NNS and testing cost under first-line RNA-based NGS approach

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tumour Type** | **Number Needed to Screen** | | **Incremental Cost to identify one patient** | **Incremental Cost-effectiveness Ratio** |
| **First Line** | **Confirmatory** |
| **Tumours in Trial** | | | | |
| Appendix | 30.83 | 1.04 | £10,791 | £12,954 |
| Breast | 1428.57 | 0.00 | £500,000 | £600,240 |
| Cholangiocarcinoma | 1000.00 | 0.00 | £350,000 | £420,168 |
| Colorectal | 833.33 | 0.00 | £291,667 | £350,140 |
| GIST | 78.13 | 0.00 | £27,344 | £32,826 |
| Infantile fibrosarcoma | 1.36 | 1.00 | £350 | £420 |
| MASC | 1.08 | 1.00 | £215 | £258 |
| Melanoma | 476.19 | 0.00 | £166,667 | £200,080 |
| NSCLC | 1111.11 | 0.00 | £388,889 | £466,853 |
| Pancreatic | 384.62 | 0.00 | £134,615 | £161,603 |
| Thyroid | 220.19 | 1.31 | £457 | £549 |
| Soft tissue sarcoma | 108.70 | 0.00 | £38,043 | £45,670 |
| **Tumours not represented in trial** | | | | |
| Cervix | 303.03 | 0.00 | £106,061 | £127,324 |
| Congenital Mesoblastic Nephroma | 2.03 | 1.00 | £350 | £421 |
| Gastro-esophageal junction | 1000.00 | 0.00 | £350,000 | £420,168 |
| HNSCC | 263.16 | 0.00 | £92,105 | £110,571 |
| High grade glioma | 2000.00 | 0.00 | £700,000 | £840,336 |
| Neuroendocrine | 333.33 | 0.00 | £116,667 | £140,056 |
| Ovarian | 400.00 | 0.00 | £140,000 | £168,067 |
| Papillary thyroid tumour | 7.52 | 0.00 | £2,632 | £3,159 |
| Paediatric high grade glioma | 23.27 | 1.03 | £361 | £433 |
| Paediatric melanoma | 11.10 | 1.01 | £355 | £426 |
| Prostate | 400.00 | 0.00 | £140,000 | £168,067 |
| Renal cell carcinoma | 400.00 | 0.00 | £140,000 | £168,067 |
| Salivary gland | 58.14 | 0.00 | £20,349 | £24,428 |
| Secretory breast carcinoma | 1.09 | 1.00 | £218 | £262 |
| Sinonasal adenocarcinoma | 400.00 | 0.00 | £140,000 | £168,067 |
| Uterine | 1000.00 | 0.00 | £350,000 | £420,168 |

##### Exhaustive Approach: treating at diagnosis

Under the most exhaustive approach where DNA-based NGS is used as a first-line test, followed by confirmatory NGS, the incremental costs of testing are significantly higher than under the other two strategies. The current reimbursement of WGS, the incremental cost associated with identifying an individual who is eligible for a Trk-inhibitor is lowest in IFS (£350). As there is no genomic testing currently available for patients with cholangiocarcinoma, identifying one patient eligible for a Trk-inhibitor is the most costly (£739,827) within the trial population. Based on the distribution of tumour types within the trial, the average incremental testing costs associated with identifying one individual eligible for treatment equals £133,601.

Under the real-world distribution, this increases £205,652 per patient identified. The average incremental cost to identify one individual eligible for treatment, including the tumour types unrepresented in the trial is £198,993. The associated ICER was £238,887 per QALY gained.

Table 21: Summary of NNS and testing cost under exhaustive testing strategy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tumour Type** | **Number Needed to Screen** | | **Incremental Cost to identify one patient** | **Incremental Cost-effectiveness Ratio** |
| **First Line** | **Confirmatory** |
| **Tumours in trial** | | | | |
| Appendix | 30.83 | 1.04 | £18,498 | £22,206 |
| Breast | 1761.49 | 3.46 | £616,523 | £740,123 |
| Cholangiocarcinoma | 1233.05 | 2.72 | £739,827 | £888,148 |
| Colorectal | 1027.54 | 2.44 | £359,638 | £431,739 |
| GIST | 96.33 | 1.13 | £57,799 | £69,387 |
| Infantile fibrosarcoma | 1.36 | 1.00 | £350 | £420 |
| MASC | 1.33 | 1.00 | £597 | £717 |
| Melanoma | 587.16 | 1.82 | £205,508 | £246,708 |
| NSCLC | 1370.05 | 2.92 | £479,518 | £575,652 |
| Pancreatic | 474.25 | 1.66 | £284,549 | £341,595 |
| Thyroid | 220.19 | 1.31 | £457 | £549 |
| Soft tissue sarcoma | 134.03 | 1.19 | £46,909 | £56,314 |
| **Tumours not represented in trial** | | | | |
| Cervix | 373.65 | 1.52 | £224,190 | £269,136 |
| Congenital Mesoblastic Nephroma | 2.03 | 1.00 | £350 | £421 |
| Gastro-esophageal junction | 1233.05 | 2.72 | £739,827 | £888,148 |
| HNSCC | 324.49 | 1.45 | £194,691 | £233,723 |
| High grade glioma | 2466.09 | 4.45 | £1,479,655 | £1,776,296 |
| Neuroendocrine | 411.02 | 1.57 | £246,609 | £296,049 |
| Ovarian | 493.22 | 1.69 | £172,626 | £207,235 |
| Papillary thyroid tumour | 9.27 | 1.01 | £3,245 | £3,895 |
| Paediatric high grade glioma | 23.27 | 1.03 | £361 | £433 |
| Paediatric melanoma | 11.10 | 1.01 | £355 | £426 |
| Prostate | 493.22 | 1.69 | £295,931 | £355,259 |
| Renal cell carcinoma | 493.22 | 1.69 | £295,931 | £355,259 |
| Salivary gland | 71.69 | 1.10 | £43,013 | £51,637 |
| Secretory breast carcinoma | 1.34 | 1.00 | £605 | £726 |
| Sinonasal adenocarcinoma | 493.22 | 1.69 | £295,931 | £355,259 |
| Uterine | 1233.05 | 2.72 | £739,827 | £888,148 |

##### Exhaustive Approach: treating at line of therapy.

If, testing were to be done at diagnosis of advanced or metastatic cancer and larotrectinib is offered in the appropriate position in the treatment pathway, the average incremental testing cost to identify one individual eligible for treatment is higher compared to when treatment is offered as first line therapy. The annual population eligible for treatment is lower (n = 94), thus the average incremental cost of testing to identify one individual eligible for treatment is estimated to be £516,635.

Table 22: Summary of NNS and testing cost under exhaustive testing strategy when testing is done at diagnosis and treatment is provided at the appropriate line of therapy

|  |  |  |  |
| --- | --- | --- | --- |
| **Tumour Type** | **Proportion eligible based on treating at line of therapy** | **Incremental Cost to Identify NTRK fusion patient** | **Incremental Cost-Effectiveness Ratio** |
| **Tumours in Trial** | | | |
| Appendix | 30% | £4,050 | £4,862 |
| Breast | 30% | £2,055,076 | £2,467,078 |
| Cholangiocarcinoma | 30% | £2,466,091 | £2,960,494 |
| Colorectal | 30% | £1,198,794 | £1,439,129 |
| GIST | 30% | £192,663 | £231,289 |
| Infantile Fibrosarcoma | 90% | £389 | £467 |
| MASC | 60% | £995 | £1,195 |
| Melanoma | 30% | £685,025 | £822,359 |
| NSCLC | 60% | £799,196 | £959,419 |
| Pancreatic | 30% | £948,497 | £1,138,651 |
| Soft tissue sarcoma | 90% | £508 | £610 |
| Thyroid | 30% | £156,364 | £187,712 |
| **Tumours not represented in trial** | | | |
| Cervix | 30% | £747,300 | £897,119 |
| Congenital Mesoblastic Nephroma | 30% | £1,168 | £1,402 |
| Gastro-oesophageal junction | 30% | £2,466,091 | £2,960,494 |
| HNSCC | 30% | £648,971 | £779,077 |
| High Grade Glioma | 60% | £2,466,091 | £2,960,494 |
| Neuroendocrine | 30% | £822,030 | £986,831 |
| Ovarian | 30% | £575,421 | £690,782 |
| Papillary thyroid tumour | 30% | £10,816 | £12,985 |
| Paediatric High Grade Glioma | 30% | £1,203 | £1,444 |
| Paediatric melanoma | 30% | £1,183 | £1,420 |
| Prostate | 30% | £986,436 | £1,184,197 |
| Renal cell carcinoma | 30% | £986,436 | £1,184,197 |
| Salivary gland | 30% | £143,377 | £172,122 |
| Secretory Breast Carcinoma | 60% | £1,008 | £1,211 |
| Sinonasal adenocarcinoma | 30% | £986,436 | £1,184,197 |
| Uterine | 30% | £2,466,091 | £2,960,494 |

### Implications for the cost-effectiveness of Trk-inhibitors

The costs associated with additional testing for targets are likely to be substantial and will have significant bearing on the cost-effectiveness of histology-independent technologies. Tumour specific costs of identifying relevant fusions are also likely to represent a significant source of heterogeneity due to the variable frequency of targets across tumour types. This heterogeneity in testing costs is likely to mean that for some tumour types where NTRK fusions are rare, NTRK testing will not be cost-effective. Opportunities to share testing costs across multiple health care technologies may reduce the cost burden of molecular testing on a specific healthcare technology, potentially increasing the financial viability of testing. It is however, currently unclear what mechanism would be used to share testing costs across multiple technologies.

## Model structure and extrapolation

Partitioned survival modelling (PSM) is the most common modelling approach used for NICE appraisals of interventions for advanced or metastatic cancers.180 This approach uses survival analysis of observed time-to-event endpoints (TTE) to derive state membership estimates. As estimates of mean survival times are required for cost-effectiveness analysis, parametric models are fitted to the observed TTE endpoints to extrapolate the observed survival data over an appropriate time horizon. The choice of appropriate parametric models to extrapolate the observed data are usually based on a series of assessments, including: visual inspection of the Kaplan-Meier curves and log cumulative hazard plots; visual fit of extrapolated models to observed data; statistical fit based on goodness of fit statistics; and clinical plausibility of the extrapolation.181 In a PSM, PFS and OS are usually extrapolated independently and directly inform the state membership for the ‘Progression-free’ and ‘Death’ states over time, respectively. The difference between PFS and OS allows the proportion of patients in the progressed health state to be estimated.

The use of PSM presents several challenges for the assessment of histology-independent products. Firstly, the more heterogeneous overall population may make it more challenging to fit a single conventional parametric curve. Secondly, the immaturity of the PFS and OS data will result in considerable uncertainty surrounding the extrapolated curves. This may lead to wide variation in the resulting predictions for survival models that have similar goodness of fit to the observed data. Each of these challenges is now considered in more detail.

Standard parametric models can include proportional hazards-based models (exponential, Weibull and Gompertz), and the accelerated failure time models (log-normal, log-logistic and generalised gamma). However, the additional heterogeneity arising from the inclusion of different tumour sites is likely to result in more complex hazard functions which may not be appropriately captured using standard parametric distributions. Consequently, the use of flexible parametric models, mixture models or response-based models may be required.182

Flexible parametric approaches directly model the effect of time on the hazard function using splines. The splines are used to form a series of polynomial distributions joined by ‘knots’. Changes in the modelled hazard function at specific time points can be accommodated by using different polynomial distributions between each knot. The number of knots determines the number of parameters required to model the hazard function. In a simple case with zero knots, these models are the same as conventional parametric distributions. By definition, these approaches provide greater flexibility compared to conventional parametric modelling. However, they also present potential challenges for histology-independent appraisals. Firstly, the approach captures heterogeneity via the effect of time on the hazard function. This may not be appropriate for achieving accurate projections of PFS and OS where the main source of heterogeneity is the difference in the natural history between different tumour sites. Secondly, the flexible parametric approach extrapolates beyond the data using only the final segment of the curve. The small number of patients and immaturity of the PFS and OS data may mean that survival projections are particularly unreliable in the final segment.

While the impact of the inclusion of tumour sites which may have different natural histories on the hazard function might be accommodated by flexible parametric approaches, the inclusion of multiple subsets of patients may provide evidence of different survival distributions within the observed TTE data. Parametric mixture models can be used to capture heterogeneity within a population by using two (or more) distinct distributions. While the use of a mixture model may provide a more appropriate approach to capture between-tumour heterogeneity, there may remain challenges in determining how many mixes are appropriate and whether the predicted long-term hazards are plausible from the resulting mixture. There also remain issues regarding the application of different HRQoL or cost estimates for individual tumour sites as the different mixture distributions are not explicitly assigned to any individual tumour site or grouping. Hence, while mixture models provide an approach to account for heterogeneity within the TTE endpoints, they do not provide a basis for accounting for heterogeneity in other inputs which may impact on cost-effectiveness estimates.

Another approach which has been proposed to account for heterogeneity in TTE endpoints is the use of response-based landmark models. This approach models survival conditional on response status identified at a predefined response evaluation landmark time based on a clinical definition of response. Survival is modelled from the landmark point to avoid the problem of ‘immortal time’ bias arising from the fact that responders have to survive to the point at which response is assessed. Separate survival curves are then fitted to the different response categories. Intuitively this approach appears particularly aligned to the appraisal of histology-independent technologies where response measures are used as the primary endpoint. This approach allows for a distinction to be made in the HRQoL of responders and non-responders (and between individual tumour sites), as well as allowing for potential differences in the costs of care. However, there may be challenges in determining whether a single landmark time point is appropriate and how uncertainty around this should be dealt with. Although different response time points can be accommodated within the survival analysis using a time varying covariate, inevitably this will increase the complexity of the economic model and may require individual patient simulation approaches. There also remain issues concerning the potentially small number of patients recruited in the underpinning studies and immaturity of the PFS and OS data. Further subdividing patients into responder categories may result in more uncertain survival predictions. In addition, while separate survival curves may better account for heterogeneity within the survival data, the approach does not resolve the fundamental problem of immaturity in these endpoints.

Although several approaches exist to account for heterogeneity in survival endpoints, they all have several important limitations in the context of histology-independent appraisals. Firstly, the use of single ‘full population’ ICER, across multiple tumour sites with potentially different treatment effectiveness, comparators, costs and HRQoL, will be difficult to interpret. A single ICER may conceal significant variation in the tumour specific ICERs, driven by a combination of factors, including observable variability in relative effectiveness between tumour types. Ignoring these differences could mean that a treatment which is not cost-effective for the total population (combining all subgroups) may be cost-effective in specific subgroups. Conversely, a treatment which appears cost-effective for the total population may not be cost-effective in particular subgroups. Given the amount of heterogeneity associated with a histology-independent appraisal, estimating the average cost-effectiveness for the full patient population covered in the scope may not provide enough information to decision-makers about whether the drug is potentially cost-effective across all subgroups.

Secondly, the approaches rely on extrapolations of the observed survival data which will potentially be immature at the time of initial appraisal such that the resulting predictions will be highly uncertain. Different survival models which appear to fit the observed data equally well may lead to significant variation in longer-term survival predictions. Consequently, it is unlikely that a single survival distribution (or a single specification of a more flexible parametric, mixture model or response-based approach) will adequately characterise uncertainties over the longer-term extrapolation period. To more formally account for the uncertainty surrounding choice of survival distribution, a model averaging approach may be needed.183, 184 This approach involves the parameterisation of uncertainty surrounding the choice of distribution, incorporating all plausible distributions as part of a weighted distribution. Uncertainty in the probabilistic analysis will then reflect both the parametric uncertainty associated within each distribution and the uncertainty surrounding the choice of preferred method. However, such an approach presents additional challenges in the context of histology-independent appraisals where the external validation of survival projections from a heterogeneous population including multiple tumour types will be difficult and expert elicitation maybe required to determine the weights to be applied as part of any weighted distribution. Furthermore, this heterogeneity will also result in the ‘at-risk’ population changing over time. That is, tumour types with poorer prognosis will experience events earlier than patients with a more favourable prognosis. Hence, the composition of the population will likely change significantly over the extrapolation period. This limits the appropriateness of applying a single ‘average’ utility or cost to the population within the model.

The greater immaturity in PFS/OS for trials which are powered on response endpoints may present challenges to fitting reliable survival distributions. In these circumstances, surrogate relationships may be needed to link response-based outcomes (e.g. ORR, DoR) to longer-term estimates of PFS and/or OS. Although Section 4 highlighted a range of alternative approaches that could be used, the lack of any clear pattern by cancer type inevitably presents challenges for using a surrogate-based modelling approach to a model which includes a heterogeneous mix of patients.

Given the importance of exploring the impact of heterogeneity more explicitly for decision-making, approaches are needed which can accommodate different sources of heterogeneity within the overall population, which can more appropriately estimate the average cost-effectiveness for the full patient population covered in the scope, and which facilitates assessment of whether the drug is potentially cost-effective across all subgroups. The BHM framework provides a potential approach which can more fully explore the potential heterogeneity in effects across tumours. The BHM allows assessments to be made for each tumour type, as well as a pooled assessment across all tumour types, accounting for the potential lack of uniformity of effect across tumours. An additional advantage of this framework is the ability to predict the response probability that would be expected in a ‘new’ tumour type (i.e. a tumour that is not represented in the trial data), which will give a measure of the uncertainty in the response rates in tumour types in the target population but for which no data are available (see Section 6.1).

Heterogeneity in time to event outcomes (PFS, OS) can be explored using the BHM in a similar way to those presented for response outcomes.44 The model assumes a common parametric distribution for each tumour type, but with a different location parameter. Information on this parameter can be borrowed across the different tumours, according to an estimated heterogeneity parameter. The results from this model would be different distributions of PFS or OS for each tumour type which could be incorporated in the economic model in order to further explore how heterogeneity in outcomes by tumour type influences the expected ICERs. Although the BHM can borrow information across tumour types and is designed to allow inferences with few events per tumour type, it is unclear whether this type of model would provide useful results given the immaturity of the survival data, the small number of patients in most tumour types, the expected lack of exchangeability of the survival outcomes and the potential for requiring informative prior distribution.

To address concerns regarding the maturity of the TTE endpoints, BHM could be applied to specific landmark survival time points (e.g. 6 or 12 months) for which more robust data exists, with surrogate relationships employed to predict longer term survival conditional on survival up to these specific time points. Alternatively, BHM could be applied to the response data itself, where fewer observations are required on response outcomes to draw meaningful conclusions about differences between tumour types. These response assessments could then be applied to conditional PFS and OS distributions from the overall population or linked to external surrogate relationships. However, as reported in Chapter 4, the use of external surrogate relationships would require the use of surrogate multivariable statistical models to estimate the final outcome (OS/PFS) which may not specifically relate to the different tumours or a specific biomarker population.

While 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 of and importance of heterogeneity. This would appear more appropriate than ignoring this heterogeneity within initial assessments. Importantly, such assessments may also help in guiding further data collection and prioritising specific subgroups where existing evidence may be scarce and/or where these exploratory analyses indicate potentially important impacts on the likely cost-effectiveness of a new treatment within the full population.

## Summary and implications

The previous sections identified a number of particular challenges for histology-independent appraisals and have explored alternative approaches which might be used to investigate and account for different sources of heterogeneity and uncertainty. Although not comprehensive, we have focused on areas of evidence and analysis anticipated to be most challenging for the appraisal of histology-independent products. Given the nature of these challenges, it is likely that a range of alternative approaches will be required to address different sources of heterogeneity. The implications for the assessments of cost-effectiveness and uncertainty will also need to be made explicit. Equally important is the need to ensure that these assessments present results in a manner which can help inform NICE decisions, both in determining the appropriateness of different recommendations and identifying key uncertainties which might be used to inform and prioritise the value of further data collection.

Chapter 7 presents a potential framework which could be used to inform approval and research policies for histology-independent products, including NICE decision-making and CDF data collection arrangements.

# A decision framework to inform approval and research policies for histology-independent technologies

An exemplar case study was developed to illustrate the nature of the assessments that could be used to evaluate the cost-effectiveness of a new histology-independent treatment. Based on these assessments, a framework is proposed to help inform approval and research policies for histology-independent technologies. A brief summary of the case study is presented in Section 7.1. Further details are reported separately in Appendix 12.

## Exemplar case study

The case study considers a hypothetical Trk-inhibitor (Drug X) for the treatment of solid tumours that harbour an NTRK gene fusion compared to the current standard of care (SoC). Although the case study draws on clinical evidence from an existing Trk-inhibitor, specifically the response outcomes and the BHM reported in Section 6.1 for larotrectinib, all other inputs are based on stylised assumptions. Importantly, the purpose of the case study is not to make any recommendations concerning the likely cost-effectiveness of any existing or new histology-independent treatment. Instead, the aim is to illustrate the nature and sequence of assessments that could potentially be used to help inform NICE approval decisions and CDF data collection arrangements.

The economic model uses a landmark response-based structure (see Section 6.5) which incorporates separate PFS and OS distributions, conditioned on response status in the overall study population. That is, the same conditional PFS and OS distributions assumed for responders and non-responders are applied to each individual histology. The use of conditional PFS and OS data therefore assumes a perfect surrogate relationship between response outcomes and PFS and OS endpoints, which is the same across all tumour types. Hence, heterogeneity in PFS and OS across individual histologies is assumed in the case study to be entirely mediated through different response rates.

The use of a response-based modelling approach necessitates additional assumptions compared to a situation in which robust TTE data are available for the overall population and each specific subgroup of interest. Equally, there may be a range of alternative modelling approaches which could be developed based on landmark survival times and/or alternative surrogate relationships. The purpose of the case study is not to make specific recommendations regarding the model structure and associated parameter assumptions but rather to present a more general framework to demonstrate how heterogeneity within the overall population could potentially be explored within a cost-effectiveness analysis and how the results could be presented to more appropriately inform alternative policy decisions. However, it should be acknowledged that assessments of heterogeneity in survival outcomes at the point of intial marketing authorisation may be challenging unless these are linked to a surrogate outcome (e.g. response, DoR), for which more robust assessments of heterogeneity are likely to be feasible.

The model structure consists of three mutually exclusive health states: (i) progression-free disease, (ii) progressed-disease, and (iii) death. State occupancy in the model is derived using a dual-partitioned survival approach which uses PFS curves to partition OS into those patients in progression-free and those with progressed disease, based on response status at a specific landmark time point.

Survival for Drug X is calculated as a weighted average of the responder and non-responder survival curves based on the overall response rate assumed in the analysis. Survival in the standard of ccare (SoC) arm was modelled assuming a 0% response. The case study therefore also makes a strong assumption that effectiveness for SoC management is the same across all tumour types and is equal to the conditional PFS and OS estimates derived from non-responders to Drug X.

In line with the NICE reference case, the model considers a NHS and Personal Social Services perspective in terms of capturing costs and QALYs and discounts both using a 3.5% discount rate. Results are presented over a lifetime (30-year) time horizon.

Response rates used in the analysis were based on the BHM analysis of the larotrectinib FDA data (Table 13). By linking the BHM estimates for response rates to conditional OS and PFS estimates, the case study model explores the implications for cost-effectiveness of heterogeneity in the overall population by considering individual histology specific estimates of cost-effectiveness alongside estimates for the overall population.

Stylised input parameters were used for all other economic model parameters and are summarised in Table 23. The acquisition cost for Drug X in the case study was assumed to be priced at a level such that the ICER in the overall population would be close to the upper limit of NICE’s end of life threshold range (circa £50,000 per QALY gained). As a number of separate scenarios are presented in the case study, the estimate of the acquisition cost was derived from the scenario which was considered to best represent a base-case scenario. This scenario included testing costs and estimates of the effectiveness in tumour sites which were not represented in the clinical evidence base. For a more detail description of the underlying assumptions and justification for the parameters, see Appendix 13.

Table 23: Input parameters included in the economic model

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Value** |  | **95% CI** |
| Effectiveness (months)\* |  |  |  |
| Median PFS |  |  |  |
| Responders | 24 |  | [21.6; 26.4] |
| Non-responders | 6 |  | [5.4; 6.6] |
| Median OS |  |  |  |
| Responders | 36 |  | [32.4; 39.6] |
| Non-responders | 12 |  | [10.8; 13.2] |
| Utilities |  |  |  |
| Progression-free |  |  |  |
| Drug X | 0.79 |  | [0.71;0.87] |
| SoC | 0.72 |  | [0.65;0.79] |
| Post-progression |  |  |  |
| Drug X | 0.64 |  | [0.57;0.71] |
| SoC | 0.64 |  | [0.57;0.71] |
| Costs (£/month) |  |  |  |
| Drug acquisition costs |  |  |  |
| Drug X | £1250 |  | - |
| SoC | £20 |  | - |
| Health state costs\*\* |  |  |  |
| Progression-free | £350 |  | [£315; £385] |
| Post-progression | £500 |  | [£450; 550] |
| Terminal care cost | £6878 (one-off cost) |  | - |
| \*It is assumed the survival function of responders and non-responders follows an exponential distribution  \*\*Health state costs are assumed to be the cost of care excluding treatment costs per individual per month | | | |

The model results are based on a probabilistic sensitivity analysis (PSA), which was implemented using 10,000 samples.

## Histology specific ICERs and overall cost-effectiveness

The case study starts with an assessment of cost-effectiveness based on the trial population and excludes testing costs. Issues around the generalisability of the trial population and the impact of including testing costs are then explored.

Table 24 presents the mean total costs, QALYs and ICERs associated with the histology-independent technology (Drug X) and SoC for each histology included in the trial. Mean survival with the SoC is less than two years for all individual histologies and Drug X is expected to increase life expectancy by greater than 3 months. This suggests that the end of life criteria has been met and so the ICER for all histologies should be compared against a maximum threshold of £50,000 per additional QALY.4

The ICERs estimated for the individual histologies range from £27,213 to £37,930 per QALY gained. The large differences in response rates, ranging between 29.9% and 93.3%, appear to have only a moderate effect on the ICER estimates reported across individual histology sites. The reason for this is that the overall cost of Drug X is assumed to be closely related to the expected survival outcomes of treatment, specifically the duration of PFS. As the response rate increases (or decreases), the duration of treatment also increases (or decreases) such that the total cost of the treatment is closely related to the expected survival outcomes. For treatment regimens which are given for a fixed duration, as opposed to a treat until disease progression (or unacceptable toxicity) strategy, the impact of heterogeneity in the response data would be expected to have a greater impact on the ICER estimates across individual histologies. Similarly, in situations in which heterogeneity in the surrogate relationship is also evident across tumour sites, a greater impact on the ICER estimates across individual histologies would be expected.

Table 24: Histology specific incremental cost-effectiveness ratios (ICERs)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subgroup** | **Per-patient level** | | |  |
|  |  | **Cost** | **QALYs** | **ICER** |
| Sarcoma | Drug X | £61,314 | 2.70 | £27,520 |
| SoC | £14,471 | 0.99 | - |
| Salivary | Drug X | £58,697 | 2.58 | £27,969 |
| SoC | £14,471 | 0.99 | - |
| IFS | Drug X | £63,332 | 2.79 | £27,213 |
| SoC | £14,471 | 0.99 | - |
| Thyroid | Drug X | £62,615 | 2.76 | £27,318 |
| SoC | £14,471 | 0.99 | - |
| Lung | Drug X | £55,032 | 2.41 | £28,721 |
| SoC | £14,471 | 0.99 | - |
| Melanoma | Drug X | £46,963 | 2.03 | £31,267 |
| SoC | £14,471 | 0.99 | - |
| Colon | Drug X | £38,667 | 1.65 | £36,857 |
| SoC | £14,471 | 0.99 | - |
| GIST | Drug X | £61,234 | 2.69 | £27,535 |
| SoC | £14,471 | 0.99 | - |
| Cholangiocarcinoma | Drug X | £34,261 | 1.45 | £43,658 |
| SoC | £14,471 | 0.99 | - |
| Appendix | Drug X | £37,773 | 1.61 | £37,859 |
| SoC | £14,471 | 0.99 | - |
| Breast | Drug X | £37,768 | 1.61 | £37,863 |
| SoC | £14,471 | 0.99 | - |
| Pancreas | Drug X | £37,751 | 1.61 | £37,930 |
| SoC | £14,471 | 0.99 | - |

A ‘histology-independent’ recommendation is defined here as the approval of Drug X for use in any histology which exhibits the specific biomarker (e.g. NTRK). If a histology-independent approval is saught then it is necessary to consider the ‘average’ or ‘pooled’ ICER across all histologies. Table 25 illustrates how a pooled ICER is calculated with the frequency of each histology based on the relative histology frequency observed in the trial (see Table 14, Section 6.3.1 for details).

Table 25: Calculating a pooled ICER based on trial histology frequency

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Observed outcomes** | |  | **Weighted consequences** |  |
|  | **ΔCost** | **ΔQALYs** | **Frequency** | **ΔCost** | **ΔQALYs** |
| Sarcoma | £46,844 | 1.70 | 20.00% | £9,369 | 0.34 |
| Salivary gland | £44,227 | 1.58 | 21.82% | £9,649 | 0.35 |
| IFS | £48,861 | 1.80 | 12.73% | £6,219 | 0.23 |
| Thyroid | £48,144 | 1.76 | 9.09% | £4,377 | 0.16 |
| Lung | £40,561 | 1.41 | 7.27% | £2,950 | 0.10 |
| Melanoma | £32,492 | 1.04 | 7.27% | £2,363 | 0.08 |
| Colon | £24,197 | 0.66 | 7.27% | £1,760 | 0.05 |
| GIST | £46,763 | 1.70 | 5.45% | £2,551 | 0.09 |
| Cholangiocarcinoma | £19,791 | 0.45 | 3.64% | £720 | 0.02 |
| Appendix | £23,302 | 0.62 | 1.82% | £424 | 0.01 |
| Breast | £23,297 | 0.62 | 1.82% | £424 | 0.01 |
| Pancreas | £23,280 | 0.61 | 1.82% | £423 | 0.01 |
| **Sum** |  |  |  | **£41,227** | **1.44** |
|  |  |  |  |  |  |
|  |  |  |  | **Pooled ICER** | **£28,573** |

The pooled ICER is simply a weighted average of the additional mean total costs of Drug X (ΔCost = £41,227) divided by a weighted average of the additional QALYs (ΔQALYs = 1.44), resulting in an ICER of approximately £28,573 per QALY.

This analysis illustrates that the pooled cost-effectiveness of Drug X depends on the frequency and distribution of the individual histologies. The frequency of particular histologies in the target population will ultimately be determined by the testing strategy implemented in clinical practice. Depending on the testing strategy and the expected distribution of histologies in the target population, the pooled ICER may alter. For example, evidence on the expected prevalence of NTRK fusions in specific histologies suggests that the distribution of histologies expected in clinical practice may differ significantly from that observed in the trial.

Table 26 shows the relative frequency of histologies expected in clinical practice assuming routine screening of all histologies. The table also illustrates how the pooled ICER can change based on differences in the expected distribution of histologies in the target population compared to those observed within the trial.

Table 26: Calculating a pooled ICER based on histology frequency in the target population

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Observed outcomes** | |  | **Weighted consequences** |  |
|  | **ΔCost** | **ΔQALYs** | **Frequency** | **ΔCost** | **ΔQALYs** |
| Sarcoma | £46,844 | 1.70 | 3.93% | £1,840 | 0.07 |
| Salivary gland | £44,227 | 1.58 | 1.80% | £796 | 0.03 |
| IFS | £48,861 | 1.80 | 21.89% | £10,693 | 0.39 |
| Thyroid | £48,144 | 1.76 | 5.01% | £2,412 | 0.09 |
| Lung | £40,561 | 1.41 | 13.37% | £5,424 | 0.19 |
| Melanoma | £32,492 | 1.04 | 2.08% | £675 | 0.02 |
| Colon | £24,197 | 0.66 | 18.39% | £4,450 | 0.12 |
| GIST | £46,763 | 1.70 | 3.01% | £1,407 | 0.05 |
| Cholangiocarcinoma | £19,791 | 0.45 | 0.27% | £53 | 0.00 |
| Appendix | £23,302 | 0.62 | 12.78% | £2,978 | 0.08 |
| Breast | £23,297 | 0.62 | 3.87% | £902 | 0.02 |
| Pancreas | £23,280 | 0.61 | 13.61% | £3,169 | 0.08 |
| **Sum** |  |  |  | **£34,798** | **1.15** |
|  |  |  |  |  |  |
|  |  |  |  | **ICER** | **£30,364** |

The pooled ICER based on the distribution of histologies expected in clinical practice is £30,364 per QALY. This is marginally higher than the estimate based on the distribution of tumour sites reported in the trial data. The evidence indicates that more common tumour sites such as colon and pancreas which have low frequency of NTRK fusions may be under-represented in the trial population and conversely certain rarer tumour sites with high frequency of NTRK fusions are potentially over-represented (e.g. sarcoma, salivary gland).

This example illustrates the importance of understanding the frequency of histologies expected in the target population and the necessity of modelling histology specific cost and health consequences. When the expected distribution of histologies is expected to differ between the trial and the target population, failure to account for this could result in a biased estimate of the pooled ICER. The magnitude of any bias will depend on the extent of heterogeneity in relevant model inputs between tumour sites.

### Screening to identify eligible patients

The previous analyses did not include the costs of identifying the population of patients with the biomarker of interest. However, a variety of tests and testing strategies may be required to identify eligible patients. As a result, the cost of patient identification may vary significantly across histologies. Indeed, even if homogeneity in all other model inputs is assumed, the cost-effectiveness estimates will inevitably vary based on differences in the costs of identifying patients with the specific biomarker. Consequently, when evaluating the overall cost-effectiveness of a technology it is necessary to consider the joint costs and benefits of the testing/treatment strategy.185, 186

Table 27 updates the previous results by including an arbitrary per patient testing cost of £50. The results clealy demonstrate that even a low per patient testing cost can result in significant variation in the ICER estimates across individual histologes.

Table 27: Including testing costs into histology specific incremental cost-effectiveness ratios (ICERs)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Per-patient level** | | |  |  |  |  |  |
|  |  | **Cost (ex. Testing)** | **QALYs** | **Frequency of mutation** | **Number needed to screen** | **Cost of testing (£50/test)** | **Cost (inc. testing)** | **ICER** |
| Sarcoma | Drug X | £61,314 | 2.70 | 0.56% | 178.57 | £8,929 | £70,243 | £32,765 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Salivary | Drug X | £58,697 | 2.58 | 92.90% | 1.08 | £54 | £58,751 | £28,003 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| IFS | Drug X | £63,332 | 2.79 | 90.90% | 1.10 | £55 | £63,387 | £27,244 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Thyroid | Drug X | £62,615 | 2.76 | 0.92% | 108.70 | £5,435 | £68,049 | £30,402 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Lung | Drug X | £55,032 | 2.41 | 0.09% | 1111.11 | £55,556 | £110,588 | £68,060 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Melanoma | Drug X | £46,963 | 2.03 | 0.21% | 476.19 | £23,810 | £70,773 | £54,178 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Colon | Drug X | £38,667 | 1.65 | 0.12% | 833.33 | £41,667 | £80,334 | £100,326 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| GIST | Drug X | £61,234 | 2.69 | 1.28% | 78.13 | £3,906 | £65,140 | £29,836 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Cholangiocarcinoma | Drug X | £34,261 | 1.45 | 0.10% | 1000.00 | £50,000 | £84,261 | £153,956 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Appendix | Drug X | £37,773 | 1.61 | 4.00% | 25.00 | £1,250 | £39,023 | £39,889 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Breast | Drug X | £37,768 | 1.61 | 0.07% | 1428.57 | £71,429 | £109,196 | £153,952 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |
| Pancreas | Drug X | £37,751 | 1.61 | 0.07% | 1428.57 | £71,429 | £109,180 | £154,304 |
| SoC | £14,471 | 0.99 | - | - | £14,471 | - |

Tumour specific costs of identifying biomarker positive patients are also likely to represent a significant source of heterogeneity due to the variable frequency of targets across tumour types. This is evident in the individual ICER estimates which now show much greater variation across different tumour sites compared to the previous analysis which excluded per patient testing costs.

The key variable driving the testing costs and the ICER estimates for the test/treat strategy is the number needed to screen. For now, we assume that the test is perfect, it correctly classifies all individual as having or not having the mutation, i.e. there are no false positives or false negatives. In this situation the number needed to screen is 1 divided by the expected frequency of the mutation in each histology. For histologies in which the mutation is very common, ‘high frequency histologies’ e.g. salivary, there is a very low number needed to screen as almost every person (92.9%) screened has the mutation. The opposite is the case for ‘low frequency histologies’ in which the mutation is rare. In pancreatic cancer, 1429 people (i.e. 1/0.07%) need to be screened to identify one individual with the mutation. A testing cost of £50 per test increases the overall costs of Drug X by £71,429, from £37,751 to £109,180 in pancreatic cancer. This increases the ICER from £37,930 to £154,304 in pancreatic cancer. The ICER for each histology has increased but in the histologies with moderate to high frequency of mutation, the increase in the ICER is more modest.

The above analysis assumes that the test (or testing strategy) is perfect, however if this is not the case then this will result in patients being misclassified. Patients who do not have the mutation will be classified as having the mutation (false positives) and patients who have the mutation will be missed (false negatives). Such misclassifications may have important implications for costs and health.144, 186 The number of false positives and negatives will depend on the testing strategy, test characteristics (sensitively/specificity) and the frequency of the mutation in each histology.161, 168, 187 The possibility of misclassification presents two tasks: 1) calculating the correct ICER given a specific test or testing strategy; and, 2) choosing the optimal test or testing strategy. Both of these tasks require estimates of the costs and QALYs associated with false positives and false negatives which will likely differ by histology. For costly new treatments and those with significant side effects, false positives may have substantial consequences. The scale of consequences associated with false negatives will depend on the additional benefits of treatment. This means that the consequences of missing a potential patient (false negative) will be larger in those histologies in which the treatment results in larger QALY benefits.

## The value of heterogeneity and population health

The preceding sections show how heterogeneity in treatment effectiveness and testing costs can be explored using pooled ICERs and individual histology ICERs. However, ICERs have an important limitation; they do not give an indication of the scale of consequences for population health. Understanding the benefits and costs of treatment at a population level will help in understanding the consequences of decision making in the presence of heterogeneity and uncertainty.

To understand the implications of heterogeneity for population health requires that benefits and costs are expressed in health or monetary equivalents, using net health benefits (NHBs) or net monetary benefits (NMBs). The same information used to provide ICER estimates can also be expressed as the per-patient NHB (or NMBs) which includes benefits, harms and NHS/Personal Social Services costs.188-190 The NHB is the difference between any health gained with the intervention and the health forgone elsewhere in the health-care system (i.e. due to the need to displace existing treatments and services to fund a new and more costly treatment), all expressed in QALY terms. NMB is equivalent but everything is expressed in monetary terms.

Table 28 illustrates how NHB and NMB are calculated given an assumed threshold of £50,000 per QALY. Testing costs are now included in this analysis.

Table 28: Calculating per person net health effects for Drug X

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subgroup** | **Per patient level** |  | **£50,000 per QALY threshold** |  |  |
|  | **ΔCost** | **ΔQALYs** | **Health foregone (ΔCost/£50,000)** | **NHB (ΔQALY - health foregone)** | **NMB (ΔQALYs x £50,000 – ΔCost)** |
| Sarcoma | £55,772 | 1.70 | 1.12 | 0.59 | £29,337 |
| Salivary gland | £44,281 | 1.58 | 0.89 | 0.70 | £34,784 |
| IFS | £48,916 | 1.80 | 0.98 | 0.82 | £40,859 |
| Thyroid | £53,579 | 1.76 | 1.07 | 0.69 | £34,539 |
| Lung | £96,117 | 1.41 | 1.92 | -0.51 | -£25,505 |
| Melanoma | £56,302 | 1.04 | 1.13 | -0.09 | -£4,342 |
| Colon | £65,863 | 0.66 | 1.32 | -0.66 | -£33,039 |
| GIST | £50,670 | 1.70 | 1.01 | 0.68 | £34,245 |
| Cholangiocarcinoma | £69,791 | 0.45 | 1.40 | -0.94 | -£47,125 |
| Appendix | £24,552 | 0.62 | 0.49 | 0.12 | £6,223 |
| Breast | £94,725 | 0.62 | 1.89 | -1.28 | -£63,961 |
| Pancreas | £94,709 | 0.61 | 1.89 | -1.28 | -£64,020 |

For sarcoma, the additional per patient costs of £55,772 can be represented as 1.12 QALYs (in NMB terms ≈ £55,572/£50,000) in health foregone elsewhere in the health system, based on a NICE threshold of £50,000 per QALY. This can then be compared to the additional benefits of 1.7 QALYs, resulting in an overall positive NHB of approximately 0.59 QALYs (≈1.7-1.12 QALYs) per person treated in this histology. Hence, for each sarcoma patient treated with Drug X, the overall gain to the health system is expected to be 0.59 QALYs per annum. However, for certain other histologies (e.g. colon), the additional health gained with Drug X is more than offset by health forgone elsewhere. This means that for every colon cancer patient who receives Drug X, it is expected that 0.66 QALYs will be lost per annum elsewhere in the health system.

The advantage of NHBs and NMBs is that they can be used to help understand the population level consequences of alternative policy decisions. Understanding the scale of population consequences requires information on the number of patients who are expected to be treated by histology. This will depend on the incidence (number of new cases per year) and prevalence (number of current cases) of the mutation for each histology. It will also depend on the screening strategy used to identify cases and where in the treatment pathway Drug X is used. To simplify the case study, we assume only incident cases and a perfect screening strategy meaning that all patients who can potentially benefit are correctly identified. The expected population health consequences of approving Drug X are shown in Table 29.

Table 29: Calculating population net health effects for Drug X

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subgroup** | **Per patient level £50,000 per QALY threshold** |  | **Population level** |  |
|  | **Health foregone (ΔCost/£50,000)** | **NHB (ΔQALY - health foregone)** | **Incidence** | **NHB, QALYs (NMB, £)** |
| Sarcoma | 1.12 | 0.59 | 5 | 2.88 (£144,046) |
| Salivary gland | 0.89 | 0.70 | 2 | 1.56 (£78,201) |
| IFS | 0.98 | 0.82 | 27 | 22.35 (£1,117,570) |
| Thyroid | 1.07 | 0.69 | 6 | 4.32 (£216,219) |
| Lung | 1.92 | -0.51 | 17 | -8.52 (-£426,230) |
| Melanoma | 1.13 | -0.09 | 3 | -0.23 (-£11,276) |
| Colon | 1.32 | -0.66 | 23 | -15.19 (-£759,374) |
| GIST | 1.01 | 0.68 | 4 | 2.57 (£128,728) |
| Cholangiocarcinoma | 1.40 | -0.94 | 0.3 | -0.31 (-£15,727) |
| Appendix | 0.49 | 0.12 | 16 | 1.99 (£99,383) |
| Breast | 1.89 | -1.28 | 5 | -6.19 (-£309,616) |
| Pancreas | 1.89 | -1.28 | 17 | -21.78 (-£1,089,034) |
| Sum |  |  | 125 | -17 (-£827,110) |

The number of patients with sarcoma who express the biomarker is approximately 5 per year. This means that treating identified sarcoma patients with the Drug X is expected to result in a gain of (5 x 0.59 ≈) 2.88 QALYs per year to the health system when compared to SoC. This contrasts with treating biomarker positive patients with colon or pancreatic cancer. Using Drug X in these populations is expected to result in a loss of 15.19 and 21.78 QALYs respectively per year.

By summing up the yearly NHB across all histologies, Table 29 shows that Drug X is expected to result in an overall loss of approximately 17 QALYs per year. This implies that a histology-independent approval for Drug X is not expected to be cost-effective. Although Drug X appears cost-effective in some individual histologies (e.g. sarcoma and IFS), the overall consequences of approving for all histologies would result in an overall annual loss of health to the health system. The analyses illustrate the importance of information on the relative frequency of histologies expected in the target population.

### Histology-dependent recommendations and the value of heterogeneity

The assessments presented in Table 29 can also be used to compare the population consequences of making different policy recommendations. Decision makers such as NICE have the option of different approval policies:

1. No stratification: Histology-independent approval
2. Partial stratification: Approval in a clinically defined set of histologies
3. Full stratification: Approval only in histologies in which cost-effective is demonstrated

These policies will determine the type of recommendations which are feasible and the relevant health consequences that need to be considered. The following section deals only with approval policies based on expected values without addressing the impact of uncertainty in decision making. Uncertainty, the need for further evidence and alternative mechanisms to reduce the risk of decision making are considered in subsequent sections.

### No stratification: Histology-independent approval

This represents an ‘all or nothing’ approval policy in which the intervention is approved for all histologies or for none. There is no stratification of decision making by histology. In this case the relevant metric is the pooled ICER (or pooled NHB/NMB equivalent) across all histologies. Based on the results of Table 29, Drug X would not be approved for use any histology as the pooled NHB is negative (correspondingly the pooled ICER would be higher than the £50,000 per QALY threshold). The health system is expected to lose approximately 17 QALYs per year if Drug X was granted a histology-independent approval.

However, a further consideration when making histology-independent decisions is that some histologies which may harbour the mutation of interest may not be directly observed in the evidence base at the time of decision making, despite the inclusion/exclusion criteria of the trial permitting their inclusion. As these patients may be treated in clinical practice, consideration should be given to the potential impact of considering histolgies which are not represented in the trial data. The larger the incidence of unrepresented NTRK fusion positive histologies, the greater the influence this can have on decision making. In this case study it is estimated that there are 151 NTRK fusion positive cases in the set of unobserved histologies each year (see Table 16, Section 6.3.4 for details). This is larger than the 125 cases in the observed set of histologies represented in the trial and should be explicitly considered if a histology-independent approval is sought.

If, as in this case study, the economic model is developed around the probability or degree of response in each histology and a BHM has been used to analyse response, then the predictive distribution could be used to estimate response in the unobserved histologies. This assumes that the effects for the unobserved histologies are exchangeable with the observed histologies. For Drug X, the predictive distribution for response in unobserved histologies has a mean response probability of 57% and is highly uncertain with a 95% credible interval from 1% to 100%. If estimates for the remaining parameters (e.g. quality of life, testing costs etc.) can be sourced from the literature or generalised from the observed histologies, then ICER and NHB estimates can also be estimated for unrepresented histologies.

The results in Table 30 include the impact of including unrepresented histologies. To simplify the case study, we collapse all unobserved histologies into one ‘unrepresented’ histology category. The response probability comes from the predictive distribution from the BHM, costs and quality of life are assumed to be the same across all observed and unobserved histologies. The average testing costs for unrepresented histologies were estimated to be £14,322 per patient tested. This estimate was based on observational data on NTRK fusion prevalence (see Appendix 12 for details).

Table 30: Incorporating unrepresented histologies into estimate of population net health effects (NHBs)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subgroup** | **Per patient level £50,000 per QALY threshold** | | **Population level** | |
|  | **Health foregone (ΔCost/£50,000)** | **NHB (ΔQALY - health foregone)** | **Incidence** | **NHB, QALYs (NMB, £)** |
| Sarcoma | 1.12 | 0.59 | 5 | 2.88 (£144,046) |
| Salivary gland | 0.89 | 0.70 | 2 | 1.56 (£78,201) |
| IFS | 0.98 | 0.82 | 27 | 22.35 (£1,117,570) |
| Thyroid | 1.07 | 0.69 | 6 | 4.32 (£216,219) |
| Lung | 1.92 | -0.51 | 17 | -8.52 (-£426,230) |
| Melanoma | 1.13 | -0.09 | 3 | -0.23 (-£11,276) |
| Colon | 1.32 | -0.66 | 23 | -15.19 (-£759,374) |
| GIST | 1.01 | 0.68 | 4 | 2.57 (£128,728) |
| Cholangiocarcinoma | 1.40 | -0.94 | 0.3 | -0.31 (-£15,727) |
| Appendix | 0.49 | 0.12 | 16 | 1.99 (£99,383) |
| Breast | 1.89 | -1.28 | 5 | -6.19 (-£309,616) |
| Pancreas | 1.89 | -1.28 | 17 | -21.78 (-£1,089,034) |
| Unrepresented | 0.97 | 0.15 | 151 | 22.33 (£1,116,748) |
| Sum |  |  | 276 | 5.79 (£289,638) |

After taking account of the unobserved histologies, Drug X is now estimated to be cost-effective in the overall population with positive NHBs. In this example, a histology-independent approval, including an assessment of the potential impact of unrepresented tumour sites, would result in an expected overall gain to the health system of approximately 5.79 QALYs (NMB≈£290,000) per year. Treating individuals with histologies unrepresented in the trial data is expected to result in positive NHB, given the assumptions made here. This is due to the relatively high mean response rate (57%) predicted by the BHM.

Although it may be challenging to identify data to inform benefits in unrepresented histologies, consideration to the magnitude and potential impact of these histologies should be explicitly considered.

### Partial stratification: Approval in a defined set of histologies

This is similar to the previous approval policy; however, in this case the intervention is approved only for a clinically defined set of histologies i.e. there is partial stratification of decision making by histology. The relevant metric here is the pooled ICER (or pooled NHB equivalent) for the defined subset of histologies. The basis for selecting a subset of histologies can be based on theoretical and/or empirical grounds. For example, Section 2.6 highlighted comments from the SAG to EMA for larotrectinib that appeared to differentiate the strength of the biological rationale and the available clinical evidence for several specific tumour types (e.g. IFS, salivary gland/MASC, congenital mesoblastic nephroma and GIST). For these specific tumour types, the SAG concluded that efficacy has been established in the absence of available treatments of proven efficacy in terms of convincing clinical efficacy endpoints and that clinical decisions to use larotrectinib were justified.

To illustrate the implications of a policy decision based on partial stratification, we assume there is sufficient grounds to consider restricting an approval decision for Drug X to only those patients with IFS, salivary gland and GIST. Evidence for patients with congential mesoblastic mephroma was not available at the time of the FDA assessment, hence these patients are not included in the data used to inform the BHM.

Table 31: Decision making with partial stratification

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subgroup** | **Per patient level £50,000 per QALY threshold** |  | **Population level** |  |
|  | **Health foregone (ΔCost/£50,000)** | **NHB, QALYs (NMB, £)** | **Incidence** | **NHB, QALYs (NMB, £)** |
| Salivary gland | 0.89 | 0.70 (£34,800) | 2 | 1.56 (£78,200) |
| IFS | 0.98 | 0.82 (£40,850) | 27 | 22.35 (£1,117,600) |
| GIST | 1.01 | 0.68 (£34,250) | 4 | 2.57 (£128,700) |
| Sum |  |  | 33 | 26.49 (£1,324,500) |

As shown in Table 31, a decision to approve Drug X in only these 3 individual histologies is expected to result in an overall annual gain to the health system of 26.49 QALYs. Although partial stratification results in fewer patients receiving Drug X compared to a full histology independent approval (i.e. 33 patients annually vs 276 patients), there would be an overall gain to the health system from a policy decision based on partial stratification. This gain is equivalent to approximately 20.7 QALYs per annum (i.e. 26.49-5.79 QALYs). In other words, a policy to fully approve a histology independent product could result in an annual loss of 20.7 QALYs to the health system compared to an optimised approval decision based on a partial stratification approach.

The majority of the gains from partial stratification are achieved by avoiding the approval of Drug X in histologies with high testing costs and relatively large incidence (e.g. lung, colon, breast and pancreatic cancer) for which Drug X does not appear to be cost-effective based on current evidene. A further advantage of partial stratification over no stratification is that assumptions about unrepresented histologies can be avoided in decision making. However, a disadvantage of partial stratification is a potential increase in monitoring costs required to prevent the use of Drug X outside of its subset of approved histologies.146

### Full stratification: Approval only in histologies in which cost-effectiveness is demonstrated

This is a fully histology-dependent approval policy in which the technology is restricted for use only in those histologies in which it has been shown to be potentially cost-effective based on expected ICER/NHB estimates. Given the ICER/NHB estimates presented in Table 29, Drug X appears potentially cost-effective in the following histologies: sarcoma; salivary gland; IFS; thyroid; GIST and appendix. These are the histologies in which NHBs are greater than zero. Equivalently they each have ICERs below £50,000 per QALY gained. Taking the sum of the NHB across each of these histologies results in an overall annual gain of 35.68 QALYs to the health system from a fully stratified approval decision for Drug X. The expected number of patients treated annually based on full stratification is estimated to be 60 patients.

The additional value of distinguishing between different types of patients represents the value of heterogeneity.145-147 In this example, the value of heterogeneity represents the difference between the NHB of a fully stratified recommendation compared to a histology-independent recommendation with no stratification. This difference is equivalent to 29.89 QALYs (≈35.68 - 5.79 QALYs) per year.

Exploring the value of heterogeneity may help to inform NICE committees of the consequences of alternative policy options, both in terms of the expected number of patients that would be eligible to receive a specific new treatment but also in terms of their overall consequences to the health system. While a histology-independent approval might be considered appropriate on the basis that this results in an overall positive annual NHB compared to rejecting the technology, it is also important to consider the potential consequences of such an approval policy compared to a more restrictive or optimised recommendation. In this case study, there appear to be significant gains to the health system that could be achieved by an optimised recommendation. Importantly, an approval decision based on partial stratification using only 3 individual histologies appears to confer approximately 74% of the gains that are potentially achieved based on a full stratification policy.

## Uncertainty and decision making

The previous sections have discussed decisions about the approval of technologies. However, decision makers such as NICE also need to consider the risk associated with decision making under uncertainty. Given the limitations in study design and sample size, there will always be uncertainties about the cost and health consequences associated with different treatment options. All ICERs and NHBs discussed previously will be associated with uncertainty. This means that through the central estimate of the ICER/NHB indicates that a treatment is cost-effective, there is also a risk that the treatment is not cost-effective. For example, a treatment which meets the end of life criteria may have a central ICER estimate of £45,000 per QALY and so is expected to be cost-effective. However, due to uncertainty there may be a 40% chance that the true ICER is above £50,000 per QALY. The health losses associated with this eventuality is the risk of decision making under uncertainty.

Uncertainties can be divided into two categories; those that arise from assumptions inherent in constructing models (structural uncertainties) and those which are a result of imprecision in parameter estimates due to limited sample size (parameter uncertainties). Previous research has shown how uncertainties associated with imprecision can be addressed through further data collection or pricing schemes.191, 192 In this section we will show how these approaches can be used to reduce risk in decision making for histology-independent technologies, in addition we show how stratified decision making represents an additional approach to managing risk associated with uncertainty.

### The consequences of uncertainty

This section introduces value of information (VOI) as a framework to quantify the health effects of uncertainty. VOI analyses can provide decision makers with metrics to help understand the drivers of decision uncertainty and assess alternative strategies which could be used to manage this risk. The uncertainty associated with a histology-independent decision (‘no stratification’) is illustrated below. The implications for partial and full stratification will be addressed in subsequent sections.

The NHB results previously reported in Table 30 are illustrated graphically in Figure 11. In addition, uncertainty around the expected (mean) estimates of NHB are represented using a 95% confidence interval. This is computed from the mean and 95% percentiles of the PSA for each histology. Figure 11 also plots the patient level pooled NHB. This is analogous to the pooled ICER reported in Table 25 and Table 26. The pooled NHB is a weighted average of the NHB associated with different histologies. Weights come from the incidence of NTRK positive histologies reported in Table 30 with a perfect screening strategy assumed.

Figure 11: Net health benefits per person across histologies with uncertainty



The figure shows that the Drug X is expected to result in additional NHB in sarcoma, salivary, IFS, thyroid and GIST with the 95% interval not crossing the line of equivalence with SoC. It is also expected to be cost-effective in appendix and in those histologies that are unrepresented in the trial, but this is uncertain. This uncertainty can be expressed in terms of the likelihood or probability that Drug X is not cost-effective compared to the SoC (i.e. 47% and 39% in appendix and unrepresented cancers, respectively). For lung, melanoma, colon, cholangiocarcinoma, breast and pancreatic cancer the model estimates that there is approximately 0% chance that Drug X is cost-effective.

As in the previous analysis, the pooled population represents the expected consequences of a histology-independent recommendation. Drug X is expected to result in 0.02 additional QALYs per person treated and there is a 52% chance that it is cost-effective compared to SoC. The pooled estimate relies on the relative incidence of histologies in the target population. Although uncertainty in histology incidence is not addressed quantitatively in this case study, this can be propagated through the PSA in the same manner as other uncertainties.

The health system consequences of decision making can be better informed with reference to population health. Figure 12 shows the population NHB for each histology and the pooled NHB. This is calculated by multiplying the per-person NHB for each group illustrated in Figure 11 by the incidence for each group (see Table 30).

Figure 12: Population net health effects across histologies with uncertainty



Figure 12 shows that although uncertainty in per person NHB may be similar across histologies, the consequences of approval and uncertainty vary substantially when the size of populations are taken into account. The figure shows that for many histologies (e.g. sarcoma, salivary, melanoma, breast etc.), the health consequences of approval and/or uncertainty are limited due to their small population. By contrast, the health consequences associated with the unrepresented histologies are relatively large as decisions in this group affect 151 individuals each year.

The pooled category represents the health consequences of the ‘no stratification’ approval policy i.e. a histology-independent approval. A histology-independent approval is expected to result in a gain of 5.79 QALYs per year on average (consistent with Table 30). However, the 95% interval indicates that this is highly uncertain, with approval potentially resulting in losses up to 120 QALYs per year (illustrated in Figure 12 by the lower confidence interval which extends to -120). The next section will describe how VOI methods can be used to quantify the health consequences of this uncertainty to help inform decision making and approval policies.

### Quantifying the health consequences of uncertainty

Histology-independent decision making (‘no stratification) is concerned with making approval decisions based on pooled cost-effectiveness estimates. From Figure 12, the Drug X is expected to provide an expected benefit of (0.02 x 276 ≈) 5.79 QALYs per year at the pooled population level. However, there is uncertainty about this benefit. VOI methods can be used to quantify the health consequences of uncertainty, i.e. the risk associated with decision making with current information.189, 193, 194 Uncertainty matters because it means that there is a chance of making the wrong decision. Quantifying the expected health consequences of uncertainty is achieved by multiplying the chance of making a wrong decision by the health consequences of making the wrong decision. This is illustrated in Figure 13.

Figure 13: Estimating the health consequences of uncertainty



If Drug X is more cost-effective than SoC in the pooled population then there are zero health consequences of uncertainty. The tall left-hand bar in Figure 13 shows that there is estimated to be a 52% chance that the Drug X is cost-effective in the pooled population. This corresponds a 52% chance of zero consequences of uncertainty.

Making an incorrect decision (e.g. approving the Drug X when it is not cost-effective) will have health consequences. For the pooled population there is a 48% chance that the decision to approve Drug X is incorrect. As shown in Figure 13, these health consequences are not uniform. There is a greater chance of more limited consequences compared to a smaller chance of greater consequences. Figure 13 shows there is a 21% chance of Drug X resulting in a loss of 25 QALYs per year (second bar from the left). There is a 19% chance of a loss of 75 QALYs (third bar from the left) and so on. The weighted average over this range of outcomes provides an estimate of the health consequences of uncertainty. This is estimated to be 29.52 QALYs per year, equivalent to approximately 0.108 QALYs per person.

This quantitative approach to the risks associated with uncertainty can be used to assess policy options which address this risk. In the following sections we will illustrate three approaches to managing this risk: further data collection, pricing agreements and stratified decision making.

### Managing risk through further data collection

Further data collection is one approach to reduce risk associated with uncertainty. Imprecision in parameter estimates due to limited sample size (e.g. overall survival) can be reduced by collecting data on these parameters.195

Decisions about further data collection are important because under current policy arrangements when NICE is unable to approve a technology for routine use due to parameter uncertainties it may recommend it for inclusion into the CDF if it is eligible.196 Topics that are eligible for the CDF are reimbursed for a time-limited duration following the development of a managed access agreement (MAA). The MAA consists of 1) a data collection agreement (DCA), which specifies the data that must be collected that could sufficiently resolve the parameter uncertainties identified by the appraisal committee and 2) a commercial access agreement that ensures the technology is reimbursed at a cost-effective price during the period of the MAA. A technology remains in the CDF until the data collection agreed in the DCA is complete; it then proceeds to reappraisal and exits the fund.

The MAA covers the entire eligible population determined by the NICE guidance which means that entry to the CDF is equivalent to an ‘approval with research’ decision which is reassessed after the data collection period (usually two years).196, 197 The assessments required to inform the suitability of an ‘approval with research’ decision over ‘only in research’, approve and reject are covered in detail elsewhere.194 Explicit consideration of these assessments could aid the transparency of CDF entry requirements. However, it is beyond the scope of this report to suggest reforms to CDF processes or determine the appropriate size of the CDF budget. These issues have been commented on elsewhere and require further research.197-199

The aim here is to provide a framework to understand how the CDF, in its current form, can help to address the risk associated with histology-independent technologies. The intention is to demonstrate how a unified decision framework could enable CDF data collection arrangements to be considered alongside other risk reduction strategies (e.g. pricing schemes and stratified decision making).

**Decision uncertainty resolved by the Cancer Drugs Fund (CDF)**

Previously, the value of resolving all uncertainty was estimated to be 29.52 QALYs per year (NMB≈£1.48 million), therefore this is an upper bound for the risk which can be resolved through further research each year. However, there are many sources of uncertainty in any model; for example, uncertainties in baseline risks, health state costs, health-related quality of life etc. Different types of research will potentially be required to inform different model parameters. For example, observational survey research may be sufficient to address uncertainties about health-related quality of life in specific disease states, whereas randomised research may be required to resolve uncertainties in the relative effects of interventions.195 Research on particular parameters will resolve more or less uncertainty depending on how central they are to the decision between the treatment alternatives. This means that research on some parameters is more valuable than on others.

The upper bound for the value of additional research on specific parameters (or set of parameters) can be calculated using an extension of VOI methods. These are called expected value of partial perfect information (EVPPI) methods.194, 200 In order to estimate the value of resolving uncertainty in a specific parameter, the EVPPI method estimates the payoff (in QALYs or £) from clinical decision if the parameter of interest was known with certainty compared to the payoff if that parameter remained uncertain. The difference between these two scenarios is the EVPPI. This decomposes the overall upper bound for the value of research into the value of resolving uncertainty in specific parameters (or sets of parameters).

To illustrate EVPPI analysis using the case study, consider the case in which only information on overall survival (for responders and non-responders) could be collected through CDF arrangements. This may be due to organisational or time constraints. Estimating EVPPI using the Gaussian Process method suggested by Strong and colleagues,200 the upper bound for the value of research on overall survival is 12.16 QALYs (NMB≈£0.6 million) per year. This can be compared to 29.52 QALYs per year (NMB≈£1.48 million) which is the value of resolving the uncertainty associated with all parameters in the model. EVPPI methods provide a more accurate assessment of the risk which can be resolved with particular data collection strategies. This same approach can be applied to any uncertainties which are parameterised in a decision model. As shown earlier, the distribution of histologies in practice can influence the cost-effectiveness of Drug X when making histology-independent recommendations. If uncertainty about the distribution of histologies can be parameterised, then EVPPI methods can be used to understand the value of research to resolve these uncertainties.

These methods can be used to help prioritise data collection. Although the CDF financial resource constraint is softened by the expenditure control mechanism, the real resources required to coordinate and quality control data collection are limited.196 In the case where high quality data on certain parameters is challenging to collect through the CDF, EVPPI methods can be used to understand the risk which can be resolved by collecting data on these parameters. This can be used to determine: 1) whether there is any value in collecting data on a specific parameter; 2) whether the benefits of the additional information is sufficient to justify the additional costs of collecting the data; 3) whether other approaches such as pricing schemes or stratification would be more appropriate to resolve the decision risks.

EVPPI methods are an important extension to VOI analysis in decision making. However, EVPPI estimates are still upper bounds for the value of additional research on individual parameters. This is because EVPPI assumes that uncertainty in the parameter of interest is completely resolved, i.e. it is the value of research if an infinite sample size was collected. Expected value of sample information (EVSI) methods relax this assumption by assessing the value of commissioning research with finite sample sizes.195, 201, 202

### Managing risk through pricing schemes

The NICE process allows for consideration of a variety of pricing schemes, including: patient access schemes, commercial access agreements and flexible pricing.203 These schemes can facilitate pricing arrangements such as simple discounts or more complex ‘pay-for-performance’ arrangements. In this section we illustrate the effect of a simple discount and a pay-for-performance scheme on uncertainty and the expected value of a technology.

Simple discounts have been identified in previous research as an effective approach for payers to reduce the risk of approving technologies which are not cost-effective.191, 192, 197 Reducing price of a technology which is expected to be cost-effective has two implications: 1) the value of implementing the technology will increase due to the resources saved; 2) the risk of the technology not being cost-effective will decrease. Under the current price (£1,250 per month), a histology-independent recommendation for Drug X is expected to result in 5.79 additional QALYs per year. However, due to uncertainty in parameters there is also a 48% chance that Drug X is not cost-effective. As shown previously, the expected health consequences of this uncertainty have been estimated to be 29.52 QALYs per year.

With a 10% simple discount (£1,125 per month), the expected value of Drug X is estimated to increase to 21.4 QALYs per year. Furthermore, the risk of Drug X not being cost-effective is reduced to 42%. The potentially negative health consequences associated with the uncertainty are reduced to 23.91 QALYs per year (a reduction in risk of 5.61 QALYs). A 20% simple discount (£1,000 per month) increases the expected value to 30.04 QALYs per year and reduces the consequences of uncertainty to 21.3 QALYs per year (a reduction in risk of 8.22 QALYs).

To illustrate the use of more complex pricing schemes, we also implemented a pay-for-performance scheme. In this scenario the undiscounted cost of Drug X (£1,250 per month) is only incurred if a patient responds to treatment. This has two impacts on risk and cost-effectiveness. First, this acts in a similar manner to a price discount because the average response is expected to be approximately 60% according to the BHM (i.e. reducing the effective price by 40%). Second, the risk associated with Drug X not resulting in the expected outcomes is now shifted from the payer to the company.192 These two impacts reduce the health consequences of uncertainty to the health system. With this pricing scheme, the expected value of Drug X increases to 51.95 QALYs per year and reduces the consequences of uncertainty (i.e. the expected risk) to 9.35 QALYs per year (a reduction in risk of 20.17 QALYs).

The examples here illustrate how alternative pricing approaches can be used to increase the expected value of a technology as well as impacting the risk and consequences associated with uncertainty.

### Managing risk through stratified decision making

The previous discussion described how uncertainty can be addressed when making histology-independent decisions in which the technology is expected to be used across all histologies (no stratification). In this section we discuss how to apply these same principles under partial and full stratification.

##### Partial stratification

The previous sections described how uncertainty and associated risk could be managed through further data collection and pricing schemes. Both assumed that the health technology would be approved for use in all histologies or none, i.e. histology-independent approval decisions. In this section we discuss stratification as an additional approach to reducing risk in approval decisions for products with a histology-independent marketing authorisation.

In the case of partial stratification the intervention is approved only for a clinically defined set of histologies. The relevant metric for decision making is the pooled NHB for the subset of histologies of interest. Therefore, it is the uncertainty in this pooled NHB that is of relevance to decision making. Figure 14 illustrates the uncertainty in pooled NHB for the approval of Drug X in IFS, salivary gland and GIST only.

Figure 14: Population net health effects with uncertainty for partially stratified decision making



Figure 14 shows that Drug X is expected to provide positive NHB in each of IFS, salivary gland and GIST individually. Implementing Drug X in this subset is expected to result in approximately 26.5 additional QALYs per year over SoC (this corresponds to Table 33). Figure 14 graphically represents the uncertainty in this estimate. The 95% interval for the pooled effect is far from the line of equivalence between Drug X and SoC indicating that there is not much uncertainty in cost-effectiveness in this subset of histologies. Given the model assumptions, it is estimated that there is now a 0% risk that Drug X is not cost-effective. This means that the risk in approving Drug X has been eliminated, without the need to carry our additional research or wait for research to report. As such, a routine commissioning decision maybe considered appropriate for this specific subset of tumour sites.

##### Full stratification

Under full stratification there is the option to make different decisions for different histologies. This is a fully histology-dependent approval policy in which the technology is restricted for use only in those histologies in which it has been shown to be cost-effective. Figure 15 illustrates the population level uncertainties in making fully stratified recommendations.

Figure 15: Population net health effects with uncertainty for fully stratified decision making



Approval in sarcoma, salivary gland, IFS, thyroid, GIST and appendix are all expected to provide positive NHB. Approval of Drug X in the remaining histologies is expected to result in a loss of population health. The figure also shows that uncertainty about health benefits (or losses) differ across histologies. The 95% uncertainty bounds cross the line of equivalence for melanoma and appendix only.

Because separate approval decisions can be made for each histology, the risk associated with decision making should be estimated for each histology separately. The risk associated with uncertainty in each histology is reported in Table 32 along with the expected annual value of approving the treatment for each histology.

Table 32: Benefits of approval and further research for fully stratified decision making

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Population level £50,000 per QALY threshold** | | |
| **Subgroup** | **Incidence** | **Health impact of uncertainty per year, QALYs (NMB, £)** | **Health impact of approval per year, QALYs (NMB, £)** |
| Sarcoma | 5 | 0.02 (£1,216) | 2.88 (£144,046) |
| Salivary gland | 2 | 0 (£113) | 1.56 (£78,201) |
| IFS | 27 | 0.01 (£577) | 22.35 (£1,117,570) |
| Thyroid | 6 | 0.01 (£665) | 4.32 (£216,219) |
| Lung | 17 | 0.03 (£1,398) | -8.52 (-£426,230) |
| Melanoma | 3 | 0.23 (£9,950) | -0.23 (-£11,276) |
| Colon | 23 | 0.02 (£1,045) | -15.19 (-£759,374) |
| GIST | 4 | 0.01 (£407) | 2.57 (£128,728) |
| Cholangiocarcinoma | 0 | 0 (£2) | -0.31 (-£15,727) |
| Appendix | 16 | 1.1 (£5,5078) | 1.99 (£99,383) |
| Breast | 5 | 0 (£0) | -6.19 (-£309,616) |
| Pancreas | 17 | 0 (£0) | -21.78 (£1,089,034) |
| Sum | 125 | 1.41 (£70,452) |  |

Table 32 shows that, of the histologies included, the largest risks are associated with decisions about appendix and melanoma. This is because these two histologies have uncertainty bounds in Figure 15 which cross the line of equivalence. Since further research appears most valuable in these histologies, this may help prioritise further data collection. EVPPI assessments can be applied to these histologies to understand the parameters which are driving uncertainty. For other histologies, it is clear that Drug X is cost-effective (e.g. IFS) or it is not cost-effective (e.g. pancreas) based on current evidence and so there is appears limited risk of making the wrong decision and so little value in further research.

Table 32 also illustrates that the total risk associated with decision making has reduced for stratified decision making compared to no stratification. The expected health consequences of uncertainty were 29.52 QALYs per year for no stratification. This was zero for partial stratification and approximately 1.41 QALYs per year for full stratification (implying a reduction of 28.11 QALYs).

The change in uncertainty (as measured by VOI) from less stratification to more stratification has been called the ‘dynamic value of heterogeneity’ in the literature.146 It should be noted that increasing stratification may increase or decrease the uncertainty in decision making. When the characteristic that treatment is being stratified by is important in explaining heterogeneity (such as histology in the case study), then stratification will increase the value of implementing the treatment while reducing the risk of making an incorrect decision. This is because variability in outcomes is translated into heterogeneity. However, if a stratification characteristic contains little information to distinguish outcomes then uncertainty may increase with stratification due to sample splitting.146

### Comparing approaches to risk management in histology-independent technologies

When making a histology independent approval decision, with current evidence and without any discount, Drug X appeared cost-effective based on expected values but the health consequences of uncertainty were estimated to be 29.52 QALYs per year. Three approaches to risk management were explored: further data collection, pricing schemes and stratified decision making.

The upper bound for value of further data collection on overall survival was expected to be 12.16 QALYs, this is compared to a reduction in risk of 5.61 QALYs from a 10% price discount, 8.22 QALYs from a 20% discount, 20.17 QALYs from a pay-for-performance scheme and a reduction of 28.11 QALYs from stratification.

The magnitude of uncertainty resolved through data collection will depend on which parameters can be informed by feasible research. Due to institutional or ethical constraints, data collection may not be possible for some parameters and this places limits on this approach to risk management.

As discussed previously, the degree of uncertainty resolved through stratifying by histology will depend on the importance of histologies in explaining heterogeneity in cost-effectiveness. In cases in which cost-effectiveness does not vary significantly across histologies, the risk reduction from stratification will be lower. There may also be additional costs associated with (partially or fully) stratified recommendations, for example, the costs of monitoring clinician behaviour to ensure that treatments are not being used in histologies for which they do not have approval.145, 146 In principle this cost can be incorporated into the analysis of alternative policy options, however reliable data to predict these costs may be difficult to find.204

When considering the impact of pricing schemes, the magnitude of risk reduction will depend on the pricing arrangement.192 For simple discounts, the risk reduction will increase with the scale of the price reduction. Neither a 10% nor a 20% simple discount reduced the risk of approval as much as either further data collection or stratified decision making in the case-study. A pay-for-performance scheme reduced the risk more than further data collection but not as much as stratified decision making. It should be noted that, when comparing price reductions (or stratification) to data collection, it is important to take account of the fact that data collection takes time to report whereas the other risk management policies can theoretically begin immediately.

Pricing schemes and stratified decision making can also increase the value of approving technologies in addition to addressing risk. A 10% and 20% discount increased the value of Drug X from 5.79 to 21.4 and 30.04 additional QALYs per year respectively. Partial and full stratification increased this to 26.49 and 35.68 QALYs per year respectively. The gain in value from stratification is a result of making more optimised decisions. This has been called the ‘static value of heterogeneity’ in the literature.146 The pay-for-performance scheme increased the potential value of approval by the greatest extent. It resulted in an additional 51.95 QALYs per year from approval of Drug X. This gain is due mostly to the substantial discount implied by the pay-for-performance scheme. Because these approaches to risk management (further data collection, pricing schemes and stratified decision making) are not mutually exclusive, each one can be used in combination to address the risk of approving a technology which is not cost-effective.191

## Discussion

We have illustrated a framework for decision making which takes account of uncertainty and heterogeneity associated with histology-independent technologies. The aim was to outline assessments which can help support NICE and CDF decision making, both in making approval decisions and in managing risks associated with uncertainties.

It is evident that heterogeneity in the cost-effectiveness of histology-independent technologies can arise from a number of sources and that these should be explicitly considered when making decisions. Even if clinical outcomes were identical across individual histologies, differences in the costs of identification can lead to important cost-effectiveness differences between individual histologies. In situations in which the target population is expected to differ from the trial population (i.e.in terms of the distribution of histology types), then explicit modelling of heterogeneity will be required to support NICE decision making. If there exist any histologies which are unrepresented in the trial population then consideration will be required to the potential costs and health consequences in unrepresented histologies along with their frequency in the target population to support a histology-independent approval.

The framework explored the health consequences associated with three different approval policies: no stratification (histology-independent approval), partial stratification and full stratification. This demonstrated the potential health gains from making stratified decisions. As discussed above, modelling the costs and health consequences associated with heterogeneity will often be required to make histology-independent decisions. This means that the assessments and assumptions required for stratified decision making will often be the same as those required for histology-independent decision making. Furthermore, because partially and fully stratified decision making allows for approval only in the subset of histologies for which there is observed data, these stratified approaches can be less dependent on strong assumptions. This is because they avoid the requirements to estimate ICERs/NHBs for unrepresented histologies.

The role of stratified decision making was also illustrated as an approach to reducing the risk associated with uncertainty. This was compared to two other approaches to risk management: further data collection and pricing schemes.191, 193, 201 This analysis showed that each approach can reduce risk associated with uncertainty. Stratified decision making was shown to be the most effective policy for risk reduction in the case study. The factors which determine the magnitude of uncertainty resolved by each approach were discussed and it was highlighted that these factors will differ across histology-independent technologies. The policy, or combination of policies, chosen in a specific scenario will depend on procedural feasibility and the characteristics of a given proposal.

### Limitations of the analysis and directions for future research

A limitation of the analysis in this section is that ‘unrepresented histologies’ are included as a homogenous group. In reality there may be significant heterogeneity between different unrepresented histologies. The sections on stratified decision making assumed that Drug X could only be approved in represented treatments. Theoretically, this need not be the case. If unrepresented histologies were not treated as a homogenous group but were considered individually it is likely that for some Drug X would be expected to be cost-effective and for others it would not. The uncertainty surrounding each would also differ. If approval in individual unrepresented histologies was feasible, the decision uncertainty remaining after full stratification would be larger than reported in Section 7.4.5. This is because the uncertainty reported for fully stratified decision making (1.41 QALYs per year) only considers uncertainty in the represented histologies. Including uncertainty in unrepresented histologies would necessarily increase this.

For the sake of clearly illustrating the core principles of decision making under uncertainty other simplifying assumptions were made. Namely, one-off infrastructure costs, population prevalence and test uncertainty were not explicitly modelled, these assumptions should be relaxed in future research.194

One-off infrastructure costs are relevant in calculating per-person testing costs. In the case study we have assumed a one off testing cost of £50 per individual tested. However, testing approaches based on NGS may require large up-front investments in infrastructure. A recommended approach to incorporate capital costs, such as testing infrastructure, is to divide the one-off expenditure by the total population of patients who are expected to use the infrastructure.194, 205 For histology independent technologies this includes individuals across a range of histologies and over the expected lifetime for the infrastructure. This has several potential important implications for decision making.

The first is that any stratification of approval by histology will necessarily mean that testing costs will be spread over a smaller number of patients. This will have the effect of increasing per-person testing costs when treatments are approved for subsets of histologies, reducing the expected health gains associated with stratification. Second, if reimbursement decisions are changed before the end of the assumed lifetime of the one-off infrastructure investment and some proportion of these costs are not recoverable then this has important implications for decision making under uncertainty. The presence of significant irrecoverable costs increases the costs associated with initially implementing then subsequently removing a technology from general use. Taking account of these costs will tend to favour more conservative approaches to decision making which demand less uncertainty before a treatment is approved for widespread use.194 This has implications for the CDF as MAAs stipulate approval of technologies for the entire eligible population determined by the NICE guidance alongside research.196 Explicit consideration of significant irrecoverable costs in this context will make the costs of inclusion into the CDF more transparent.

A third implication of investment costs is that testing infrastructure such as NGS may provide a basis for the use of other health technologies which use the same infrastructure. This means that the population of patients who are expected to use the infrastructure extends across all treatments and indications expected to use the infrastructure.

A further simplification of the case study was that it was assumed that the test used to identify eligible patients was perfect i.e. it results in zero false positives and zero false negatives. The reality of testing will differ in two ways: 1) the accuracy of a test may not be perfect and will therefore misclassify a certain proportion of patients; 2) the false positive and false negative rate will be estimated with uncertainty, meaning that the rate of misclassification may not be known with certainty. For point 1, the consequences of misclassification and the analytical approaches to deal with it have been discussed in Section 7.2.1. For the second point, if uncertainty in false positive and false negative rate can be parameterised then the health consequences of this uncertainty can be managed using the same EVPPI methods as illustrated in the case study.

The case study was also built upon a simplified surrogate relationship between response and survival. Survival was assumed to be determined by response and conditional on response or non response it was assumed to be homogenous across histologies. The aim of this model was to link heterogeneity in response to heterogeneity in costs and health outcomes. However, as discussed in Section 4, the relationship between response and survival is highly uncertain, variable and may be very weak. Further research is required to better inform how surrogate outcomes, such as response, can be linked to costs and health outcomes.

Section 7.4.6 compared data collection, pricing schemes and stratified decision making as alternative approaches to manage risk and increase the health impact of decision making. Considering the full range of options has important implications for price negotiations in histology-independent technologies. The health impacts of stratified decision making could be used as a benchmark in negotiating discounts required for histology-independent approval. For example, in order to obtain a histology-independent approval the reimbursement decision maker could require a pricing scheme sufficient to reduce risk to the level that would exisit under stratified decision making. Any approval policy will create a specific set of incentives for research and pricing strategy.206, 207 Further research is required to understand the incentives provided by current arrangements and the potential benefits of changes to policy.

Finally, the case-study focused on histology as the the main source of heterogeneity. However, heterogeneity could be explored using a range of alternative charcteristics 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 utilised and combined in different ways in decision making.146 How best to decide on which characteristics to utilise in decision making and how they should interact is a complex question which requires further research.

# Recommendations for practice and further research

Drawing on the research findings, recommendations are provided relating to three distinct areas:

1. The types of analysis and evidence required to inform decisions regarding histology-independent drugs by NICE.

2. Potential changes to the NICE methods guide for technology appraisals or additional requirements relating to histology-independent drugs.

3. Priorities for methodological research.

## Types of analyses and evidence required to inform decisions regarding histology-independent products

##### Treatment effectiveness

Complex innovative study designs are increasingly used to improve the efficiency of the drug development process and to speed up regulatory approval and access of drugs with new mechanisms of action. Adaptive basket trials are particularly suited to assess efficacy of histology-independent drugs, although their reliance on surrogate outcomes, small sample sizes and mostly uncontrolled designs pose challenges for HTA. Adequately designed and analysed basket studies which assess the homogeneity of outcomes and allow borrowing of information across baskets where appropriate, are recommended. In particular the use of comparative and randomised designs and primary outcomes that can adequately predict the clinical outcomes of interest is recommended where feasible.

The potential for heterogeneity in treatment effects, either across tumour types or across other characteristics, is likely to be an important issue in the appraisal of histology independent technologies. Careful consideration should be given to the appropriateness of the assumptions of homogeneity of treatment effects and NICE committees should expect to see an exploration of this assumption in company submissions. Bayesian hierarchical methods, which are frequently used in the analysis of basket trials, may provide a useful vehicle with which to explore any heterogeneity. Where there is evidence of heterogeneity in treatment effects and estimates of cost-effectiveness, consideration should be given to optimised recommendations.

##### Counterfactual

Generating a counterfactual is likely to be challenging in the context of histology independent technologies and, in the absence of randomised evidence, it is likely that no single approach will be able to provide robust estimates of relative effectiveness. Companies developing histology independent technologies therefore should be encouraged to consider several alternatives. Consideration should be given to the relative strengths and weaknesses of these alternatives when evaluating the most appropriate comparison. Evidence on the prognostic and predicative performance of the biomarkers should also be considered where possible, though it is recognised that such data may be limited at the time of submission.

##### Generalisability

Trial evidence available to support approval of histology independent technologies may differ substantially from the patients eligible for treatment in practice. Significant differences may for example be seen in the distribution of tumour types, positioning of the technology and subsequent treatments received. The potential for heterogeneity in treatment effects means that differences between the trial population and the eligible population may have an important impact on estimates of cost-effectiveness. Where possible, it is important that such differences are properly accounted for. Consideration of the differences between the trial population and eligible population should also be bourne in mind when considering an appropriate counterfactual data set.

Trial evidence supporting histology independent technologies may not offer complete coverage of the eligible population. As such, there may be no effectiveness evidence supporting a proportion of the eligible population. Appropriate consideration should be given to these unrepresented tumour types in the appraisal of histology independent technologies. BHM, may be able to provide an estimate of the distribution of treatment effects in this population. Data collection plans, where considered appropriate, should consider the potential for collecting evidence in unpresented tumours to better inform estimates of effect. Consideration should also be given to the fact that unrepresented tumours are not a single tumour type and may be heterogeneous. As such, blanket approval or collection of data in unrepresented tumours may not be appropriate.

##### Genomic testing

Genomic testing is likely to be integral to identify patients eligible for histology-independent therapy. Genomic testing costs may vary substantially across tumour types and therefore represent an important potential source of heterogeneity that should be appropriately considered. It is possible that some tumour types will not be cost-effective on the basis of genomic testing costs alone. Current NICE guidance provides that testing should be included where they are necessary to support a new healthcare technology. Investment in universal provision of genomic testing, however, generates challenges to this model as some testing strategies may be used to identify multiple potential targets. In principle, it may therefore be appropriate to apportion testing costs over several technologies. It is currently unclear how this should be done or who should make such judgments.

##### Model structure

Alternative sources of heterogeneity that may impact cost-effectiveness estimates (e.g. baseline risk, treatment effect, costs and HRQoL) should be explicitly acknowledged and appropriately reflected in any economic model. Where an economic analysis is developed using a partitioned survival analysis approach based on the direct extrapolation of TTE endpoints, appropriate exploration of the validity of pooling PFS and OS across pre-specified subgroups and histologies (where data permits) should be undertaken (e.g. separate presentation of Kaplan-Meier curves and landmark PFS and OS rates). The process of internal and external validation should be clearly described.

BHM approaches may be useful to support the validity of pooling PFS and OS data. Where there is substantive evidence of heterogeneity in treatment effects, consideration should be given to alternative model structures that are better able to reflect this heterogeneity including the use of landmark response approaches. If such a model is used, evidence supporting the proposed surrogate-final relationship should be presented and uncertainty surrounding the surrogate relationships included in the model should be fully characterised. While concerns remain regarding the validity of response as a surrogate for PFS and OS, a surrogate-based modelling approach informed by predictions from meta-analyses which capture all relevant uncertainty may be preferable to the extrapolation of heavily censored and potentially confounded PFS and OS data. The BRMA approach outlined in DSU TSD 20142 is recommended to ensure that all uncertainty around the surrogate relationship is reflected in the predictions used in the model.

##### Consideration of uncertainty

Uncertainty is inherent to all decision made by NICE and other reimbursement agencies, but may be particularly acute when considering histology independent technologies due to the limitations of the underlying evidence base. When considering the implications of uncertainty, due consideration should be given to the scale and consequence of decisions, as often populations may be small with limited consequences to the overall health system. Quantification of the consequences of uncertainty using NHB may provide an important framework to add to the assessments already routinely specified with the exiting TA methods guide to quantify decision uncertainty. Routine presentation of such metrics should be considered.

## Potential changes to the NICE methods guide for technology appraisals and/or additional requirements relating to histology-independent drugs.

In practice there may be barriers to NICE making partially or fully stratified decisions. This is because the NICE TA process has been developed primarily to make approval decisions for a technology in a defined population. However, stratified decision making can be considered as a subgroup analysis in which histology is the relevant source of heterogeneity. The NICE methods guide recognises that costs and the capacity to benefit may differ across patients with differing characteristics and recommends that this should be explored as part of the reference case.4 The assessments outlined in our report are consistent with and should be supplemented by the existing NICE guidance on subgroup analysis.

The quantity of subgroups which result from fully stratified decision making could present a challenge to implementing this approach in practice. Partial stratification of approval decisions is one approach to address this. Partial stratification would reduce the number of approval decisions which must be made, compared to full stratification. A transparent and accountable process for deciding which histologies should be grouped together would be required under this approach. The process for deciding which histologies should be grouped together could be usefully informed by the criteria for defining subgroups in the NICE methods guide. According to the current process, subgroups should be based on the expectation of *“differential clinical or cost-effectiveness, biologically plausible mechanisms, social characteristics or other clearly justified factors”.*4 The relevant subgroups should be defined at the scoping stage but with the possibility of subgroup identification later in the process. This same process may be appropriate to define which histologies should be grouped together to make partially stratified approval decisions.

A further issue for NICE methods and process concerns the approval of a histology independent treatment in histologies not included in the main clinical studies. This could be considered as a specific case of a more general problem concerning the approval of treatments in populations for which there is limited or no direct evidence. This problem is faced in different forms, two of which are outlined here to provide additional context for approval decisions covering unrepresented histologies. The first scenario pertains to making decisions about treatments using unrepresentative data. For example, approval is commonly granted for populations which are only imperfectly represented in trial data. This is the problem of external validity and is common with randomised trial data as clinical trials tend to be conducted populations which differ from the population of interest.208 The second scenario is using a technology in new indications for which there is no data. For example pembrolizumab has been submitted for approval in a range of indications, including: squamous non-small-cell lung cancer, urothelial cancer, head and neck cancer among others.209-211 In this case approval may not be granted for a new indication unless there is direct evidence in the population of interest. The approval in unrepresented histologies for histology-independent technologies represents a space between these two scenarios. Approval decisions for treatments for use in unrepresented histologies will depend on context, in some cases it will be more similar to approving in a slightly different population, and in others it will be more analogous to approval in a completely new indication.

Assessments of heterogeneity in survival outcomes at the point of intial marketing authorisation maybe challenging due to data immaturity and potential confounding, unless these are more explictly linked to a surrogate outcome (e.g. response, DoR) for which more robust assessments of heterogeneity may be feasible. While BHM approaches could in theory be explored in the context of time to event endpoints (PFS, OS), the small numbers, potential for greater heterogeneity, high censoring and potential confounding remain important obstacles. However, many of the challenges associated with immaturity in time to event endpoint and potential confounding in uncontrolled Phase II studies are not restricted to histology-independent appraisals. Our review of NICE TAs for products approved with ORR as the primary endpoint, identified a potential disconnect between the regulators’ acceptance of surrogate endpoints and the limited use of surrogate relationships in the corresponding NICE appraisals. While this disconnect may reflect legitimate concerns regarding the reliability of ORR and CR as surrogates for PFS or OS, we recommend that exploration of the surrogate relationships between response-based outcomes (ORR and DoR) should be more routinely considered in economic modelling in help inform and/or validate longer term extrapolations of PFS and OS due to the likely immaturity of these endpoints. NICE will need to consider whether their existing methods guide needs to be more explicit about the challenges of uncontrolled Phase II studies and whether more specific guidance is needed concerning the role and use of surrogate endpoints in these circumstances.

The presentation of the scale of the consequences of heterogeneity and decision uncertainty using population NHB may provide an important additional approach to the assessments already routinely specified with the exiting TA methods guide. Similar arguments have been made in the context of regenerative medicines and cell therapies.192 As part of their ongoing methods review NICE should consider whether the types of metrics presented in this report should be routinely requested within company submissions.

## Priorities for future methodological research

Methods were suggested that allowed for potential sources of heterogeneity of effect across tumour type or other patient characteristics to be accounted for, whilst still allowing some degree of borrowing of strength when estimating treatment effectiveness. However, estimation of the level of heterogeneity can be poor when evidence is sparse. Approaches for considering external evidence and expert opinion to construct an informative prior distribution for the heterogeneity parameter, may be an area for further research.

Even if the heterogeneity parameter can be well estimated, it is unclear what degree of borrowing should be allowed when there is evidence of a high or very high level of heterogeneity. In particular, the implications of borrowing strength across treatment effects in the presence of very high heterogeneity and consequences for uncertainty in decision making should be researched.

So far, methods such as the BHM which allow borrowing of information have mainly been applied to response endpoints. Their extension to time-to-event endpoints and potential for adjustment for known prognostic factors and other confounders would be an interesting area for further research. In addition, further research should also consider the application of BHM approaches to surrogate relationships to determine the validity of borrowing across different subgroups and drug classes.

Given the increasing use of uncontrolled Phase II studies to support initial regulatory approval based on surrogate endpoints, further methodological research is required to determine the basis for selecting between alternative surrogate endpoints for HTA assesments and specifically the appriate basis for selecting specific landmark response and survival time points.

Given the importance of testing costs as a source of heterogeneity and the lack of a clear concensus on the appropriate basis for apportioning costs between multiple current and future targets, further methodological research should more fully establish how these costs should be appropriately included in future NICE appraisals.

# References

1. U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research. Approval Package for application number 125514Orig1s014. Trade Name: Keytruda. Generic or Proper name: pembrolizumab In. Rockville, Maryland, United States; 2017. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/125514Orig1s014.pdf>

2. European Medicines Agency. *Assessment report: VITRAKVI. International non-proprietary name: larotrectinib*: EMA; 2019. Available from: <https://www.ema.europa.eu/en/documents/assessment-report/vitrakvi-epar-public-assessment-report_en.pdf>

3. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *N Engl J Med* 2018;378:731-39.

4. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal*. London: NICE; 2013. Available from: <https://www.nice.org.uk/process/pmg9/chapter/foreword>

5. U.S. Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). *Developing Targeted Therapies in Low-Frequency Molecular Subsets of a Disease. Guidance for Industry*. Silver Spring, MD, USA: FDA; 2018. Available from: <https://www.fda.gov/media/117173/download>

6. U.S. Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) OCoEO. Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics. Guidance for Industry. In. Silver Spring, MD, USA; 2018. Available from: <https://www.fdanews.com/ext/resources/files/2018/2/09-28-18-Masterprotocols.pdf?1538162468>

7. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989;10:1-10.

8. U.S. Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for Industry Expedited Programs for Serious Conditions – Drugs and Biologics. In. Silver Spring, MD, USA; 2014. Available from: <https://www.fda.gov/files/drugs/published/Expedited-Programs-for-Serious-Conditions-Drugs-and-Biologics.pdf>

9. U.S. Department of Health and Human Services. Food and Drug Administration. *Drug Approval Package: Vitrakvi (larotrectinib)*. FDA; 2018. [cited 2020 17th January]. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210861Orig1s000_21171Orig1s000TOC.cfm>.

10. U.S. 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, Maryland, United States; 2019. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212725Orig1s000,%20212726Orig1s000Approv.pdf>

11. Kim C, Prasad V. Strength of validation for surrogate end points used in the US Food and Drug Administration's approval of oncology drugs. *Mayo Clin Proc* 2016;91:713-25.

12. European Medicines Agency. *Workshop on site and histology - Independent indications in oncology*. EMA; 2017. [cited 2020 17th January]. Available from: <https://www.ema.europa.eu/en/events/workshop-site-histology-independent-indications-oncology>.

13. European Medicines Agency. *Workshop on single-arm trials in oncology*. EMA; 2016. [cited 2020 17th January]. Available from: <https://www.ema.europa.eu/en/events/workshop-single-arm-trials-oncology>.

14. European Medicines Agency. *Guideline on the evaluation of anticancer medicinal products in man*: EMA; 2017. Available from: <https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf>

15. European Medicines Agency. *Concept paper on the revision of the guideline on the evaluation of anticancer medicinal products in man*: EMA; 2018. Available from: <https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-revision-guideline-evaluation-anticancer-medicinal-products-man_en.pdf>

16. Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chronic Dis* 1961;13:346-53.

17. Fleming TR. One-sample multiple testing procedure for Phase II clinical trials. *Biometrics* 1982;38:143-51.

18. Lin Y, Shih WJ. Adaptive two-stage designs for single-arm phase iia cancer clinical trials. *Biometrics* 2004;60:482-90.

19. Herson J. Predictive probability early termination plans for Phase II clinical trials. *Biometrics* 1979;35:775-83.

20. Thall PF, Simon R. Practical Bayesian guidelines for Phase IIB Clinical Trials. *Biometrics* 1994;50:337-49.

21. Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat Med* 1997;16:1791-802.

22. Tan S-B, Machin D. Bayesian two-stage designs for phase II clinical trials. *Stat Med* 2002;21:1991-2012.

23. Blagden SP, Billingham L, Brown LC, Buckland SW, Cooper AM, Ellis S, et al. Effective delivery of Complex Innovative Design (CID) cancer trials—A consensus statement. *Br J Cancer* 2020.

24. Redig AJ, Pasi AJ. Basket trials and the evolution of clinical trial design in an era of genomic medicine. *J Clin Oncol* 2015;33:975-77.

25. Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol* 2017;28:34-43.

26. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. *Cell* 2017;168:584-99.

27. Renfro LA, Mandrekar SJ. Definitions and statistical properties of master protocols for personalized medicine in oncology. *J Biopharm Stat* 2018;28:217-28.

28. Simon R. New designs for basket clinical trials in oncology. *J Biopharm Stat* 2018;28:245-55.

29. Lopez-Chavez A, Thomas A, Rajan A, Raffeld M, Morrow B, Kelly R, et al. Molecular profiling and targeted therapy for advanced thoracic malignancies: A biomarker-derived, multiarm, multihistology Phase II basket trial. *J Clin Oncol* 2015;33:1000-07.

30. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay J-Y, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N Engl J Med* 2015;373:726-36.

31. Le Tourneau C, Delord J-P, Gonçalves A, Gavoille C, Dubot C, Isambert N, et al. Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *The Lancet Oncology* 2015;16:1324-34.

32. Beckman RA, Antonijevic Z, Kalamegham R, Chen C. Adaptive design for a confirmatory basket trial in multiple tumor types based on a putative predictive biomarker. *Clin Pharmacol Ther* 2016;100:617-25.

33. Antonijevic Z, Beckman R, editors. *Platform trial designs in drug development: Umbrella trials and basket trials*. Boca Raton: CRD Press, 2019.

34. Freidlin B, Korn EL. Borrowing information across subgroups in Phase II trials: Is it useful? *Clin Cancer Res* 2013;19:1326-34.

35. Simon R. Critical review of umbrella, basket, and platform designs for oncology clinical trials. *Clin Pharmacol Ther* 2017;102:934-41.

36. Simon RM. Primary site idependent clinical trials in oncology. In: Antonijevic Z, Beckman R, editors. *Platform trials in drug development: Umbrella trials and basket trials*. Boca Raton: CRC Press; 2019. p. 167-80.

37. London WB, Chang MN. One- and two-stage designs for stratified phase II clinical trials. *Stat Med* 2005;24:2597-611.

38. LeBlanc M, Rankin C, Crowley J. Multiple histology Phase II trials. *Clin Cancer Res* 2009;15:4256-62.

39. Jung S-H, Chang MN, Kang SJ. Phase II cancer clinical trials with heterogeneous patient populations. *J Biopharm Stat* 2012;22:312-28.

40. Cunanan KM, Iasonos A, Shen R, Begg CB, Gönen M. An efficient basket trial design. *Stat Med* 2017;36:1568-79.

41. Simon R, Geyer S, Subramanian J, Roychowdhury S. The Bayesian basket design for genomic variant-driven phase II trials. *Semin Oncol* 2016;43:13-18.

42. Palmer AC, Plana D, Sorger PK. Comparing the efficacy of cancer therapies between subgroups in basket trials. *bioRxiv* 2018:401620.

43. Spiegelhalter DJ, Abrams KR, Myles J. *Bayesian approaches to clinical trials and Health-Care Evaluation*. New York: Wiley; 2004.

44. 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;22:763-80.

45. Thall PF, Sung H-G. Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials. *Stat Med* 1998;17:1563-80.

46. Chugh R, Wathen JK, Maki RG, Benjamin RS, Patel SR, Myers PA, et al. Phase II multicenter trial of imatinib in 10 histologic subtypes of sarcoma using a Bayesian hierarchical statistical model. 2009;27:3148-53. Available from: <https://ascopubs.org/doi/abs/10.1200/JCO.2008.20.5054>

47. Berry SM, Broglio KR, Groshen S, Berry DA. Bayesian hierarchical modeling of patient subpopulations: Efficient designs of Phase II oncology clinical trials. *Clin Trial* 2013;10:720-34.

48. Liu R, Liu Z, Ghadessi M, Vonk R. Increasing the efficiency of oncology basket trials using a Bayesian approach. *Contemp Clin Trials* 2017;63:67-72.

49. Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharm Stat* 2016;15:123-34.

50. Cunanan KM, Iasonos A, Shen R, Gönen M. Variance prior specification for a basket trial design using Bayesian hierarchical modeling. *Clin Trial* 2019;16:142-53.

51. Leon-Novelo LG, Bekele BN, Müller P, Quintana F, Wathen K. Borrowing strength with nonexchangeable priors over subpopulations. *Biometrics* 2012;68:550-58.

52. Chu Y, Yuan Y. BLAST: Bayesian latent subgroup design for basket trials accounting for patient heterogeneity. *J R Stat Soc Ser C Appl Stat* 2018;67:723-40.

53. Chu Y, Yuan Y. A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin Trial* 2018;15:149-58.

54. Fujikawa K, Teramukai S, Yokota I, Daimon T. A Bayesian basket trial design that borrows information across strata based on the similarity between the posterior distributions of the response probability. *Biometrical J* 2019.

55. Hobbs BP, Landin R. Bayesian basket trial design with exchangeability monitoring. *Stat Med* 2018;37:3557-72.

56. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* 2010;363:809-19.

57. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* 2012;483:100-03.

58. Heinrich MC, Joensuu H, Demetri GD, Corless CL, Apperley J, Fletcher JA, et al. Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. *Clin Cancer Res* 2008;14:2717-25.

59. Hudis CA. Trastuzumab — mechanism of action and use in clinical practice. *N Engl J Med* 2007;357:39-51.

60. Fleming GF, Sill MW, Darcy KM, McMeekin DS, Thigpen JT, Adler LM, et al. Phase II trial of trastuzumab in women with advanced or recurrent, HER2-positive endometrial carcinoma: A Gynecologic Oncology Group study. *Gynecol Oncol* 2010;116:15-20.

61. Gatzemeier U, Groth G, Butts C, Van Zandwijk N, Shepherd F, Ardizzoni A, et al. Randomized phase II trial of gemcitabine–cisplatin with or without trastuzumab in HER2-positive non-small-cell lung cancer. *Ann Oncol* 2004;15:19-27.

62. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012;31:2973-84.

63. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. *Health Technol Assess* 2009;13:1-72.

64. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine* 1996;125:605-13.

65. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431-40.

66. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49-67.

67. Heller G. Statistical controversies in clinical research: an initial evaluation of a surrogate end point using a single randomized clinical trial and the Prentice criteria. *Ann Oncol* 2015;26:2012-16.

68. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89-95.

69. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users’ guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. *JAMA* 1999;282:771-78.

70. Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. *Int J Technol Assess Health Care* 2009;25:6-13.

71. German Institute of Quality and Efficiency in Health Care (IQWiG). *Validity of surrogate endpoints in Oncology*; 2011.

72. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC Med Res Methodol* 2012;12:27.

73. Buyse M, Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biometrical J* 2016;58:104-32.

74. Fischer A, Hernandez-Villafuerte K, Latimer N, Henshall C. *Extrapolation from Progression-Free Survival to Overall Survival in Oncology*: Office of Health Economics; 2016.

75. Davis S, Tappenden P, Cantrell A. *A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer*. University of Sheffield report for NICE Decision Support Unit; 2012.

76. Savina M, Gourgou S, Italiano A, Dinart D, Rondeau V, Penel N, et al. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: A critical review. *Critical Reviews in Oncology Hematology* 2018;123:21-41.

77. Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer* 2019;106:196-211.

78. Abdel-Rahman O. Surrogate end points for overall survival in trials of PD-(L)1 inhibitors for urinary cancers: a systematic review. *Immunotherapy* 2018;10:139-48.

79. Agarwal N, Bellmunt J, Maughan BL, Boucher KM, Choueiri TK, Qu AQ, et al. Six-month progression-free survival as the primary endpoint to evaluate the activity of new agents as second-line therapy for advanced urothelial carcinoma. *Clin Genitourin Cancer* 2014;12:130-7.

80. Agarwal SK, Mangal N, Menon RM, Freise KJ, Salem AH. Response rates as predictors of overall survival: A meta-analysis of acute myeloid leukemia trials. *J Cancer* 2017;8:1562-67.

81. Blumenthal GM, Karuri SW, Zhang H, Zhang LJ, Khozin S, Kazandjian D, et al. Overall response rate, progression-free survival, and overall survival with targeted and standard therapies in advanced non-small-cell lung cancer: Us food and drug administration trial-level and patient-level analyses. *J Clin Oncol* 2015;33:1008-+.

82. Blumenthal GM, Zhang L, Zhang H, Kazandjian D, Khozin S, Tang S, et al. Milestone analyses of immune checkpoint inhibitors, targeted therapy, and conventional therapy in metastatic non-small cell lung cancer trials: A meta-analysis. *JAMA Oncol* 2017;3:e171029.

83. Bruzzi P, Del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C, et al. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *J Clin Oncol* 2005;23:5117-25.

84. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol* 2008;26:1987-92.

85. Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: A meta-analysis. *Lancet* 2000;356:373-78.

86. Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. *J Clin Epidemiol* 2015;68:833-42.

87. Colloca G, Venturino A. Trial-level analysis of progression-free survival and response rate as end points of trials of first-line chemotherapy in advanced ovarian cancer. *Med Oncol* 2017;34:87.

88. Colloca G, Venturino A, Guarneri D. Analysis of response-related and time-to-event endpoints in randomized trials of gemcitabine-based treatment versus gemcitabine alone as first-line treatment of patients with advanced pancreatic cancer. *Clin Colorectal Cancer* 2016a;15:264-76.

89. Colloca G, Venturino A, Guarneri D. Analysis of clinical end points of randomised trials including bevacizumab and chemotherapy versus chemotherapy as first-line treatment of metastatic colorectal cancer. *Clin Oncol (R Coll Radiol)* 2016b;28:e155-64.

90. Colloca G, Vitucci P, Venturino A. Trial level analysis of prostate-specific antigen-related versus unrelated endpoints in Phase III trials of first-line and second-line medical treatments of patients with metastatic castration-resistant prostate cancer. *Clin Genitourin Cancer* 2016c;14:389-97.

91. Cremolini C, Antoniotti C, Pietrantonio F, Berenato R, Tampellini M, Baratelli C, et al. Surrogate endpoints in second-line trials of targeted agents in metastatic colorectal cancer: A literature-based systematic review and meta-analysis. *Cancer Res Treat* 2017;49:834-45.

92. Delea TE, Khuu A, Heng DYC, Haas T, Soulieres D. Association between treatment effects on disease progression end points and overall survival in clinical studies of patients with metastatic renal cell carcinoma. *Br J Cancer* 2012;107:1059-68.

93. Elia EG, Städler N, Ciani O, Taylor RS, Bujkiewicz S. Combining tumour response and progression free survival as surrogate endpoints for overall survival in advanced colorectal cancer. *arXiv preprint arXiv:1809.02935* 2018.

94. Foster NR, Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer* 2011;117:1262-71.

95. Giessen C, Laubender RP, Ankerst DP, Stintzing S, Modest DP, Schulz C, et al. Surrogate endpoints in second-line treatment for mCRC: a systematic literature-based analysis from 23 randomised trials. *Acta Oncologica* 2015;54:187-93.

96. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *Br J Cancer* 2005;93:1215-21.

97. Hamada T, Nakai Y, Isayama H, Yasunaga H, Matsui H, Takahara N, et al. Progression-free survival as a surrogate for overall survival in first-line chemotherapy for advanced pancreatic cancer. *Eur J Cancer* 2016;65:11-20.

98. Han K, Ren M, Wick W, Abrey L, Das A, Jin J, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. *Neuro-Oncology* 2014;16:696-706.

99. Hashim M, Pfeiffer BM, Bartsch R, Postma M, Heeg B. Do surrogate endpoints better correlate with overall survival in studies that did not allow for crossover or reported balanced postprogression treatments? An application in advanced non-small cell lung cancer. *Value Health* 2018;21:9-17.

100. Hotta K, Kato Y, Leighl N, Takigawa N, Gaafar RM, Kayatani H, et al. Magnitude of the benefit of progression-free survival as a potential surrogate marker in Phase 3 trials assessing targeted agents in molecularly selected patients with advanced non-small cell lung cancer: Systematic review. *Plos One* 2015;10.

101. Hotta K, Kiura K, Fujiwara Y, Takigawa N, Oze I, Ochi N, et al. Association between incremental gains in the objective response rate and survival improvement in phase III trials of first-line chemotherapy for extensive disease small-cell lung cancer. *Ann Oncol* 2009;20:829-34.

102. Ichikawa W, Sasaki Y. Correlation between tumor response to first-line chemotherapy and prognosis in advanced gastric cancer patients. *Ann Oncol* 2006;17:1665-72.

103. Imaoka H, Sasaki M, Takahashi H, Hashimoto Y, Ohno I, Mitsunaga S, et al. Progression-free survival as a surrogate endpoint in advanced neuroendocrine neoplasms. *Endocr Relat Cancer* 2017;24:475-83.

104. Imaoka H, Sasaki M, Takahashi H, Hashimoto Y, Ohno I, Mitsunaga S, et al. Alternate endpoints for Phase II trials in advanced neuroendocrine tumors. *Oncologist* 2019;24:47-53.

105. Iezzi R, Pompili M, Rinninella E, Annicchiarico E, Garcovich M, Cerrito L, et al. TACE with degradable starch microspheres (DSM-TACE) as second-line treatment in HCC patients dismissing or ineligible for sorafenib. *Eur Radiol* 2019;29:1285-92.

106. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol* 2006;7:741-6.

107. Kaufman HL, Schwartz LH, William WN, Jr., Sznol M, Fahrbach K, Xu Y, 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;144:2245-61.

108. Lee L, Wang L, Crump M. Identification of potential surrogate end points in randomized clinical trials of aggressive and indolent non-Hodgkin's lymphoma: correlation of complete response, time-to-event and overall survival end points. *Ann Oncol* 2011;22:1392-403.

109. Li J, He Q, Yu X, Khan K, Weng XW, Guan MJ. Complete response associated with immune checkpoint inhibitors in advanced non-small-cell lung cancer: a meta-analysis of nine randomized controlled trials. *Cancer Manag Res* 2019;11:1623-29.

110. Li X, Liu S, Gu H, Wang D. Surrogate end points for survival in the target treatment of advanced non-small-cell lung cancer with gefitinib or erlotinib. *J Cancer Res Clin Oncol* 2012;138:1963-9.

111. Liu L, Chen F, Zhao J, Yu H. Correlation between overall survival and other endpoints in metastatic breast cancer with second- or third-line chemotherapy: Literature-based analysis of 24 randomized trials. *Bull Cancer* 2016;103:336-44.

112. Louvet C, de Gramont A, Tournigand C, Artru P, Maindrault-Goebel F, Krulik M. Correlation between progression free survival and response rate in patients with metastatic colorectal carcinoma. *Cancer* 2001;91:2033-8.

113. Makris EA, MacBarb R, Harvey DJ, Poultsides GA. Surrogate end points for overall survival in metastatic, locally advanced, or unresectable pancreatic cancer: A systematic review and meta-analysis of 24 randomized controlled trials. *Ann Surg Oncol* 2017;24:2371-78.

114. Mangal N, Salem AH, Li MY, Menon R, Freise KJ. Relationship between response rates and median progression-free survival in non-Hodgkin's lymphoma: A meta-analysis of published clinical trials. *Hematol Oncol* 2018;36:37-43.

115. Mangal N, Salem AH, Menon RM, Freise KJ. Use of depth of response to predict progression-free survival in relapsed or refractory multiple myeloma: Evaluation of results from 102 clinical trials. *Hematol Oncol* 2018;36:547-53.

116. Moriwaki T, Yamamoto Y, Gosho M, Kobayashi M, Sugaya A, Yamada T, 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;114:881-8.

117. Mushti SL, Mulkey F, Sridhara R. Evaluation of overall response rate and progression-free survival as potential surrogate endpoints for overall survival in immunotherapy trials. *Clin Cancer Res* 2018;24:2268-75.

118. Nakashima K, Horita N, Nagai K, Manabe S, Murakami S, Ota E, et al. Progression-free survival, response rate, and disease control rate as predictors of overall survival in phase iii randomized controlled trials evaluating the first-line chemotherapy for advanced, locally advanced, and recurrent non-small cell lung carcinoma. *J Thorac Oncol* 2016;11:1574-85.

119. Nickolich M, Babakoohi S, Fu P, Dowlati A. Clinical trial design in small cell lung cancer: surrogate end points and statistical evolution. *Clinl Lung Cancer* 2014;15:207-12.

120. Nie RC, Chen FP, Yuan SQ, Luo YS, Chen S, Chen YM, et al. Evaluation of objective response, disease control and progression-free survival as surrogate end-points for overall survival in anti-programmed death-1 and anti-programmed death ligand 1 trials. *Eur J Cancer* 2019;106:1-11.

121. Pang Y, Shen Z, Sun J, Wang W. Does the use of targeted agents in advanced gastroesophageal cancer increase complete response? A meta-analysis of 18 randomized controlled trials. *Cancer Manag Res* 2018;10:5505-14.

122. Penel N, Ryckewaert T, Kramar A. What is an active regimen in carcinoma of unknown primary sites? Analysis of correlation between activity endpoints reported in phase II trials. Correlation of activity endpoints in phase II trials. *Bull Cancer* 2014;101:E19-24.

123. Petrelli F, Barni S. Surrogate end points and postprogression survival in renal cell carcinoma: an analysis of first-line trials with targeted therapies. *Clin Genitourin Cancer* 2013;11:385-9.

124. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: an analysis of the first-line phase III trials. *Med Oncol* 2014;31:776.

125. Ritchie G, Gasper H, Man J, Lord S, Marschner I, Friedlander M, et al. Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers a systematic review and meta-analysis. *JAMA Oncol* 2018;4:522-28.

126. Rose PG, Tian CQ, Bookman MA. Assessment of tumor response as a surrogate endpoint of survival in recurrent/platinum-resistant ovarian carcinoma: A Gynecologic Oncology Group study. *Gynecol Oncol* 2010;117:324-29.

127. Roviello G, Andre F, Venturini S, Pistilli B, Curigliano G, Cristofanilli M, et al. Response rate as a potential surrogate for survival and efficacy in patients treated with novel immune checkpoint inhibitors: A meta-regression of randomised prospective studies. *Eur J Cancer* 2017;86:257-65.

128. Sekine I, Tamura T, Kunitoh H, Kubota K, Shinkai T, Kamiya Y, et al. Progressive disease rate as a surrogate endpoint of phase II trials for non-small-cell lung cancer. *Ann Oncol* 1999;10:731-33.

129. Shi Q, Flowers CR, Hiddemann W, Marcus R, Herold M, Hagenbeek A, et al. Thirty-month complete response as a surrogate end point in first-line follicular lymphoma therapy: An individual patient-level analysis of multiple randomized trials. *J Clin Oncol* 2017;35:552-60.

130. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Correlation between overall survival and other endpoints in clinical trials of second-line chemotherapy for patients with advanced gastric cancer. *Gastric Cancer* 2014;17:362-70.

131. Shukuya T, Mori K, Amann JM, Bertino EM, Otterson GA, Shields PG, et al. Relationship between overall survival and response or progression-free survival in advanced non-small cell lung cancer patients treated with anti-PD-1/PD-L1 antibodies. *J Thorac Oncol* 2016;11:1927-39.

132. Siddiqui MK, Tyczynski J, Pahwa A, Fernandes AW. Objective response rate is a possible surrogate endpoint for survival in patients with advanced, recurrent ovarian cancer. *Gynecol Oncol* 2017;146:44-51.

133. Sidhu R, Rong A, Dahlberg S. Evaluation of progression-free survival as a surrogate endpoint for survival in chemotherapy and targeted agent metastatic colorectal cancer trials. *Clin Cancer Res* 2013;19:969-76.

134. Tanaka K, Kawano M, Iwasaki T, Itonaga I, Tsumura H. Surrogacy of intermediate endpoints for overall survival in randomized controlled trials of first-line treatment for advanced soft tissue sarcoma in the pre- and post-pazopanib era: A meta-analytic evaluation. *BMC Cancer* 2019;19 (1).

135. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol* 2007;25:4562-8.

136. Tsujino K, Kawaguchi T, Kubo A, Aono N, Nakao K, Koh Y, et al. Response rate is associated with prolonged survival in patients with advanced non-small cell lung cancer treated with gefitinib or erlotinib. *J Thorac Oncol* 2009;4:994-1001.

137. Tsujino K, Shiraishi J, Tsuji T, Kurata T, Kawaguchi T, Kubo A, et al. Is response rate increment obtained by molecular targeted agents related to survival benefit in the phase III trials of advanced cancer? *Ann Oncol* 2010;21:1668-74.

138. Vidaurre T, Wilkerson J, Simon R, Bates SE, Fojo T. Stable disease is not preferentially observed with targeted therapies and as currently defined has limited value in drug development. *Cancer J* 2009;15:366-73.

139. Wilkerson J, Fojo T. Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. *Cancer J* 2009;15:379-85.

140. Zer A, Prince RM, Amir E, Abdul Razak A. Evolution of randomized trials in advanced/metastatic soft tissue sarcoma: End point selection, surrogacy, and quality of reporting. *J Clin Oncol* 2016;34:1469-75.

141. Zhu R, Lu D, Chu YW, Chai A, Green M, Zhang N, et al. Assessment of correlation between early and late efficacy endpoints to identify potential surrogacy relationships in non-hodgkin lymphoma: A literature-based meta-analysis of 108 Phase II and Phase III studies. *AAPS Journal* 2017;19:669-81.

142. Bujkiewicz S, Achana F, Papanikos T, Riley RD, Abrams KR. *Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. Report by The Decision Support Unit*. Sheffield: Decision Support Unit, ScHARR, University of Sheffield; 2019. Available from: <http://nicedsu.org.uk/wp-content/uploads/2019/10/TSD-20-mvmeta-final.pdf>

143. Bujkiewicz S, Jackson D, Thompson J, Turner R, Städler N, Abrams K, et al. Bivariate network meta-analysis for surrogate endpoint evaluation. *Stat Med* 2019;38:3322-41.

144. Phelps CE, Mushline AI. Focusing technology assessment using medical decision theory. *Med Decis Making* 1988;8:279-89.

145. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ* 2003;12:421-7.

146. 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;34:951-64.

147. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Med Decis Making* 2007;27:112-27.

148. Cancer Research UK. *Cancer Research UK policy statement: Patient access to molecular diagnostics and targeted medicines in England*. London; 2018. Available from: <https://www.cancerresearchuk.org/sites/default/files/access_to_molecular_diagnostic_tests_and_targeted_medicines_in_england_0.pdf>

149. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS book*. Boca Raton, FL: CRC Press; 2013.

150. R Core Team. *R: A language and environment for statistical computing (*[*https://www.R-project.org/*](https://www.R-project.org/)*)*. [program] Vienna, Austria: R Foundation for Statistical Computing; 2019.

151. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* 2005;12:1-16 (<http://www.jstatsoft.org/v12/i03>). Available from: <http://www.jstatsoft.org/v12/i03>

152. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998;7:434-55.

153. Gelman A, Rubin DB. Inferences from iterative simulation using multiple sequences. *Statistical Science* 1992;7:457-72.

154. National Institute for Health and Care Excellence (NICE). *Larotrectinib for treating NTRK fusion-positive advanced solid tumours*: National Institute for Health and Care Excellence; 2019.

155. Bongarzone I, Vigneri P, Mariani L, Collini P, Pilotti S, Pierotti MA. RET/NTRK1 rearrangements in thyroid gland tumors of the papillary carcinoma family: correlation with clinicopathological features. *Clin Cancer Res* 1998;4:223-28.

156. Prasad ML, Vyas M, Horne MJ, Virk RK, Morotti R, Liu Z, et al. NTRK fusion oncogenes in pediatric papillary thyroid carcinoma in northeast United States. *Cancer* 2016;122:1097-107.

157. Vokuhl C, Nourkami‐Tutdibi N, Furtwängler R, Gessler M, Graf N, Leuschner I. ETV6–NTRK3 in congenital mesoblastic nephroma: A report of the SIOP/GPOH nephroblastoma study. *Pediatric Blood and Cancer* 2018;65:e26925.

158. Phillippo DM, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams and N. J. Welton. . *NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE*. Sheffield, UK, ScHARR, University of Sheffield.; 2016. Available from: <https://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/Population-adjustment-TSD-FINAL.pdf>

159. 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.

160. Hatswell AJ, Sullivan WG. Creating historical controls using data from a previous line of treatment - Two non-standard approaches. *Stat Methods Med Res* 2019;0:1-10.

161. Hsiao SJ, Zehir A, Sireci AN, Aisner DL. Detection of tumor NTRK gene fusions to identify patients who may benefit from tyrosine kinase (TRK) inhibitor therapy. *J Mol Diagn* 2019;21:553-71.

162. NHS England. *National Genomic Test Directory*. 2019. [cited 2019 15th May]. Available from: <https://www.england.nhs.uk/publication/national-genomic-test-directories/>.

163. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10:789.

164. Morganti S, Tarantino P, Ferraro E, D'Amico P, Viale G, Trapani D, Duso BA, Curigliano G. [Complexity of genome sequencing and reporting: Next generation sequencing (NGS) technologies and implementation of precision medicine in real life.](https://www.ncbi.nlm.nih.gov/pubmed/30661654) *Crit Rev Oncol Hematol*. 2019 Jan;133:171-182

165. Eifert C, Powers RS. From cancer genomes to oncogenic drivers, tumour dependencies and therapeutic targets. *Nature Reviews Cancer* 2012;12:572.

166. Sheikine Y, Kuo FC, Lindeman NI. Clinical and technical aspects of genomic diagnostics for precision oncology. *J Clin Oncol* 2017;35:929-33.

167. Solomon JP, Hechtman JF. Detection of NTRK fusions: merits and limitations of current diagnostic platforms. *Cancer Res* 2019 Jul 1;79(13):3163-3168

168. Penault-Llorca F, Rudzinski ER, Sepulveda AR. Testing algorithm for identification of patients with TRK fusion cancer. *J Clin Pathol* 2019;72:460-67.

169. Wright C. Long Noncoding RNAs and Cancer. In: Gray S, editor. *Epigenetic Cancer Therapy*: Acadmic Press; 2015. p. 91-114.

170. Marchiò C, Scaltriti M, Ladanyi M, Iafrate AJ, Bibeau F, Dietel M, et al. ESMO recommendations on the standard methods to detect NTRK fusions in daily practice and clinical research. *Ann Oncol* 2019.

171. Sigal DS, Bhangoo MS, Hermel JA, Pavlick DC, Frampton G, Miller VA, et al. Comprehensive genomic profiling identifies novel NTRK fusions in neuroendocrine tumors. *Oncotarget* 2018;9:35809-12.

172. Okamura R, Boichard A, Kato S, Sicklick JK, Bazhenova L, Kurzrock R. Analysis of NTRK alterations in pan-cancer adult and pediatric malignancies: implications for NTRK-targeted therapeutics. *JCO Precision Oncology* 2018;2018.

173. Center for Drug Evaluation and Research. *NDA Multidisciplinary Review and Evaluation NDA 210861 and NDA 211710 VITRAKVI (larotrectinib)*: U.S. Food and Drug Administration, ; 2016. Available from: <https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210861Orig1s000_211710Orig1s000MultidisciplineR.pdf>

174. Amatu A, Sartore-Bianchi A, Siena S. NTRK gene fusions as novel targets of cancer therapy across multiple tumour types. *ESMO Open* 2016;1:e000023.

175. Williams HL, Walsh K, Diamond A, Oniscu A, Deans ZC. Validation of the Oncomine™ focus panel for next-generation sequencing of clinical tumour samples. *Virchows Archiv* 2018;473:489-503.

176. Solomon JP, Linkov I, Rosado A, Mullaney K, Rosen EY, Frosina D, et al. NTRK fusion detection across multiple assays and 33,997 cases: diagnostic implications and pitfalls. *Modern Pathology* 2019;33:38-46.

177. Makretsov N, He M, Hayes M, Chia S, Horsman DE, Sorensen PHB, et al. A fluorescence in situ hybridization study of ETV6‐NTRK3 fusion gene in secretory breast carcinoma. *Genes Chromosomes Cancer* 2004;40:152-57.

178. Scottish Science Advisory Council *Informing the Future of Genomic Medicine in Scotland*; Edinburgh, SSAC, 2019. Available from:

<https://www.scottishscience.org.uk/sites/default/files/article-attachments/Genomics%20Short%20Report.pdf>

179. CADTH. *Larotrectinib (Vitrakvi) for Neurotrophic Tyrosine Receptor Kinase (NTRK) Positive Solid Tumours: pan-Canadian Oncology Drug Review. Final Economic Guidance Report* 2019.

180. Woods B, Sideris E, Palmer S, Latimer N, Soares M. *Partitioned survival analysis for decision modelling in health care: A critical review*. Sheffield: Decision Support Unit, ScHARR; 2017. Available from: <https://nicedsu.org.uk/wp-content/uploads/2017/06/Partitioned-Survival-Analysis-final-report.pdf>

181. Latimer NJ. *NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data*. Sheffield: Report by the Decision Support Unit; 2011. Available from: <http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf>

182. 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;37:1129-38.

183. Price MJ, Welton NJ, Briggs AH, Ades AE. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. *Value Health* 2011;14:205-18.

184. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. *Med Decis Making* 2011;31:662-74.

185. Australian Government. Department of Health. *Co-dependent and Hybrid Technologies*. [cited 2020 17th January]. Available from: <https://www1.health.gov.au/internet/hta/publishing.nsf/Content/co-1>.

186. Soares MO, Walker S, Palmer SJ, Sculpher MJ. Establishing the value of diagnostic and prognostic tests in health technology assessment. *Med Decis Making* 2018;38:495-508.

187. Albert CM, Davis JL, Federman N, Casanova M, Laetsch TW. TRK fusion cancers in children: a clinical review and recommendations for screening. *J Clin Oncol* 2019;37:513-24.

188. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;18:S68-S80.

189. Briggs A, Claxton K, Sculpher MJ. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.

190. Hettle R, Corbett M, Hinde S, Hodgson R, Jones-Diette J, Woolacott N, et al. The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. *Health Technology Assessment* 2017;21:1-+.

191. Grimm S, Strong M, Brennan A, Wailoo A. *Framework for analysing risk in health technology assessments and its application to managed entry agreements. Report By The Decision Support Unit*. Sheffield: Decision Support Unit, ScHARR, University of Sheffield; 2016. Available from: <http://nicedsu.org.uk/wp-content/uploads/2018/05/DSU-Managed-Access-report-FINAL.pdf>

192. Hettle R, Corbett M, Hinde S, Hodgson R, Jones-Diette J, Woolacott N, et al. The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. *Health Technol Assess* 2017;21:1-204.

193. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;18:341-64.

194. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, et al. A comprehensive algorithm for approval of health technologies with, without, or only in research: The key principles for informing coverage decisions. *Value Health* 2016;19:885-91.

195. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making* 2004;24:207-27.

196. NHS England Cancer Drugs Fund Team. *Appraisal and Funding of Cancer Drugs from July 2016 (including the new Cancer Drugs Fund): A new deal for patients, taxpayers and industry*. London; 2016. Available from: <https://www.england.nhs.uk/publication/cdf-sop-16/>

197. Claxton K. *Pharmaceutical Pricing: Early Access, The Cancer Drugs Fund and the Role of NICE*. York: Centre for Health Economics, University of York; 2016. Available from: <https://www.york.ac.uk/media/che/documents/policybriefing/Drug_prices.pdf>

198. Grieve R, Abrams K, Claxton K, Goldacre B, James N, Nicholl J, et al. Cancer Drugs Fund requires further reform. *BMJ* 2016;354:i5090.

199. Morrell L, Wordsworth S, Schuh A, Middleton MR, Rees S, Barker RW. Will the reformed Cancer Drugs Fund address the most common types of uncertainty? An analysis of NICE cancer drug appraisals. *BMC Health Services Research* 2018;18.

200. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making* 2014;34:311-26.

201. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: A fast, nonparametric regression-based method. *Med Decis Making* 2015;35:570-83.

202. Willan AR, Pinto EM. The value of information and optimal clinical trial design. *Stat Med* 2005;24:1791-806.

203. National Institute for Health and Care Excellence. *Guide to the processes of technology appraisal* NICE; 2018. Available from: <https://www.nice.org.uk/process/pmg19/resources/guide-to-the-processes-of-technology-appraisal-pdf-72286663351237>

204. Kim DD, Basu A. New metrics for economic evaluation in the presence of heterogeneity: Focusing on evaluating policy alternatives rather than treatment alternatives. *Med Decis Making* 2017;37:930-41.

205. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press; 2015.

206. Claxton K. OFT, VBP: QED? . *Health Econ* 2007;16:545-58.

207. Rothery C, Claxton K, Palmer S, Epstein D, Tarricone R, Sculpher M. Characterising uncertainty in the assessment of medical devices and determining future research needs. *Health Econ* 2017;26:109-23.

208. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 2015;16:495.

209. National Institute For Health And Care Excellence. *Pembrolizumab with carboplatin and paclitaxel for untreated metastatic squamous non-small-cell lung cancer*. London: NICE; 2019. Available from: <https://www.nice.org.uk/guidance/ta600>

210. National Institute for Health and Care Excellence. *Pembrolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable.* London: NICE; 2018. Available from: <https://www.nice.org.uk/guidance/ta522>

211. National Institute for Health and Care Excellence. *Pembrolizumab for untreated metastatic or unresectable recurrent squamous cell head and neck cancer [ID1140]*. NICE; 2020. [cited 2020 20th February]. Available from: <https://www.nice.org.uk/guidance/indevelopment/gid-ta10181>.

212. National Institute for Health and Care Excellence (NICE). *TA492: Atezolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable*; 2018. Available from: <https://www.nice.org.uk/guidance/ta492>

213. National Institute for Health and Care Excellence (NICE). *TA510: Daratumumab monotherapy for treating relapsed and refractory multiple myeloma*; 2018. Available from: <https://www.nice.org.uk/guidance/ta510>

214. National Institute for Health and Care Excellence (NICE). *TA487: Venetoclax for treating chronic lymphocytic leukaemia*; 2017. Available from: <https://www.nice.org.uk/guidance/ta487>

215. National Institute for Health and Care Excellence (NICE). *TA462: Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma*; 2017. Available from: <https://www.nice.org.uk/guidance/ta462>

216. National Institute for Health and Care Excellence (NICE). *GID-TA10268: Brigatinib for treating ALK-positive non-small-cell lung cancer after crizotinib [ID1328]*; 2019. Available from: <https://www.nice.org.uk/guidance/indevelopment/gid-ta10268>

217. National Institute for Health and Care Excellence (NICE). *TA416: Osimertinib for treating locally advanced or metastatic EGFR T790M mutation-positive non-small-cell lung cancer*; 2016. Available from: <https://www.nice.org.uk/guidance/ta416>

218. National Institute for Health and Care Excellence (NICE). *TA517: Avelumab for treating metastatic merkel cell carcinoma*; 2017. Available from: <https://www.nice.org.uk/guidance/ta517>

219. National Institute for Health and Care Excellence (NICE). *TA529: Crizotinib for treating ROS1-positive advanced non-small-cell lung cancer*; 2018. Available from: <https://www.nice.org.uk/guidance/ta529>

220. Hatswell AJ, Freemantle N, Baio GJP. Economic evaluations of pharmaceuticals granted a marketing authorisation without the results of randomised trials: a systematic review and taxonomy. *Pharmacoeconomics* 2017;35:163-76.

221. Faria R, Hernandez Alava M, Manca A, Wailoo AJ. *The use of observational data to inform estimates of treatment effectiveness for technology appraisal: methods for comparative individual patient data. Report by the Decision Support Unit*. Sheffield: Decision Support Unit, ScHARR, University of Sheffield; 2015. Available from: <http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD17-DSU-Observational-data-FINAL.pdf>

222. Bell H, Wailoo A, Hernandez M, Grieve R, Faria R, Gibson L, et al. *The use of real world data for the estimation of treatment effects in NICE decision making. Report by the Decision Support Unit*. Sheffield: Decision Support Unit, ScHARR, University of Sheffield; 2016. Available from:

<http://nicedsu.org.uk/wp-content/uploads/2018/05/RWD-DSU-REPORT-Updated-DECEMBER-2016.pdf>

223. National Institute for Health and Care Excellence (NICE). *TA540: Pembrolizumab for treating relapsed or refractory classical Hodgkin lymphoma*; 2018. Available from: <https://www.nice.org.uk/guidance/ta540>

224. Wright SJ, Newman WG, Payne K. Accounting for capacity constraints in economic evaluations of precision medicine: a systematic review. *Pharmacoeconomics* 2019;37:1011-27.

225. Lee SJ, Kim NKD, Lee S-H, Kim ST, Park SH, Park JO, et al. NTRK gene amplification in patients with metastatic cancer. *Precision Future Med* 2017;1:129-37.

226. Lange AM, Lo H-W. Inhibiting TRK proteins in clinical cancer therapy. *Cancers* 2018;10:105.

227. Office for National Statistics. *Cancer Registration Statistics, England: 2017*. In. London; 2019. Available from: <https://www.ons.gov.uk/releases/cancerregistrationstatisticsengland2017>

228. Coombes N, Elliss-Brookes L, Johnson S, Lyons J, Mackay C, Morement H, et al. *Rare and less common cancers: Incidence and Mortality in England, 2010 to 2013*. In: Network NCI, editor. London, UK: Public Health England; 2015.

229. UK and Ireland Neuroendcrine Tumour Society. *Incidence and prevalence of neuroendocrine tumours in England*. In: UKI NETS. London, UK: Endocrine Abstracts; 2017.

230. National Cancer Institute. *Gastrointestinal Stromal Tumors Treatment (Adult) (PDQ®)–Health Professional Version*. National Cancer Institute; 2018. [cited 2019 June]. Available from: <https://www.cancer.gov/types/soft-tissue-sarcoma/hp/gist-treatment-pdq>.

231. Care Quality Improvement Department RCoP. *National Lung Cancer Audit annual report 2018 (for the audit period 2017)*. In: Physicians RCo, editor. London, UK: Healthcare Quality Improvement Partnership (HQIP) 2019.

232. World Population Review. *Western European Countries 2020. Western European Countries*. 2019. [cited 2019 12th December]. Available from: <http://worldpopulationreview.com/countries/western-european-countries/>.

233. Marmor S, Portschy PR, Tuttle TM, Virnig BA. The rise in appendiceal cancer incidence: 2000–2009. *J Gastrointest Surg* 2015;19:743-50.

234. Tsuchiya N, Sawada Y, Endo I, Saito K, Uemura Y, Nakatsura T. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol* 2015;21:10573.

235. Sethi R, Kozin E, Remenschneider A, Meier J, VanderLaan P, Faquin W, et al. Mammary analogue secretory carcinoma: update on a new diagnosis of salivary gland malignancy. *Laryngoscope* 2014;124:188-95.

236. Orbach D, Rey A, Cecchetto G, Oberlin O, Casanova M, Thebaud E, et al. Infantile fibrosarcoma: management based on the European experience. *J Clin Oncol* 2009;28:318-23.

237. Gooskens SL, Houwing ME, Vujanic GM, Dome JS, Diertens T, Coulomb‐l'Herminé A, et al. Congenital mesoblastic nephroma 50 years after its recognition: a narrative review. *Pediatric blood & cancer* 2017;64:e26437.

238. Deen MH, Burke KM, Janitz A, Campbell J. Cancers of the thyroid: overview and statistics in the United States and Oklahoma. *The Journal of the Oklahoma State Medical Association* 2016;109:333.

239. Austin M, Xing Y, Hayes-Jordan A, Lally K, Cormier J. Melanoma incidence rises for children and adolescents: an epidemiologic review of pediatric melanoma in the United States. *J Pediatr Surg* 2013;48:2207.

240. Cancer Research UK. *Breast cancer incidence by stage at diagnosis* Cancer Research UK; 2014. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence>.

241. Thrumurthy SG, Thrumurthy SSD, Gilbert CE, Ross P, Haji A. Colorectal adenocarcinoma: risks, prevention and diagnosis. *BMJ* 2016;354:i3590.

242. Cancer Research UK. *Bowel cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer>.

243. Amelio JMS, Ruzafa JC, Desai K, Tzivelekis S, Muston D, Khalid JM, et al. Prevalence of gastrointestinal stromal tumour (GIST) in the United Kingdom at different therapeutic lines: an epidemiologic model. *BMC Cancer* 2014;14:364.

244. Skálová A, Vanecek T, Sima R, Laco J, Weinreb I, Perez-Ordonez B, et al. Mammary analogue secretory carcinoma of salivary glands, containing the ETV6-NTRK3 fusion gene: a hitherto undescribed salivary gland tumor entity. *Am J Surg Pathol* 2010;34:599-608.

245. Luk PP, Selinger CI, Eviston TJ, Lum T, Yu B, O’Toole SA, et al. Mammary analogue secretory carcinoma: an evaluation of its clinicopathological and genetic characteristics. *Pathology* 2015;47:659-66.

246. Cancer Research UK. *Melanoma skin cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/incidence>.

247. UK PC. *Types of Pancreatic Cancer*. 2018. [cited 2019 16th December]. Available from: <https://www.pancreaticcancer.org.uk/information-and-support/facts-about-pancreatic-cancer/types-of-pancreatic-cancer/>.

248. Cancer Research UK. *Pancreatic cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/incidence#heading-Three>

249. Cancer Research UK. *Soft tissue sarcoma incidence statistics*. 2010. [cited 2019 27th February]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma/incidence#heading-Four>.

250. American Cancer Society. *Cancer Facts and Figures*. 2017. [cited 2020 24th January]. Available from: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html>.

251. Cancer Research UK. *Cervical cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/incidence#heading-Three>.

252. National Institute for Health and Care Excellence. *Pembrolizumab for previously treated oesophageal or gastrooesophageal junction cancer Draft scope*. London: NICE; 2018. Available from: <https://www.nice.org.uk/guidance/gid-ta10322/documents/draft-scope-post-referral>

253. Cancer Research UK. *Oesophageal cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence>.

254. Cancer Research UK. *Head and neck cancers statistics*. 2016. [cited 2019 Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers>.

255. Public Health England. *Glioblastoma in England 2007-2011*; 2016.

256. Cancer Research UK. *Ovarian cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence>

257. Farrimond J. *Analysis of the cancer registry combined database for use with the Brain and CNS Registry*: National Cancer Intelligence Network. Eastern Cancer Registration and Information Centre; 2010. Available from: <http://www.ncin.org.uk/search/brain+cns+registry>

258. Cancer Research UK. *Children's cancer incidence statistics*. Cancer Research UK; 2014. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/childrens-cancers>.

259. Brzeziańska E, Karbownik M, Migdalska-Sęk M, Pastuszak-Lewandoska D, Włoch J, Lewiński A. Molecular analysis of the RET and NTRK1 gene rearrangements in papillary thyroid carcinoma in the Polish population. *Mut Res/Fundamental Molecular Mechanisms Mutagenesis* 2006;599:26-35.

260. Cancer Research UK. *Prostate cancer incidence by stage at diagnosis* Cancer Research UK; 2016. [cited 2019 Jun]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Three>.

261. Cancer Research UK. *Kidney Cancer: Stages, types and grades*. Cancer Research UK; 2016. [cited 2019 Dec]. Available from: <https://www.cancerresearchuk.org/about-cancer/kidney-cancer/stages-types-grades/types-grades>.

262. Horowitz DP, Sharma CS, Connolly E, Gidea-Addeo D, Deutsch I. Secretory carcinoma of the breast: results from the survival, epidemiology and end results database. *Breast* 2012;21:350-53.

263. Jacob JD, Hodge C, Franko J, Pezzi CM, Goldman CD, Klimberg VS. Rare breast cancer: 246 invasive secretory carcinomas from the National Cancer Data Base. *J Surg Oncol* 2016;113:721-25.

264. Rushton L, Bagga S, Bevan R, Brown T, Cherrie J, Holmes P, et al. *The burden of occupational cancer in Great Britain*. London: Health & Safety Executive; 2010.

265. Cancer Research UK. *Cancer of the Unknown Primary (CUP): stages and types*. 2017. [cited 2019 16th December]. Available from: <https://www.cancerresearchuk.org/about-cancer/cancer-unknown-primary-cup>.

266. Demetri GD, Paz-Ares L, Farago AF, Liu SV, Chawla SP, Tosi D, et al. Efficacy and safety of entrectinib in patients with NTRK fusion-positive tumours: Pooled analysis of STARTRK-2, STARTRK-1, and ALKA-372-001. *Ann Oncol* 2018;29:ix175, LBA4.

267. Hyman D, van Tilburg C, Albert C, Tan D, Geoerger B, Farago A, et al. 445PD Durability of response with larotrectinib in adult and pediatric patients with TRK fusion cancer. *Ann Oncol* 2019;30:mdz244. 007.

268. National Institute for Health and Clinical Excellence. *Brigatinib for treating ALK-positive nonsmall-cell lung cancer after crizotinib [ID1328]. Committee papers*. London: NICE; 2019. Available from: <https://www.nice.org.uk/guidance/ta571/documents/committee-papers>

269. Georghiou T, Bardsley M. *Exploring the cost of end of life care*. London: Nuffield Trust; 2014.

# Appendices

Appendix 1: List of regulatory sources

|  |  |
| --- | --- |
| Workshop on site and histology - Independent indications in oncology | https://www.ema.europa.eu/en/events/workshop-site-histology-independent-indications-oncology |
| Workshop on single-arm studies in oncology | https://www.ema.europa.eu/en/events/workshop-single-arm-trials-oncology |
| Developing Targeted Therapies in Low-Frequency Molecular Subsets of a Disease Guidance for Industry | https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ UCM588884.pdf |
| Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics Guidance for Industry | https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/ Guidances/UCM621817.pdf |
| Guidance for Industry Expedited Programs for Serious Conditions – Drugs and Biologics | https://www.fda.gov/downloads/Drugs/Guidances/UCM358301.pdf |
| Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure | https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/ucm613636.htm |
| Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics | https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoyInformation/Guidance/UCM071590.pdf |
| Tissue agnostic therapies in oncology. Regulatory considerations for orphan drug designation | https://www.fda.gov/downloads/NewsEvents/MeetingsConferencesWorkshops/UCM598186.pdf |
| Essential considerations for successful qualification of novel methodologies | https://www.ema.europa.eu/documents/other/essential-considerations-successful-qualification-novel-methodologies\_en.pdf |
| Scientific guidelines on biostatistics (e.g. investigation of subgroups in clinical trials, multiplicity issues in clinical trials, extrapolation of efficacy and safety in medicine development, methodological issues in confirmatory clinical trials planned with an adaptive design) | https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety/biostatistics |
| Predictive biomarker-based assay development in the context of drug development and lifecycle | https://www.ema.europa.eu/en/predictive-biomarker-based-assay-development-context-drug-development-lifecycle |
| Guideline on the evaluation of anticancer medicinal products in man | https://www.ema.europa.eu/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5\_en.pdf |
| Appendix 4 to the guideline on the evaluation of anticancer medicinal products in man | https://www.ema.europa.eu/documents/scientific-guideline/evaluation-anticancer-medicinal-products-man-appendix-4-condition-specific-guidance-rev2\_en.pdf |

Appendix 2: Summary of trials

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Product** | **Indications** | **Clinical Evidence** | | | | |
| **Study** | **Clinical Outcomes** | **Study Design** | **Patient Population** | **Biomarker Screening** |
| **Merestinib** | Solid tumours | NCT02920996 | **MET cohort:**  **Primary**:   1. ORR (up to 2 years) [MET cohort]   **Secondary**:   1. OS rate (up to 2 years) [MET cohort] 2. PFS rate (2 years) [MET cohort] 3. Duration of response (up to 2 years) [MET cohort]   Safety ( 2 years) [all participants] | Design: Phase II (separate cohorts specified)  Estimated enrolment: n=25  Allocation approach: Non-randomised  Study start: 11 November 2016  Estimated primary completion: October 2020  Estimated study completion: March 2024 | NSCLC with MET exon 14 mutation or solid tumours with an NTRK rearrangement | **MET cohort**: must have a MET exon 14 mutation as confirmed by targeted NextGen Sequencing using the DFCI/BWH OncoPanel or another CLIA-certified method. Participants whose NSCLC specimens contain actionable genetic mutations/alterations (e.g. ALK/EGFR) should receive appropriate targeted therapies prior to enrollment in the trial.   * **NTRK cohort**: must have an NTRK1, 2, or 3 rearrangement as confirmed by targeted NextGen Sequencing using the DFCI/BWH OncoPanel or another CLIA-certified method. |
| **Avelumab plus Talazoparib** | Locally advanced (primary or recurrent) or metastatic solid tumours | NCT03330405 | **Primary**:   1. Safety (28 days) 2. Overall response (24 months)   **Secondary**:   1. Pharmacokinetics (15 days) 2. Immunogenicity (15 days) 3. Overall response (24 months) 4. PSA or CA-125 tumour marker (24 months) 5. PD-L1 levels (24 moths) 6. Time to tumour response (24 months) 7. Duration of response (24 months) 8. PFS (24 months) 9. OS (24 months) 10. PSA response (24 months) | Design: Phase Ib/II (separate cohorts specified)  Estimated enrolment: n=242  Allocation approach: Non-Randomised  Study start: 19 October 2017  Estimated primary completion: 28 March 2020  Estimated study completion: 28 March 2020 | Patients with locally advanced (primary or recurrent) or metastatic solid tumours, including non-small cell lung cancer (NSCLC), triple negative breast cancer (TNBC), hormone receptor positive (HR+) breast cancer, recurrent platinum sensitive ovarian cancer, urothelial cancer (UC), and castration resistant prostate cancer (CRPC). | N/A |
| Locally advanced or metastatic RAS-mutant solid tumours | NCT03637491 | **Primary**:   1. Safety (28 days) 2. Confirmed objective response (24 months)   **Secondary**:   1. Pharmacokinetics (12 months) 2. Objective response (24 months) 3. Time to tumour response (24 months) 4. Duration of response (24 months) 5. OS (24 months) 6. PFS (24 months) 7. Pharmacokinetics (3 months) 8. Biomarker levels (PD-L1, tumour mutational burden and DNA damage repair) | Design: Phase Ib/II  Estimated enrolment: n=127  Allocation approach: Randomised  Study start: 15 August 2018  Estimated primary completion: 1 May 2022  Estimated study completion: 7 November 2022 | Patients with locally advanced or metastatic KRAS- or NRAS-mutant non-small cell lung cancer, pancreatic ductal adenocarcinoma, or other KRAS- or NRAS-mutant solid tumours. | Positive KRAS or NRAS mutation status as determined using a validated test performed in a College of American Pathologists (CAP)/Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory (or other comparable local or regional certification) |
| Solid tumours with a BRCA or ATM defect | NCT03565991 | **Primary**:   1. Confirmed objective response (24 months)   **Secondary:**   1. Confirmed objective response by the investigator (24 months) 2. Time to tumour response (24 months) 3. Duration of response (24 months) 4. PFS (24 months) 5. OS (24 months) 6. PSA and CA-125 response (24 months) 7. Circulating tumour cell level 8. Pharmacokinetics (up to 24 months) 9. Immunogenicity (up to 24 months) | Design: Phase II (separate cohorts specified)  Estimated enrolment: n=200  Allocation approach: Non-Randomised  Study start: 18 June 2018  Estimated primary completion: 8 March 2021  Estimated study completion: 2 December 2022 | Patients with locally advanced or metastatic solid tumours with a BRCA or ATM defect | Not given |
| **LOXO-195** | Solid tumours | NCT03215511 | **Primary**:   1. MTD 2. Best overall response (up to 2 years)   **Secondary:**   1. Safety (up to 24 months) 2. Overall response (24 months) 3. Pharmacokinetics (5 months) 4. Duration of response (up to 24 months) 5. PFS (up to 24 months) 6. OS (up to 24 months) 7. Clinical benefit rate (up to 24 months | Design: Phase I/II (separate cohorts specified)  Estimated enrolment: n=93  Allocation approach: Non-Randomised  Study start: 10 July 2017  Estimated primary completion: August 2019  Estimated study completion: 18 May 2026 | Patients with unresectable or metastatic solid tumours and progressed or intolerant to prior Trk-inhibitor. | NTRK gene fusions will be identified via a CLIA certified (or equivalent) laboratory. Exception: Patients with Infantile Fibrosarcoma (IFS) and congenital mesoblastic nephroma (CMN) may be enrolled based on ETV6+ FISH test without identifying NTRK3 |
| **TPX - 0005** | Advanced solid tumours harbouring ALK, ROS1, or NTRK1-3 rearrangements | NCT03093116 | **Primary:**   1. Maximum tolerated dose (28 days of first dose) 2. Recommended Phase 2 Dose (28 days of first dose) 3. ORR (2 to 3 months after treatment start)   **Secondary:**   1. Effect of food on AUC (2 to 3 months after treatment start) 2. Time to response (3 years) 3. DoR (3 years) 4. CBR (3 years) 5. PFS (3 years) 6. OS (3 years) 7. Intracranial ORR (3 years) 8. CNS PFS (3 years) | Design: Phase I/II (separate cohorts specified)  Estimated enrolment: 450 participants  Allocation approach: Non-randomised  Study start date: 27 February 2017  Estimated primary completion date: January 2021  Estimated study completion date: December 2021 | Patients >= 18 years with histologically or cytologically confirmed confirmed locally advanced or metastatic solid tumour (including non-Hodgkin lymphoma) harmoring ALK, ROS1, NTRK1-3 gene rearrangement. | Not given |
| **Sunitinib** | Refractory solid tumours | NCT02691793 | **Primary:**   1. PFS (24 months)   **Secondary**   1. ORR (24 months) 2. TTP (24 months) 3. OS (24 months) 4. Number of subjects with AE (24 months) | Design: Phase 4  Estimated enrolment: 25 participants  Allocation approach: Non-Randomised  Study start date: 20 November 2017  Estimated primary completion: December 2018  Estimated study completion: December 2018 | Patients >=19 years with RET fusion positive or FGFR2 fusion/other FGFR mutation refractory solid tumour and/or specific sensitivity to Sunitinib by Avatar scan that has progressed following standard therapy or that has not responded to standard therapy or for which there is no standard therapy. | Not given |
| Advanced Rare Tumours | NCT01396408 | **Primary**   1. OR (every 4 weeks)   **Secondary**   1. DoR/TTP/PFS/OS (48 months) 2. Translational research (48 months) 3. Safety (daily up to 4 weeks after treatment) | Design: Phase II  Estimated enrolment: 137  Allocation approach: Non-randomised  Study start date: 14 July 2011  Primary completion date: July 2015  Study completion date: December 2019 | Patients >=16 years with histologically or cytologically confirmed advanced rare tumours:   * Vascular sarcomas * Clear cell ovary carcinomas * Thyroid carcinoma * Neuro-endocrine tumours * Adrenocorticocarcinoma * Thymic carcinoma * Hepatocellular carcinoma | Not given |
| **Olaparib** | Advanced (unresectable and/or metastatic) cancers | NCT03742895 | **Primary:**   1. ORR (up to 53 months)   **Secondary:**   1. DOR (up to 53 months) 2. OS (up to 53 months) 3. PFS (up to 53 months) 4. AEs (up to 53 months) 5. Time to earliest progression by cancer antigen-125 (up to 53 months) | Design: Phase II  Estimated enrolment: 370 participants.  Allocation approach: Non-Randomised  Study start date: 12 December 2018  Primary completion date: 30 April 2023  Study completion date: 30 April 2023 | Patients >= 18 years with multiple types of advanced cancer (unresectable and/or metastatic) that: 1) have progressed or been intolerant to standard of care therapy; and 2) are positive for homologous recombination repair mutation (HRRm) or homologous recombination deficiency (HRD). | Not given |
| Advanced solid tumours, non-Hodgkin lymphoma, or histiocytic disorders with defects in DNA Damage Repair Genes | NCT03233204 | **Primary:**   1. ORR (up to 4 years)   **Secondary:**   1. PFS (up to 4 years) 2. Toxicity (up to 4 years) 3. PK (up to 4 years)   **Other**   1. Change in tumour genomic profile (up to 4 years) | Design: Phase II  Estimated enrolment: 49 participants  Allocation approach: Non-Randomised  Study start date: 24 July 2017  Estimated primary completion date: 30 September 2024  Estimated Study Completion date: 30 September 2024 | Patients 1-21 years with solid tumours, non-Hodgkin lymphoma, or histiocytic disorders with defects in deoxyribonucleic acid (DNA) damage repair genes that have spread to other places in the body and have come back or do not respond to treatment | Not given |
| Glioma, cholangiocarcinoma, or solid tumours with IDH1 or IDH2 mutations | NCT03212274 | **Primary:**   1. ORR (up to completion of course 8)   **Secondary:**   1. PFS (up to 1 year) 2. AE (up to 1 year) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 145 participants  Allocation approach: Non-Randomised  Study Start Date: March 30, 2018  Primary Completion Date: July 31, 2019  Study Completion Date: July 31, 2019 | Patients >=18 years diagnosed with a glioma, cholangiocarcinoma or other solid malignant tumour that has progressed despite standard therapy, or for which no effective standard therapy exists, with biopsy-confirmed evidence of an IDH1 or IDH2 mutation associated with neomorphic activity of the encoded proteins;; only specific mutations that lead to a neomorphic phenotype will be eligible for enrollment, and include IDH1: R132V, R132G, R132S, R132L, R132C and R132H; IDH2: R140W, R140L, R140Q, R172W, R172G, R172S, R172M, R172K | Patients must have IDH1 or IDH2 mutation which must be detected in a clinical accredited laboratory using a Food and Drug Administration (FDA)-approved molecular test or a validated deoxyribonucleic acid (DNA)-based assay conducted in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory |
| Advanced cancer with a confirmed BRCA1 and/or BRCA2 mutation | NCT01078662 | **Primary:**   1. Tumour response rate (maximum up to 29 months)   **Secondary:**   1. ORR (up to 29 months) 2. PFS (up to 29 months) 3. OS (up to 29 months) 4. OS-12 months 5. DOR (up to 29 months) 6. Disease control rate at week 16 | Design: Phase II  Estimated enrolment: 299 participants  Allocation approach: Non-Randomised  Study Start Date: February 21, 2010  Primary Completion Date: July 31, 2012  Study Completion Date: December 31, 2019 | Patients >=18 years with malignant solid tumours for which no standard treatment exists and with confirmed documented deleterious or suspected deleterious BRCA mutation (ovarian, breast, prostate, pancreatic, advanced tumours). | Not given |
| Patients with tumours harbouring damaging mutations in Homologous - DNA repair (HDR) genes or mutations such as ATM, CHK2, MRN (MRE11/NBS1/RAD50), CDKN2A/B and APOBEC | NCT02576444 | **Primary:**   1. ORR (change from baseline to 16 weeks) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 64 participants  Allocation approach: Non-Randomised  Study Start Date: November 2015  Primary Completion Date: March 2020  Study Completion Date: March 2020 | Patients >=18 years with histologically documented metastatic cancer (not hematologic malignancies). | Molecular testing or appropriate IHC results from CLIA-certified laboratories used for patient eligibility should be obtained from the most recent tumour biopsy (baseline tumour biopsies and on-progression tumour biopsies are optional) |
| Relapsed or refractory tumour | NCT02813135 | **Primary:**   1. ORR (56 days) 2. TTP (56 days) | Phase I/II basket  Estimated enrolment: 397 participants  Allocation approach: Non-Randomised  Study start date: 3 August 2016  Estimated primary completion: January 2022  Estimated study completion: January 2022 | Patients <18 years with haematologic or solid tumour malignancy that has progressed despite standard therapy, or for which no effective standard therapy exists | Patient must have had advanced molecular profiling (i.e. WES/WGS +/- RNAseq) of their recurrent or refractory tumour i.e. at the time of disease progression/relapse; exceptionally patients with advanced molecular profiling at diagnosis may be allowed |
| Advanced cancer whose tumour harbors a genomic variant known to be a drug target or to predict sensitivity to a drug | NCT02693535 | **Primary:**   1. ORR (at 16 weeks of treatment)   **Secondary**   1. OS (up to 3 years) | Phase II (separate cohorts specified)  Estimated enrolment: 2980  Allocation approach: Non-Randomised  Study start date: March 2016  Estimated primary completion: December 2021 | Patients =12 years with Histologically-proven locally advanced or metastatic solid tumour, multiple myeloma or B cell non-Hodgkin lymphoma who is no longer benefiting from standard anti-cancer treatment or for whom, in the opinion of the treating physician, no such treatment is available or indicated | Results must be available from a genomic test or immunohistochemistry (IHC) test for protein expression performed in a Clinical Laboratory Improvement Amendments (CLIA)-certified, College of American Pathologists (CAP) -accredited, New York State accredited (for labs offering services to residents of NY) laboratory that has registered the test with the National Institutes of Health (NIH) Genetic Test Registry or has established an integration with the TAPUR platform. The genomic or IHC test used to qualify a patient for participation in TAPUR may have been performed on any specimen of the patient's tumour obtained at any point during the patient's care at the discretion of the patient's treating physician. Genomic assays performed on cell-free DNA in plasma ("liquid biopsies") will also be acceptable if the genomic analysis is performed in a laboratory that meets the criteria described above. |
| Relapsed or Refractory Advanced Solid Tumours, Non-Hodgkin Lymphomas, or Histiocytic Disorders | NCT03155620 | **Primary:**   1. ORR (up to 4 years)   **Secondary**   1. Safety (up to 4 years) 2. PFS (up to 4 years) 3. PK (up to 4 years)   **Other**   1. Genomics (up to 4 years) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 1000 participants  Allocation approach: Non-Randomised  Study start date: 24 July 2017  Estimated primary completion date: 30 September 2027  Estimated study completion date: 30 September 2027 | Paediatric patients with solid tumours, non-Hodgkin lymphomas, or histiocytic disorders that have progressed following at least one line of standard systemic therapy and/or for which no standard treatment exists that has been shown to prolong survival | Not given |
| Cancers of unknown primary site. | NCT03498521 | **Primary:**   1. PFS (up to 48 months)   **Secondary:**   1. OS (up to 48 months) 2. ORR 3. Duration of benefit (up to 48 months) 4. AE (up to 48 months) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 790 participants  Allocation approach: Randomised  Study start date: 10 July 2018  Estimated primary completion: 25 June 2021  Estimated study completion: 25 June 2022 | Patients >=18 years with histologically-confirmed cancer of unknown primary site (CUP)(non-specific subset) according to criteria from the European Society for Medical Oncology, version 1 (ESMO v1), who have achieved disease control after 3 cycles of first-line platinum doublet induction chemotherapy. | Not given |
| Non-Hodgkin lymphopma, multiple myeloma, advanced solid tumours | NCT03297606 | **Primary:**   1. ORR (4 years)   **Secondary**   1. AE (up to 4 years) 2. PFS (up to 4 years) | Design: Phase II basket  Estimated enrolment: 720 participants  Allocation approach: Non-Randomised | Patients >=18 years with a histologically‐proven incurable metastatic solid tumour (excluding primary brain tumours), multiple myeloma or B cell non‐ Hodgkin lymphoma (excluding CLL, SLL and HCL), for whom there is no standard treatment known to prolong life, or who has refused such treatment | Not given |
| Refractory solid tumours | NCT03239015 | **Primary:**   1. ORR (2 months)   **Secondary**   1. PFS (2 months) 2. OS (1 month) 3. AE (1 month) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 60 participants  Allocation approach: Non-Randomised  Study start date: 1 January 2017  Estimated primary completion 30 June 2018  Estimated study completion: 31 December 2019 | Patients 18-75 years with Malignant solid tumours diagnosed histologically. Common solid tumour patients have no any standard choice after multiple line of therapy; Rare solid tumour did not have any standard recommended treatment | NGS |
| Advanced solid tumours | NCT02029001 | **Primary:**   1. Induction progression-free rate 2. PFS (up to 36 months)   **Secondary**   1. ORR (over induction period) 2. OS 3. QoL (QLQ-C30) 4. Safety   **Other**   1. DoR 2. Cost-effectiveness | Design: Phase II (separate cohorts specified)  Estimated enrolment: 560 participants  Allocation approach: Randomised  Study start date: March 2014  Primary completion date: January 2020  Study completion date: October 2022 | Patients >=18 years with histologically or cytologically confirmed diagnosis of metastatic or locally advanced and unresectable solid tumour of any type, not amenable to curative treatment. Concerning primitive tumours of the central nervous system (CNS), all histological types of malignant tumours (including parenchymal and meningeal tumours) are eligible. | Not given |
| Advanced solid tumour, multiple myeloma, or non-Hodgkin lymphoma | NCT02925234 | **Primary:**   1. % Patients treated based on molecular profile (6 months after treatment initiation) 2. Objective tumour response (6 months) 3. Stable disease (6 months) 4. AE >=G3 (6 months)   **Secondary**   1. PFS (up to 1 year) 2. OS (up to 1 year) 3. Duration of treatment (6 months) | Design: Phase II (separate cohorts specified)  Estimated enrolment: 400 participants  Allocation approach: Non-Randomised  Study start date: August 2016  Estimated primary completion: August 2019  Estimated study completion: December 2019 | Patients >=18 years with a histologically-proven locally advanced or metastatic solid tumour, multiple myeloma or B cell non-Hodgkin lymphoma who is no longer benefitting from standard anti-cancer treatment or for whom no such treatment is available or indicated. | Results must be available from a tumour genomic or protein expression test. Eligible tests may include any of the following technologies: fluorescence in situ hybridization (FISH), polymerase chain reaction (PCR), comparative genomic hybridization (CGH), next generation sequencing (NGS) or immunohistochemistry (IHC). The test may have been performed on the primary tumour or a metastatic deposit, in a diagnostic laboratory or within the context of another CPCT study, and must reveal a potentially actionable variant. |
| **LOXO-292** | Advanced Solid Tumours, RET Fusion-Positive Solid Tumours, and Medullary Thyroid Cancer | NCT03157128 | **Primary:**   1. Dosage 2. ORR (up to 2 years)   **Secondary**   1. AE (2 years) 2. ORR (2 years) 3. ORR/DOR/CBR/PFS/OS (2 years)   **Other**   1. Genomics 2. HRQoL (QLQ-C30) | Design: Phase I/II  Estimated enrolment: 870  Allocation approach: Non-randomised  Study start date: 9 May 2017  Estimated primary completion date: August 2019  Estimated study completion date: December 2019 | Patients >=12 years with advanced solid tumours, including RET fusion-positive solid tumours, Medullary Thyroid Cancer, and other tumours with RET activation. | Not given |
| **Epacadostat with pembrolizumab** | Advanced or metastatic solid tumours | NCT03085914 | **Primary:**   1. Phase 1: Safety and tolerability (up to 27 months) 2. Phase 2: ORR (up to 24 months)   **Secondary**   1. Phase 1: ORR (up to 24 months) 2. Phase 2: safety and tolerability (up to 27 months) 3. DOR (up to 24 months) 4. PFS (up to 24 months) | Design: Phase I/II (separate cohorts specified)  Estimated enrolment: 70 participants  Allocation approach: Non-Randomised  Study start date: 2 May 2017  Estimated primary completion date: October 2019  Estimated study completion date: January 2020 | Patients >= 18 years of age histologically or cytologically confirmed diagnosis of selected advanced or metastatic solid tumours. | Not given |
| Advanced or metastatic malignancies | NCT03277352 | **Primary:**   1. Phase 1: AE (up to 18 months) 2. Phase 2: ORR/CRR (up to 18 months)   **Secondary**   1. Disease Control Rate (18 months) 2. DOR (18 months) 3. Duration of disease control (18 months) 4. PFS (18 months) 5. OS (at 1 and 2 years) | Design: Phase I/II  Estimated enrolment: 10 participants  Allocation approach: Non-randomised  Study start date: 21 November 2017  Estimated primary completion date: March 2020  Estimated study completion date: May 2020 | Patients >= 18 years with locally advanced or metastatic disease; locally advanced disease must not be amenable to resection with curative intent. | Not given |
|  | Advanced solid tumours | NCT02959437 | **Primary:**   1. Phase I: AE (up to 18 months) 2. Phase II: ORR (up to 18 months)   **Secondary**   1. Phase I ORR 2. Phase II AE 3. PFS (up to 18 months) 4. DOR (up to 18 months) | Design: Phase I/II (separate cohorts specified)  Estimated enrolment: 70 participants  Allocation approach: Non-Randomised  Study start date: 26 January 2017  Estimated primary completion date:15 February 2019  Estimated study completion date: 9 July 2020 | Patients >=18 years with histologically or cytologically confirmed advanced or metastatic solid tumours that have failed prior standard therapy (disease progression; subject refusal or intolerance is also allowable). Part 1 is a dose-escalation assessment to evaluate the safety and tolerability of the combination therapies. Once the recommended doses have been determined, subjects with previously treated NSCLC, microsatellite-stable colorectal cancer (CRC), head and neck squamous cell carcinoma, urothelial carcinoma, and melanoma will be enrolled into expansion cohorts in Part 2 | Not given |
| **Durvalumab with Tremelimumab** | Advanced solid and haematological cancers | NCT03837899 | **Primary:**   1. Recommended Phase II dose in patients receiving chemotherapy (15 months) 2. Safety and tolerability (up to 4 years) 3. ORR (up to 4 years)   **Secondary:**   1. PK (15 months) | Design: Phase I/II  Estimated enrolment: 158 participants  Allocation approach: Non-Randomised  Study start date: 7 March 2019  Estimated primary completion date:18 October 2021  Estimated study completion date: 20 March 2023 | Paediatric patients (up to 17 years) with solid tumours, which must have progressed or be refractory to standard therapies. | Not given |
| Advanced rare solid tumours | NCT02938793 | **Primary:**   1. Antitumour Activity (24 months) 2. AEs (24 months)   **Secondary:**   1. Expression of PD-1 | Design: Phase II  Estimated enrolment: 50 participants  Allocation approach: Non-Randomized  Study start date: 1 December, 2016  Estimated primary completion date: 28 February 2020  Estimated study completion date: 31 December 2021 | Adult patients with diagnosis of a rare advanced solid malignancy meeting EORTC criteria. Subjects must have failed or been ineligible for standard treatment options, if available | Not given |
| Advanced malignancies | NCT02978482 | **Primary:**   1. Plasma concentration 2. AEs 3. ORR (12 months after last patient is dosed or withdrawn, or study is discontinued)   Secondary   1. Anti-drug antibody 2. CR/PR/Stable disease/Progressive disease (6 months after last patient is dosed) 3. OS (12 months after last evaluable patient is first dosed) | Phase I/II (separate cohorts specified)  Estimated enrolment: 26 participants  Allocation approach: Non-Randomised  Study start date: 1 December 2016  Primary completion date: 28 July 2018  Study completion date: 13 May 2019 | Chinese adult patients with histologically or cytologically confirmed advanced and/or metastatic solid tumours other than HCC refractory or intolerable to existing standard of treatment. | Not given |
| Advanced solid malignancies | NCT03084471 | **Primary:**   1. Safety: AEs   **Secondary**   1. Safety: treatment-related adverse events and treatment discontinued/interrupted 2. OS (up to 5 years following date of first patient initiation) | Phase I/II  Estimated enrolment: 1,200 participants  Allocation approach Non-Randomised  Study start date: 17 April 2017  Estimated primary completion date: 26 March 2023  Estimated study completion date: 26 March 2023 | Adult patients with a life expectancy of at least 12 weeks and no prior exposure to anti PD-1 or anti PD-L-1. | Not given |
| Somatically hyper mutated recurrent solid tumours | NCT03911557 | **Primary**   1. Time-to-progression ratio (2 years)   **Secondary**   1. PFS (2 years) | Phase II basket  Estimated enrolment 48 participants  Allocation approach: Non-Randomised  Estimated study start date: May 2019  Estimated primary completion date: June 2023  Estimated study completion date: June 2024 | Adult patients with relapsed/refractory solid tumour patients (not previously treated with anti-PD-1/PD-L1 or anti-CTLA-4 immunotherapy), whose tumours expressed a high tumour mutational burden (TMB) or moderate TMB. | Next generation sequencing |
| Advanced rare tumours | NCT02879162 | **Primary**   1. ORR (48 months)   **Secondary**   1. AEs 2. Time-to-progression (48 months) 3. PFS (38 months) 4. Response duration (48 months) | Design: Phase II  Estimated enrolment: 160 participants  Allocation approach: Non-Randomised  Study start date: 19 October 2016  Estimated primary completion date: August 2020  Estimated study completion date: December 2020 | Patients 16 years and older with histologically and/or cytologically confirmed cancer that advanced/metastatic/recurrent or unresectable and for which no curative therapy exists. The list includes: salivary carcinoma, carcinoma of unknown primary site with tumour infiltrating lymphocytes and/or expressing PD-L1, mucosal melanoma, acral melanoma, osteosarcoma, undifferentiated pleomotphic sarcoma, clear cell carcinoma of the ovary, squamous cell carcinoma of the anal canal. | Not given |
| **BLU-667** | Solid tumours | NCT03037385  (EudraCT Number 2016-004390-41) | **Primary**:   1. Tolerability (12 months) 2. Safety (24 months) 3. ORR (up to 2 years)   **Secondary**   1. DOR, PFS, OS (up to 2 years) 2. RET gene status and correlation between RET gene status and ORR, DOR and DCR (up to 2 years) 3. Pharmacokinetics (4 months) 4. Pharmacodynamics (12 months) | Design: Phase I/II (separate cohorts specified)  Estimated enrolment: n=360  Allocation approach: Non-Randomised  Study start: 15 December 2015  Estimated primary completion: 28 August 2023  Estimated study completion: 28 August 2023 | Enrolling patients with medullary thyroid cancer, RET-altered NSCLC and other RET-altered solid tumours | Not given |
| **Atezolizumab** | Solid tumours | NCT02458638 | **Primary:**   1. NPR (18 weeks)   **Secondary**   1. NPR (24 weeks) 2. ORR (24 weeks) 3. BOR (24 weeks) 4. DoR (24 weeks) 5. PFS (24 weeks) 6. TTP (24 weeks) 7. OS (24 weeks) 8. Safety (24 weeks) | Design: Phase II  Estimated enrolment: n=477  Allocation approach: Non-Randomised  Study start: 15 December 2015  Estimated primary completion: 14 April 2018  Estimated study completion: 13 December 2019 | Enrolling adults patients with advanced solid tumours who have received at least one line of prior systemic therapy or for which no alternative therapy to prolong survival exists. | N/A |
| **Cobimetinib** | Solid tumours | NCT02639546 | **Primary**:   1. Safety (1 month) 2. Dosing (1 month) 3. Pharmacokinetics (12 months) 4. Percentage of patients with OR (6.75 years) 5. PFS (6.75 years)   **Secondary**   1. DOR (up to 6.75 years) 2. OS (up to 6.75 years) | Design: Phase I/II  Estimated enrolment: n=50  Allocation approach: Non- Randomised  Study start: 20 May 2016  Estimated primary completion: 21 February 2023  Estimated study completion: 21 February 2023 | Enrolling paediatric and young adult participants with solid tumours with known or potential kinase pathway activation (RAS/RAF/MEK/ERK pathway involvement) for which standard therapy has proven to be ineffective or intolerable or for which no curative standard-of-care treatment options exist. | N/A |
| **Crizotinib** | Solid and liquid tumours | NCT01524926  (CREATE) | **Primary**:   1. Antitumour activity   **Secondary**   1. Safety 2. PFS 3. DCR 4. OS 5. DOR | Design: Phase II (separate cohorts specified)  Estimated enrolment: n=582  Allocation approach: Non-Randomised  Study start: September 2012  Estimated primary completion: December 2017  Estimated study completion: December 2018 | Enrolling patients with advanced tumours induced by causal alterations of ALK and/or MET | Not given |
| Solid tumours | NCT02034981 | **Primary**:   1. ORR (8 weeks)   **Secondary**   1. Safety (up to 2.5 years) 2. DCR (4 months) 3. DOR 4. PFS 5. OS | Design: Phase II (separate cohorts specified)  Estimated enrolment: n=246  Allocation approach: Non- Randomised  Study start: August 2013  Estimated primary completion: June 2018  Estimated study completion: July 2022 | Enrolling patients harbouring an alteration on ALK, MET or ROS1 | Not given |

AEs, adverse events; BOR, best overall response; DCR, disease control rate; DOR, duration of response; HRQoL, health-related quality of life; NPR, non-progression rate; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PK/PD, pharmacokinetics/pharmacodynamics; TTP, time to progression

Appendix 3: MEDLINE search strategy

Search Strategy (March 2019):

1 \*Neoplasms/

2 (cancer$ or neoplasm$ or tumour$ or tumour$ or malignan$ or oncology or lymphoma$ or sarcoma$ or melanoma$ or myeloma$ or carcinoma$).tw.

3 1 or 2

4 tumour response$.tw.

5 tumour response$.tw.

6 objective response$.tw.

7 ORR.tw.

8 "duration of response$".tw.

9 dor.tw.

10 response rate$.tw.

11 complete response$.tw

12 overall response$.tw

13 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12

14 3 and 13

15 Regression analysis/

16 regression.tw.

17 relationship.tw.

18 correlation.tw.

19 prediction.tw.

20 association.tw.

21 15 or 16 or 17 or 18 or 19 or 20

22 14 and 21

23 endpoint$.tw.

24 end point$.tw.

25 (surrogate or surrogacy).tw.

26 23 or 24 or 25

27 22 and 26

28 progression-free survival/

29 "progression free survival".tw.

30 "overall survival".tw.

31 (pfs or os).tw.

32 "time to progression".tw.

33 ttp.tw.

34 28 or 29 or 30 or 31 or 32 or 33

35 27 and 34

36 limit 35 to (english language and humans)

Appendix 4: Table of study characteristics (ordered by cancer type then author)

| **Reference** | **Cancer** | **Surrogate outcome** | **Final outcome** | **Stage** | **Line** | **Treatment** | **N studies** | **N patients** | **Study types** | **Publication/ search years** | **Data type** | **Response criteria** | **Absolute association** | **Treatment effect association** | **STE reported** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Agarwal 201780 | Acute myeloid leukemia | ORR  CR | OS | Various | 1st | Systemic | 20† | NR | RCT + SA | 2004-2016 | AD | NR | Y |  |  |
| Moriwaki 2016116 | Biliary tract | ORR | OS | Advanced | 1st | Chemo | 17† | 2040 | RCT | Up to 2015 | AD | NR |  | Y |  |
| Bruzzi 200583 | Breast | ORR | OS | Metastatic | All | Chemo | 10 | 2126 | RCT | 1991-2001 | IPD | WHO (8), ECOG (1), NR (1) |  | Y |  |
| Burzykowski 200884 | Breast | ORR | PFS  OS | Metastatic | 1st | Chemo | 11 | 3953 | RCT | 1999-2008 | IPD | WHO |  | Y |  |
| Hackshaw 200596 | Breast | ORR  CR | OS | Metastatic | 1st | Chemo | 42\* | 9163 | RCT | 1966-2005 | AD | NR |  | Y |  |
| Liu 2016111 | Breast | ORR\* | OS | Metastatic | 2nd + 3rd | Chemo | 24 | 8617 | RCT | 1999 to 2014 | AD | NR | Y |  |  |
| Petrelli 2014124 | Breast | ORR | OS | Metastatic or advanced | 1st | Targeted + chemo | 20† | 10138† | RCT | 2000 to 2012 | AD | NR | Y |  |  |
| Buyse 200085 | Colorectal | ORR | OS | Advanced | 1st | Chemo | 25 | 3791 | RCT | Collected 1990-1996 | IPD | WHO |  | Y |  |
| Ciani 201586 Elia 201893 | Colorectal | ORR | PFS  OS | Advanced or metastatic | All | Systemic | 33 | NR | RCT | 2003-2013 | AD | RECIST or WHO |  | Y | Y |
| Colloca 2016b89 | Colorectal | ORR  DoR | OS | Metastatic | 1st | Bevacizumab + chemo | 11 | NR | RCT | 2000-2014 | AD | RECIST |  | Y |  |
| Giessen 201595 | Colorectal | ORR | OS | Metastatic | 2nd | Chemo | 22 | 10509 | RCT | 2000-2013 | AD | RECIST (17), WHO (5) | Y |  |  |
| Cremolini 201791 | Colorectal | ORR | OS | Metastatic | 2nd | Targeted | 20\* | 7571 | RCT | To 2015 | AD | NR |  | Y |  |
| Johnson 2006106 | Colorectal | ORR | OS | Metastatic | 1st | Chemo | 146† | 35337† | RCT | To 2005 | AD | NR (very few RECIST) |  | Y |  |
| Louvet 2001112 | Colorectal | ORR | PFS  OS | Metastatic | 1st | Various | 29 | 13498 | RCT | 1990 to 2000 | AD | NR | Y |  |  |
| Sidhu 2013133 | Colorectal | ORR | OS | Metastatic | 1st (most) | Chemo +/- targeted | 24† | 20438† | RCT | 2000 to 2011 | AD | NR |  | Y |  |
| Tang 2007135 | Colorectal | ORR | OS | Metastatic | 1st | Chemo | 39 | 18668 | RCT | 1990 to 2005 | AD | NR | Y | Y |  |
| Tsujino 2010137 | Colorectal | ORR | PFS  OS | Advanced | NR | Targeted | 7 | NR | RCT | Up to 2009 | AD | NR |  | Y |  |
| Ichikawa 2006102 | Gastric | ORR | TTP  OS | Advanced | 1st | Chemo | 25 | 4593 | RCT | NR | AD | WHO, SWOG, RECIST, Japan | Y |  |  |
| Shitara 2014130 | Gastric | ORR\* | OS | Advanced | 2nd + 3rd | Chemo | 64 | 4286 | RCT + SA | 2002 to 2012/2013 | AD | NR | Y |  |  |
| Pang 2018121 | Gastroesophageal | ORR\*  CR | OS | Advanced | 1st + 2nd | Targeted | 18 | 7892 | RCT | Up to 2018 | AD | RECIST | Y |  |  |
| Han 201498 | Glioblastoma | ORR | OS | Unclear | Various | Various | 91† | 7125† | RCT + SA | 1991-2012 | AD | NR ("standard criteria") | Y |  |  |
| Blumenthal 201782 | Lung (NSCLC) | ORR | PFS  OS | Metastatic | Various | Chemo, immune or targeted | 25 | 20013† | RCT | 2003-2016 | AD | RECIST or WHO |  | Y |  |
| Blumenthal 201581 | Lung (NSCLC) | ORR | PFS  OS | Metastatic | Various | Chemo or targeted | 14 | 12567† | RCT | 2003-2013 | AD | RECIST (11) or WHO (3) |  | Y |  |
| Hashim 201899 | Lung (NSCLC) | ORR | OS | Advanced | 2nd + subsequent | Various | 140 | 41725 | RCT | To 2016 | AD | NR |  | Y | Y |
| Hotta 2015100 | Lung (NSCLC) | ORR | OS | Advanced | Various | Targeted | 18 | 7633† | RCT | 2003-2014 | AD | NR |  | Y |  |
| Ito 2019105 | Lung (NSCLC) | ORR | PFS  OS | Advanced | Various | Immune checkpoint inhibitors (PD-(L)1) | 7 | 3752† | RCT | NR | AD | NR | Y | Y |  |
| Johnson 2006106 | Lung (NSCLC) | ORR | OS | Advanced | 1st | Chemo | 191† | 44125† | RCT | To 2005 | AD | NR (very few RECIST) |  | Y |  |
| Li 2019109 | Lung (NSCLC) | ORR\*  CR | OS | Advanced | 1st + 2nd | Immune checkpoint inhibitors | 5† | 4803† | RCT | Up to 2018 | AD | RECIST | Y |  |  |
| Li 2012110 | Lung (NSCLC) | ORR | OS | Advanced | 1st + 2nd | Targeted | 60 | 9903 | RCT + SA | Up to 2011 | AD | RECIST (52), WHO (10) | Y |  |  |
| Nakashima 2016118 | Lung (NSCLC) | ORR | OS | Advanced, locally advanced and recurrent | 1st | Chemo | 44 | 22709 | RCT | 2005 to 2015 | AD | RECIST |  | Y |  |
| Ritchie 2018125 | Lung (NSCLC) | ORR\* | PFS  OS | Advanced | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 8 | NR | RCT | 2000 to 2017 | AD | NR | Y | Y |  |
| Roviello 2017127 | Lung (NSCLC) | ORR | PFS  OS | Unclear | Various | Immune checkpoint inhibitors | 7\* | 3369\* | RCT | Up to 2017 | AD | RECIST or mWHO |  | Y |  |
| Sekine 1999128 | Lung (NSCLC) | ORR | OS | Unclear | Various | Chemo | 42 | 1935 | SA +1 RCT | 1988-1997 | AD | WHO | Y |  |  |
| Shukuya 2016131 | Lung (NSCLC) | ORR | OS | Advanced | All | a) Immune checkpoint inhibitors (PD-(L)1)  b) Chemo (docetaxel) | a) 10†  b) 22† | NR | RCT + SA | 2012 to 2016 | AD | RECIST (most) | Y |  |  |
| Tsujino 2010137 | Lung (NSCLC) | ORR | PFS  OS | Advanced | NR | Targeted | 6 | NR | RCT | Up to 2009 | AD | NR |  | Y |  |
| Tsujino 2009136 | Lung (NSCLC) | ORR | PFS  OS | Advanced | NR | Targeted | 28 | 6171 | RCT + SA | To 2007 | AD | RECIST (21), WHO (9) | Y |  |  |
| Vidaurre 2009138 | Lung (NSCLC) | ORR\* | PFS  OS | Advanced, locally advanced, unresectable or metastatic | NR | Chemo or targeted | 35 | NR | RCT + SA | 2006 to 2008 | AD | NR | Y |  |  |
| Foster 201194 | Lung (SCLC) | ORR  CR | OS | Extensive-stage | 1st | Chemo | 3 RCTs (32 centres) | 596† | RCT | Trials initiated 1987-1999 | AD | NR (CR=disappearance; PR ≥50% reduction |  | Y |  |
| Hotta 2009101 | Lung (SCLC) | ORR | OS | Extensive disease | 1st | Chemo | 48 | 8779 | RCT | 1990-2008 | AD | WHO (23), ECOG (2), RECIST (1), Japan (1), or NR |  | Y |  |
| Nickolich 2014119 | Lung (SCLC) | ORR  CR  PR | PFS  OS | Limited or extensive disease | 1st + 2nd + maintenance | Various | 66† | 8471† | RCT + SA | 1983 to 2010 | AD | NR | Y |  |  |
| Mangal 2018115 (myeloma) | Multiple myeloma | ORR\*  CR  VGPR or CR | PFS | Relapsed / refractory | 2nd + subsequent | Various | 79† | 13322† | RCT + SA | 1999 to 2016 | AD | IMWG | Y |  |  |
| Imaoka 2019104 | Neuroendocrine | ORR | PFS | Advanced | Various | Systemic | 22 | 1310 | RCT + SA | 1996-2016 | AD | RECIST (20), WHO (2) | Y |  |  |
| Imaoka 2017103 | Neuroendocrine | ORR | OS | Advanced | Various | Systemic | 20 | 2530 | RCT + SA | 1996-2016 | AD | NR | Y |  |  |
| Lee 2011108 | NHL (aggressive) | CR | PFS  OS | Unclear | 1st | Chemo | 36† | 16103† | RCT | 1990-2009 | AD | NR |  | Y |  |
| Lee 2011108 | NHL (indolent) | CR | PFS  OS | Unclear | 1st | Chemo | 15† | 5128† | RCT | 1990-2009 | AD | NR |  | Y |  |
| Mangal 2018114 (NHL) | NHL | ORR\*  CR | PFS | Stage III/IV >75% in most cohorts | Various | Various | 73 | 6071 | RCT + SA | 1996 to 2015 | AD | NR | Y |  |  |
| Shi 2017129 | NHL (indolent; follicular) | CR 30mo  CR 24mo | PFS | Unclear | 1st | Chemo or immuno (induction or maintenance) | 13 | 3837 | RCT | 1990 to 2011 | IPD | NR (CR= disappearance) |  | Y | Y |
| Zhu 2017141 | NHL (indolent; follicular) | CR | PFS | Unclear | NR | Chemo, immune or targeted | 13 | NR | RCT + SA | 1993 to 2013 | AD | NR | Y |  |  |
| Zhu 2017141 | NHL (mantle cell) | CR | PFS | Unclear | NR | Chemo, immune or targeted | NR | NR | RCT + SA | 1993 to 2013 | AD | NR | Y |  |  |
| Colloca & Venturino 201787 | Ovarian | ORR  CR | PFS  OS | Advanced | 1st | Chemo | 29 | NR | RCT | 1990-2016 | AD | WHO (24), RECIST (8) |  | Y |  |
| Rose 2010126 | Ovarian | ORR\* | PFS  OS | Recurrent / platinum-resistant | 2nd | Various | 11 | 407 | SA | 1994 to 2004 | IPD | WHO (10), RECIST (1) | Y |  |  |
| Siddiqui 2017132 | Ovarian | ORR\* | PFS  OS | Advanced, recurrent | 2nd + subsequent | Chemo | 39† | 9223† | RCT | 2000 to 2015 | AD | NR | Y | Y |  |
| Colloca 2016a88 | Pancreatic | ORR  DoR | PFS  OS | Advanced or metastatic | 1st | Gemcitabine + chemo or targeted | 36\* | NR | RCT | 1997-2014 | AD | RECIST |  | Y |  |
| Hamada 201697 | Pancreatic | ORR | OS | Advanced | 1st | Chemo | 47 | 15906† | RCT | 1995-2015 | AD | NR | Y | Y |  |
| Makris 2017113 | Pancreatic (adenocarcinoma) | ORR | OS | Locally advanced, unresectable or metastatic | 1st | Chemo (gemcitabine) | 22\* | 10379\* | RCT | 2000 to 2015 | AD | NR (RR=shrinkage or disappearance) |  | Y |  |
| Colloca 2016c90 | Prostate | ORR | OS | Metastatic (castration-resistant) | 1st + 2nd | Chemo, hormonal + targeted | 17 | NR | RCT | 1995-2014 | AD | NR (CR=disappearance; PR=≥30% reduction) |  | Y |  |
| Abdel-Rahman 201878 | Renal cell | ORR | OS | Advanced | Various | Immune checkpoint inhibitors (PD-(L)1) | 4 | 1093 | RCT + SA | To 2017 | AD | RECIST | Y |  |  |
| Delea 201292 | Renal cell | ORR | OS | Metastatic | NR | Cytokine or targeted | 25\* | 10943† | RCT | 1997-2010 | AD | NR |  | Y |  |
| Petrelli 2013123 | Renal cell | ORR | PFS  OS | Metastatic | 1st | Targeted | 6† | 3188† | RCT | Up to 2011 | AD | NR | Y | Y |  |
| Tanaka 2019134 | Soft tissue sarcoma | ORR | OS | Advanced | 1st | Chemo | 27† | 6156† | RCT | 1974 to 2017 | AD | NR |  | Y |  |
| Zer 2016140 | Soft tissue sarcoma | ORR | OS | Advanced or metastatic | All | Systemic | 52† | 9762† | RCT | 1974 to 2014 | AD | NR |  | Y |  |
| Penel 2014122 | Unknown primary | ORR\* | PFS  OS | Unclear | NR | NR | 38† | NR | SA | 1997 to 2011 | AD | RECIST or WHO | Y |  |  |
| Abdel-Rahman 201878 | Urothelial | ORR | OS | Advanced | Various | Immune checkpoint inhibitors (PD-(L)1) | 9 | 1699 | RCT + SA | To 2017 | AD | RECIST | Y |  |  |
| Agarwal 201479 | Urothelial | ORR | OS | Advanced (operable or metastatic) | 2nd | Chemo or biologic | 10 | 560 | RCT + SA | NR | AD | RECIST | Y |  |  |
| Kaufman 2018107 | Various solid tumours | ORR | OS | Unclear | Various | Immune checkpoint inhibitors +/- chemo | 27† | 10300† | RCT | 2005-2017 | AD | RECIST or mWHO |  | Y |  |
| Mushti 2018117 | Various solid tumours | ORR\* | OS | Unclear | NR | Immune checkpoint inhibitors (PD-(L)1) | 13 | 6722 | RCT | 2014 to 2016 | AD | RECIST |  | Y |  |
| Nie 2019120 | Various solid tumours | ORR\* | OS | Advanced or recurrent | Various | Immune checkpoint inhibitors (PD-(L)1) | 43† | 15088† | RCT + SA | Up to 2018 | AD | RECIST | Y | Y |  |
| Ritchie 2018125 | Various solid tumours | ORR\* | PFS  OS | Advanced | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 20† | 10828† | RCT | 2000 to 2017 | AD | NR | Y | Y |  |
| Roviello 2017127 | Various solid tumours | ORR | PFS  OS | Unclear | Various | Immune checkpoint inhibitors | 17† | 8994† | RCT | Up to 2017 | AD | RECIST or mWHO |  | Y |  |
| Tsujino 2010137 | Various solid tumours | ORR | PFS  OS | Advanced | NR | Targeted | 18 | NR | RCT | Up to 2009 | AD | NR |  | Y | Y |
| Vidaurre 2009138 | Various | ORR\* | PFS  OS | Advanced, locally advanced, unresectable or metastatic | NR | Chemo or targeted | 143† | 6974† | RCT + SA | 2006 to 2008 | AD | NR | Y |  |  |
| Wilkerson+Fojo 2009139 | Various solid tumours | ORR | PFS  OS | Metastatic | NR | NR | 66† | NR | RCT | NR | AD | NR |  | Y |  |
| Note: Of the 63 included studies (64 refs), 8 references78, 106, 108, 125, 127, 137, 138, 141 appear on 2-3 rows as they report on 2-3 different cancer types. \*Calculated from reported data. †Unclear for individual subgroups.  AD, aggregate data; chemo, chemotherapy; CR, complete response; DoR, duration of response; ECOG, Eastern Cooperative Oncology Group; IMWG, International Myeloma Working Group (criteria); IPD, individual patient data; mo, months; mWHO, modified World Health Organisation (criteria); NHL, non-Hodgkin’s lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; PR, partial response; RCT, randomised controlled trials; RECIST, Response Evaluation Criteria In Solid Tumours; SA, single-arm studies; SCLC, small cell lung cancer; STE, surrogate threshold effect; TTP, time to progression; VGPR, very good partial response; WHO, World Health Organisation (criteria). | | | | | | | | | | | | | | | |

Appendix 5: Absolute correlation & regression results (ordered by outcome type then cancer type then author)

| **Ref** | **SO** | **FO** | **Cancer** | **Line Sub-groups** | **Treatment** | **N stds** | **N pts** | **Absolute correlation methods** | **Correlation coefficient (95% CI), p-value** | **Absolute regression Methods** | **Regression R2 (95% CI), p-value** | **Linear regression equation** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ORR vs. PFS (or TTP)** | | | | | | | | | | | | |
| Louvet 2001112 | ORR | PFS | Colorectal | 1st | Various | 29 | 13498 | Spearman (ORR vs. med PFS) | rs=0.66, p<0.0001 | LR (ORR vs. med PFS) |  | PFS = 3.2 + 0.1 \* ORR |
| Ichikawa 2006102 | ORR | TTP | Gastric | 1st | Chemo (any) | 12\* | 2144 | Spearman, wtd O(RR vs. med TTP) | rs=0.49, p<0.0001 | WLR (ORR vs. med TTP) |  | TTP = 1.73 + 0.09 \* ORR |
| Ichikawa 2006102 | ORR | TTP | Gastric | 1st | Chemo (novel) | 8\* | 1077 | Spearman, wtd (ORR vs. med TTP) | rs=0.41, p=0.018 |  |  |  |
| Ichikawa 2006102 | ORR | TTP | Gastric | 1st | Chemo (non-novel) | 7\* | 1067 | Spearman, wtd (ORR vs. med TTP) | rs=0.56, p=0.0053 |  |  |  |
| Ito 2019105 | ORR | PFS | Lung (NSCLC) | Various | Immune checkpoint inhibitors (PD-(L)1) | 6 | 3752† | a) Pearson, wtd b) Spearman, wtd (ORR vs. med PFS) | a) r=0.55, p<0.0001 b) rs=0.33, p<0.0001 | WLR R2 (ORR vs. med PFS) | R2=0.30, p=0.206 |  |
| Ito 2019105 | ORR | PFS | Lung (NSCLC) | - Various - High PD-L1 expression | Immune checkpoint inhibitors (PD-(L)1) | 7 | 1381 | a) Pearson, wtd b) Spearman, wtd (ORR vs. med PFS) | a) r=0.90, p<0.0001 b) rs=0.48, p<0.0001 | WLR R2 (ORR vs. med PFS) | R2=0.81, p=0.006 |  |
| Ritchie 2018125 | ORR | PFS | Lung (NSCLC) | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 8 | NR | Correlation (NR) (ORR vs. 6mo PFS) | r=0.85 (0.63 to 1.06), p=NR |  |  |  |
| Tsujino 2009136 | ORR | PFS | Lung (NSCLC) | NR | Targeted | 18\* | 3790\* |  |  | LR (ORR vs. med PFS) | R2=NR, p=0.001 | Slope 0.072 |
| Vidaurre 2009138 | ORR | PFS | Lung (NSCLC) | NR | Chemo or targeted | 35 | NR |  |  | Regression (NR) (ORR vs. med PFS) | R2=0.75, p<0.0001 |  |
| Nickolich 2014119 | ORR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (ORR vs. med PFS) | r=0.73, p<0.0001 |  |  |  |
| Nickolich 2014119 | ORR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (ORR vs. med PFS) | r=0.02, p=0.978 |  |  |  |
| Nickolich 2014119 | ORR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (ORR vs. med PFS) | r=0.51, p=0.013 |  |  |  |
| Mangal 2018115 (myeloma) | ORR | PFS | Multiple myeloma | 2nd + | Various | 79† | 13322† |  |  | WLR adj R2 (logit ORR vs. log med PFS) | Adj R2=0.50, p=NR |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | Various | Systemic | 22 | 1310 | Pearson (ORR vs. med PFS) | r=0.37 (-0.05 to 0.80), p=0.085 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | - Various - Published 1996-2010 | Systemic | 6\* | NR | Pearson (ORR vs. med PFS) | r= -0.08 (-0.76 to 0.60), p=0.824 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | - Various - Published 2011-2016 | Systemic | 16\* | NR | Pearson (ORR vs. med PFS) | r=0.43 (-0.07 to 0.93), p=0.095 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | Various | Cytotoxic | 9 arms | NR | Pearson (ORR vs. med PFS) | r=0.63 (0.03 to 1.22), p=0.041 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | Various | Non-cytotoxic | 18 arms | NR | Pearson (ORR vs. med PFS) | r=0.18 (-0.27 to 0.62), p=0.432 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | Various | Targeted | 19 arms | NR | Pearson (ORR vs. med PFS) | r=0.42 (-0.06 to 0.90), p=0.086 |  |  |  |
| Imaoka 2019104 | ORR | PFS | Neuroendocrine | Various | Non-targeted | 8 arms | NR | Pearson (ORR vs. med PFS) | r= -0.72 (-1.09 to -0.35), p<0.001 |  |  |  |
| Mangal 2018114 (NHL) | ORR | PFS | NHL | Various | Various | 73 | 6071 |  |  | LR adj R2 (logit ORR vs. log med PFS) | Adj R2=0.70, p=NR | log (med PFS) = 1.97 + 0.414 \* logit (ORR) |
| Rose 2010126 | ORR | PFS | Ovarian | 2nd | Various | 11 | 407 | a) Pearson b) Kendall Tau-b (ORR vs. med PFS) | a) r=0.62, p=0.044  b) r=0.48, p=0.042 |  |  |  |
| Siddiqui 2017132 | ORR | PFS | Ovarian | 2nd + | Chemo | 39† | 9223† | a) Pearson, wtd (ORR vs. med PFS) b) Pearson, unwtd (ORR vs. med PFS) | a) r=0.85, p<0.001 b) 0.76, p<0.001 | WLR R2 (ORR vs. med PFS): a) unadj b) adj | a) R2=0.72, p=NR b) adj R2=0.72, p=NR | med PFS = 2.59 + 0.12 \* ORR |
| Petrelli 2013123 | ORR | PFS | Renal cell | 1st | Targeted | 6† | 3188† | Spearman, wtd (ORR vs. med PFS) | rs=0.96, p<0.0001 |  |  |  |
| Penel 2014122 | ORR | PFS | Unknown primary | NR | NR | 38† | NR | Pearson via WLR (ORR v. med PFS) | r=0.54, p<0.0001 |  |  |  |
| Ritchie 2018125 | ORR | PFS | Various solid tumours | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 20† | 10828† | Correlation (NR) (ORR vs. 6mo PFS) | r=0.37 (0.06 to 0.95), p=NR |  |  |  |
| Vidaurre 2009138 | ORR | PFS | Various | NR | Chemo | 85 | 3982\* |  |  | Regression (NR) (ORR vs. med PFS) | R2=0.53, p<0.0001 |  |
| Vidaurre 2009138 | ORR | PFS | Various | NR | Targeted | 58 | 2992\* |  |  | Regression (NR) (ORR vs. med PFS) | R2=0.61, p<0.0001 |  |
| Vidaurre 2009138 | ORR | PFS | Various | NR | Chemo or targeted | 143† | 6974† |  |  | Regression (NR) (ORR vs. med PFS) | R2=0.56, p<0.0001 |  |
| **ORR vs. OS** | | | | | | | | | | | | |
| Agarwal 201780 | ORR | OS | Acute myeloid leukemia | 1st | Systemic | 20† | NR |  |  | WLR adj R2 (logit ORR vs. log med OS) | Adj R2=0.45, p=NR |  |
| Liu 2016111 | ORR | OS | Breast | 2nd + 3rd | Chemo | 24 | 8617 | Spearman (ORR vs. med OS) | rs=0.54 (0.29 to 0.72), p<0.0001 |  |  |  |
| Liu 2016111 | ORR | OS | Breast | - 2nd + 3rd - Previous anthracycline/taxanes | Chemo | 15\* | NR | Spearman (ORR vs. med OS) | rs=0.62 (0.32 to 0.84), p=NR |  |  |  |
| Liu 2016111 | ORR | OS | Breast | - 2nd + 3rd - Previous trastuzumab/bevacizumab | Chemo | 5\* | NR | Spearman (ORR vs. med OS) | rs=0.78 (0.19 to 1.0), p=NR |  |  |  |
| Liu 2016111 | ORR | OS | Breast | 2nd + 3rd | Chemo (taxanes) | 21\* | NR | Spearman (ORR vs. med OS) | rs=0.49 (-0.19 to 0.92), p=NR |  |  |  |
| Liu 2016111 | ORR | OS | Breast | 2nd + 3rd | Chemo (antimetabolites) | 22\* | NR | Spearman (ORR vs. med OS) | rs=-0.10, p=NR |  |  |  |
| Liu 2016111 | ORR | OS | Breast | - 2nd + 3rd - HER2-pos | Chemo | 5\* | NR | Spearman (ORR vs. med OS) | rs=0.96 (0.80 to 1.00), p=NR |  |  |  |
| Liu 2016111 | ORR | OS | Breast | - 2nd + 3rd - HER2-neg | Chemo | 3\* | NR | Spearman (ORR vs. med OS) | rs=1.00, p=NR |  |  |  |
| Petrelli 2014124 | ORR | OS | Breast | 1st | Targeted + chemo | 20† | 10138† | Spearman, wtd (ORR vs. med OS) | rs=0.61 (0.59 to 0.63), p=NR |  |  |  |
| Giessen 201595 | ORR | OS | Colorectal | 2nd | Chemo | 22 | 10509 | Pearson, wtd (log odds ORR vs. log med OS) | r=0.58 (0.38 to 0.72), p=0.003 |  |  |  |
| Louvet 2001112 | ORR | OS | Colorectal | 1st | Various | 28\* | 13284\* | Spearman (ORR vs. med OS) | rs=0.41, p=0.0009 | LR (ORR vs. med OS) |  | OS = 10.45 + 0.088 \* ORR |
| Tang 2007135 | ORR | OS | Colorectal | 1st | Chemo | 39 | 18668 | Spearman (ORR vs. med OS) | rs=0.59 (0.42 to 0.72), p<0.000001 |  |  |  |
| Ichikawa 2006102 | ORR | OS | Gastric | 1st | Chemo (any) | 25 | 4593 | Spearman, wtd (ORR vs. med OS) | rs=0.45, p<0.0001 | WLR (ORR vs. med OS) |  | OS = 5.89 + 0.08 \* ORR |
| Ichikawa 2006102 | ORR | OS | Gastric | 1st | Chemo (novel) | 11\* | 1170 | Spearman, wtd (ORR vs. med OS) | rs=0.18, p=0.12 |  |  |  |
| Ichikawa 2006102 | ORR | OS | Gastric | 1st | Chemo (non-novel) | 20\* | 3423 | Spearman, wtd (ORR vs. med OS) | rs=0.47, p<0.0001 |  |  |  |
| Shitara 2014130 | ORR | OS | Gastric | 2nd + 3rd | Chemo | 64 | 4286 | Spearman (ORR vs. med OS) | rs=0.38 (0.16 to 0.6), p=NR |  |  |  |
| Pang 2018121 | ORR | OS | Gastroesophageal | 1st + 2nd | Targeted | 18 | 7892 | Correlation (NR) (ORR vs. med OS) | r=0.86, p<0.0001 |  |  |  |
| Han 201498 | ORR | OS | Glioblastoma | Various | Various | 91† | 7125† |  |  | WLR R2 (ORR vs. med OS) | R2=0.22 (0.04 to 0.42), p=NR |  |
| Ito 2019105 | ORR | OS | Lung (NSCLC) | Various | Immune checkpoint inhibitors (PD-(L)1) | 6 | 3752† | a) Pearson, wtd b) Spearman, wtd (ORR vs. med OS) | a) r= -0.02, p=0.4564 b) rs= -0.14, p<0.0001 |  |  |  |
| Ito 2019105 | ORR | OS | Lung (NSCLC) | - Various - High PD-L1 expression | Immune checkpoint inhibitors (PD-(L)1) | 7 | 1381 | a) Pearson, wtd b) Spearman, wtd (ORR vs. med OS) | a) r=0.92, p<0.0001 b) rs=0.77, p<0.0001 | WLR R2 (ORR vs. med OS) | R2=0.84, p=0.004 |  |
| Li 2019109 | ORR | OS | Lung (NSCLC) | 1st + 2nd | Immune checkpoint inhibitors | 5† | 4803† | Pearson (ORR vs. med OS) | r=0.52, p=0.28 | LR (ORR vs. med OS) | R2=0.27, p=NR |  |
| Li 2012110 | ORR | OS | Lung (NSCLC) | 1st + 2nd | Targeted | 60 | 9903 |  |  | WLSR R2 (ORR vs. med OS) | R2=0.83, p<0.000001 |  |
| Ritchie 2018125 | ORR | OS | Lung (NSCLC) | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 8 | NR | Correlation (NR) (ORR vs. 12mo OS) | r=0.66 (0.17 to 1.08), p=NR |  |  |  |
| Sekine 1999128 | ORR | OS | Lung (NSCLC) | Various | Chemo | 42 | 1935 | Pearson (ORR vs. med OS) | r=0.62, p<0.001 |  |  |  |
| Shukuya 2016131 | ORR | OS | Lung (NSCLC) | All | Immune checkpoint inhibitors (PD-(L)1) | 10† | NR | Spearman, wtd (ORR vs. med OS) | rs=0.45, p=0.141 |  |  |  |
| Shukuya 2016131 | ORR | OS | Lung (NSCLC) | All | Chemo (docetaxel) | 22† | NR | Spearman, wtd (ORR vs. med OS) | rs=0.41, p=0.053 |  |  |  |
| Tsujino 2009136 | ORR | OS | Lung (NSCLC) | NR | Targeted | 28 | 6171 |  |  | LR (ORR vs. med OS) | R2=NR, p<0.0001 | Slope 0.258 |
| Vidaurre 2009138 | ORR | OS | Lung (NSCLC) | NR | Chemo or targeted | 35 | NR |  |  | Regression (NR) (ORR vs. med OS) | R2=0.28, p=0.0024 |  |
| Nickolich 2014119 | ORR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (ORR vs. med OS) | r=0.66, p<0.0001 |  |  |  |
| Nickolich 2014119 | ORR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (ORR vs. med OS) | r=0.40, p=0.193 |  |  |  |
| Nickolich 2014119 | ORR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (ORR vs. med OS) | r=0.44, p=0.012 |  |  |  |
| Imaoka 2017103 | ORR | OS | Neuroendocrine | Various | Systemic | 20 | 2530 | Spearman (ORR vs. med OS) | rs= -0.26 (-0.64 to 0.11), p=0.164 |  |  |  |
| Rose 2010126 | ORR | OS | Ovarian | 2nd | Various | 11 | 407 | a) Pearson b) Kendall Tau-b (ORR vs. med OS) | a) r=0.56, p=0.071  b) r=0.40, p=0.086 |  |  |  |
| Siddiqui 2017132 | ORR | OS | Ovarian | 2nd + | Chemo | 31† | 9223† | a) Pearson, wtd (ORR vs. med OS) b) Pearson, unwtd (ORR vs. med OS) | a) r=0.82, p<0.001 b) 0.71, p<0.001 | WLR R2 (ORR vs. med OS): a) unadj b) adj | a) R2=0.67, p=NR b) adj R2=0.66, p=NR | med OS = 9.48 + 0.28 \* ORR |
| Hamada 201697 | ORR | OS | Pancreatic | 1st | Chemo | 47 | 15906† | Spearman (ORR vs. med OS) | rs=0.39 (0.20 to 0.55), p<0.001 |  |  |  |
| Abdel-Rahman 201878 | ORR | OS | Renal cell | Various | Immune checkpoint inhibitors (PD-(L)1) | 4 | 1093 | Pearson (ORR vs. med OS) | r= -0.40, p=0.436 |  |  |  |
| Petrelli 2013123 | ORR | OS | Renal cell | 1st | Targeted | 6† | 3188† | Spearman, wtd (ORR vs. med OS) | rs=0.96, p<0.0001 |  |  |  |
| Penel 2014122 | ORR | OS | Unknown primary | NR | NR | 38† | NR | Pearson via WLR (ORR v. med OS) | r=0.54, p<0.0001 |  |  |  |
| Abdel-Rahman 201878 | ORR | OS | Urothelial | Various | Immune checkpoint inhibitors (PD-(L)1) | 9 | 1699 | Pearson (ORR vs. med OS) | r= -0.12, p=0.758 |  |  |  |
| Agarwal 201479 | ORR | OS | Urothelial | 2nd | Chemo or biologic | 10 | 560 | Pearson (ORR vs. 12mo OS) | r=0.37, p=0.30 | WLR R2 (ORR vs. 12mo OS): a) unadj b) adj (RE) | a) R2=0.26, p=NR b) Adj R2=0.16, p=0.1359 |  |
| Agarwal 201479 | ORR | OS | Urothelial | - 2nd - Operable | Chemo | NR | 214† | Pearson (ORR vs. 12mo OS) | r=0.78, p=NR | WLR adj R2 (ORR vs. 12mo OS) | Adj R2=0.54, p=NR |  |
| Agarwal 201479 | ORR | OS | Urothelial | - 2nd - Metastatic | Chemo | NR | 391† | Pearson (ORR vs. 12mo OS) | r= -0.018, p=NR | WLR adj R2 (ORR vs. 12mo OS) | Adj R2= -0.13, p=NR |  |
| Nie 2019120 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors (PD-(L)1) | 43† | 15088† |  |  | Squared Spearman (ORR vs. med OS) | r2s=0.29, p<0.001 |  |
| Ritchie 2018125 | ORR | OS | Various solid tumours | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 20† | 10828† | Correlation (NR) (ORR vs. 12mo OS) | r=0.08 (-0.17 to 0.70), p=NR |  |  |  |
| Vidaurre 2009138 | ORR | OS | Various | NR | Chemo | 85 | 3982\* |  |  | Regression (NR) (ORR vs. med OS) | R2=0.35, p<0.0001 |  |
| Vidaurre 2009138 | ORR | OS | Various | NR | Targeted | 58 | 2992\* |  |  | Regression (NR) (ORR vs. med OS) | R2=0.45, p<0.0001 |  |
| Vidaurre 2009138 | ORR | OS | Various | NR | Chemo or targeted | 143† | 6794† |  |  | Regression (NR) (ORR vs. med OS) | R2=0.33, p<0.0001 |  |
| **CR vs. PFS** | | | | | | | | | | | | |
| Nickolich 2014119 | CR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (CR vs. med PFS) | r=0.71, p<0.0001 |  |  |  |
| Nickolich 2014119 | CR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (CR vs. med PFS) | r=0.22, p=0.491 |  |  |  |
| Nickolich 2014119 | CR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (CR vs. med PFS) | r=0.35, p=0.116 |  |  |  |
| Mangal 2018115 (myeloma) | CR | PFS | Multiple myeloma | 2nd + | Various | 79† | 13322† |  |  | WLR adj R2 (logit CR vs. log med PFS) | Adj R2=0.47, p=NR |  |
| Mangal 2018114 (NHL) | CR | PFS | NHL | Various | Various | 73 | 6071 |  |  | LR adj R2 (logit CR vs. log med PFS) | Adj R2=0.57, p=NR | log (med PFS) = 2.38 + 0.340 \* logit (CR) |
| Zhu 2017141 | CR | PFS | NHL (indolent; follicular) | NR | Chemo, immune or targeted | 13 | NR |  |  | WLR R2: a) CR vs. med PFS b) CR vs. 3-year PFS | a) R2=0.69 (0.22 to 0.89), p=NR  b) R2=0.44, p=NR | med PFS = 0.83 + 0.46 \* CR |
| Zhu 2017141 | CR | PFS | NHL (mantle cell) | NR | Chemo, immune or targeted | NR | NR |  |  | WLR R2 (CR vs. med PFS) | R2=0.39, p=NR |  |
| **CR vs. OS** | | | | | | | | | | | | |
| Agarwal 201780 | CR | OS | Acute myeloid leukemia | 1st | Systemic | 20† | NR |  |  | WLR adj R2 (logit CR vs. log med OS) | Adj R2=0.48, p=NR |  |
| Pang 2018121 | CR | OS | Gastroesophageal | 1st + 2nd | Targeted | 18 | 7892 | Correlation (NR) (CR vs. med OS) | r=0.43, p=0.18 |  |  |  |
| Li 2019109 | CR | OS | Lung (NSCLC) | 1st + 2nd | Immune checkpoint inhibitors | 5\* | 4103\* | Pearson (CR vs. med OS) | r=0.19, p=0.75 | LR (CR vs. med OS) | R2=0.04, p=NR |  |
| Nickolich 2014119 | CR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (CR vs. med OS) | r=0.62, p<0.0001 |  |  |  |
| Nickolich 2014119 | CR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (CR vs. med OS) | r=-0.04, p=0.863 |  |  |  |
| Nickolich 2014119 | CR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (CR vs. med OS) | r=0.19, p=0.295 |  |  |  |
| **PR (or VGPR or CR) vs. PFS** | | | | | | | | | | | | |
| Nickolich 2014119 | PR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (PR vs. med PFS) | r=0.35, p=0.019 |  |  |  |
| Nickolich 2014119 | PR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (PR vs. med PFS) | r=0.70, p=0.011 |  |  |  |
| Nickolich 2014119 | PR | PFS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (PR vs. med PFS) | r=0.49, p=0.035 |  |  |  |
| Mangal 2018115 (myeloma) | VGPR or CR | PFS | Multiple myeloma | 2nd + | Various | 79† | 13322† |  |  | WLR adj R2 (VGPR or CR vs. med PFS) | Adj R2=0.64, p=NR |  |
| **PR vs. OS** | | | | | | | | | | | | |
| Nickolich 2014119 | PR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited or extensive | Various | 66† | 8471† | Pearson (PR vs. med OS) | r=0.29, p=0.018 |  |  |  |
| Nickolich 2014119 | PR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Limited disease | Various | 66† | 8471† | Pearson (PR vs. med OS) | r=0.60, p=0.009 |  |  |  |
| Nickolich 2014119 | PR | OS | Lung (SCLC) | - 1st + 2nd + maintenance - Extensive disease | Various | 66† | 8471† | Pearson (PR vs. med OS) | r=0.66, p=0.0002 |  |  |  |
| \*Calculated from reported data. †Unclear for individual subgroups.  adj, adjusted; chemo, chemotherapy; CI, confidence interval; CR, complete response; FO, final outcome; HER2, human epidermal growth factor receptor 2; log, logarithm; LR, linear regression; med, median; mo, months; NHL, non-Hodgkin’s lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; PR, partial response; r, Pearson correlation; R2, regression coefficient of determination; r2s, squared Spearman rank correlation; rs, Spearman rank correlation; SCLC, small cell lung cancer; SO, surrogate outcome; TTP, time to progression; unwtd, unweighted; VGPR, very good partial response; wtd, weighted; WLR, weighted linear regression; WLSR, weighted least squares regression. | | | | | | | | | | | | |

Appendix 6: Treatment effect correlation & regression results (ordered by outcome type then cancer type then author)

| **Ref** | **SO** | **FO** | **Cancer** | **Line Sub-groups** | **Treatment** | **N stds** | **N pts** | **Treatment effect correlation methods** | **Correlation coefficient (95% CI), p-value** | **Treatment effect regression methods** | **Regression R2 (95% CI), p-value** | **Linear regression equation** | **STE** | **IQWiG** | **BSES2** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ORR vs. PFS** | | | | | | | | | | | | | | | |
| Burzykowski 200884 | ORR | PFS | Breast | 1st | Chemo | 11 | 3953 | Spearman via LR with Plackett copula (logOR ORR vs. logHR PFS) | rs=0.96 (0.73 to 1.19), p=NR | LR |  | logHR PFS = 0.10 + 0.50 \* logOR ORR | NR | Medium+ | NE |
| Ciani 201586 Elia 201893 | ORR | PFS | Colorectal | All | Systemic | 33 | NR |  |  | LR: Adj R2 (logOR ORR vs. logHR PFS) | Adj R2=0.61 (0.27 to 0.87), p=NR | logHR PFS = -0.05 - 0.32 \* logOR ORR | NR | Medium | NE |
| Ciani 201586 Elia 201893 | ORR | PFS | Colorectal | - All - No crossover | Systemic | 7 | NR |  |  | LR: Adj R2 (logOR ORR vs. logHR PFS) | Adj R2=0.63 (0.03 to 0.99), p=NR | logHR PFS = -0.05 - 0.31 \* logOR ORR | NR | Medium | NE |
| Tsujino 2010137 | ORR | PFS | Colorectal | NR | Targeted | 7 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR PFS) | R2=0.65, p=0.029 | Slope -0.037 | NR | Medium | NE |
| Blumenthal 201782 | ORR | PFS | Lung (NSCLC) | Various | Chemo, immune or targeted | 25 | 20013† |  |  | WLR R2: a) OR ORR vs. HR PFS b) 6mo ratio ORR vs. HR PFS | a) R2=0.74 (0.55 to 0.88), p=NR b) R2=0.70 (0.50 to 0.84), p=NR |  | NR | Medium+ | NE |
| Blumenthal 201581 | ORR | PFS | Lung (NSCLC) | Various | Chemo or targeted | 14 | 12567† |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.89 (0.80 to 0.98), p=NR |  | NR | Medium+ | NE |
| Blumenthal 201581 | ORR | PFS | Lung (NSCLC) | Various | Chemo | 11 | 11701† |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.77 (0.58 to 0.96), p=NR |  | NR | Medium+ | NE |
| Ito 2019105 | ORR | PFS | Lung (NSCLC) | Various | Immune (PD-(L)1) | 6 | 3752† | a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR PFS) | a) r= -0.87, p<0.0001 b) rs= -0.97, p<0.0001 | WLR R2 (OR ORR vs. HR PFS) | R2=0.76, p=0.011 |  | NR | Medium+ | Fair |
| Ito 2019105 | ORR | PFS | Lung (NSCLC) | - Various - High PD-L1 expression | Immune checkpoint inhibitors (PD-(L)1) | 7 | 1381 | a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR PFS) | a) r=0.67, p<0.0001 b) rs=0.56, p<0.0001 | WLR R2 (OR ORR vs. HR PFS) | R2=0.45, p=0.101 |  | NR | Low | Good |
| Ritchie 2018125 | ORR | PFS | Lung (NSCLC) | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 8 | NR | Correlation (NR), wtd (OR ORR vs. HR PFS) | r=0.74 (0.38 to 1.08), p=NR |  |  |  | NR | Medium | Good |
| Roviello 2017127 | ORR | PFS | Lung (NSCLC) | Various | Immune checkpoint inhibitors | 7\* | 3369\* |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.42 (0.003 to 0.85), p=0.06 |  | NR | Low | NE |
| Tsujino 2010137 | ORR | PFS | Lung (NSCLC) | NR | Targeted | 6 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR PFS) | R2=0.94, p=0.002 | Slope -0.015 | NR | Medium+ | NE |
| Colloca & Venturino 201787 | ORR | PFS | Ovarian | 1st | Chemo | 29 | NR | Spearman (diff ORR vs. diff med PFS) | rs=0.64, p<0.001 | LR R2 (log RR ORR vs. log HR PFS) | R2=0.28, p=0.005 |  | NR | Low | NE |
| Colloca & Venturino 201787 | ORR | PFS | Ovarian | - 1st - Published 1990-2002 | Chemo | 15 | NR | Spearman (diff ORR vs. diff med PFS) | rs=0.64, p=0.018 | LR R2 (log RR ORR vs. log HR PFS) | R2=0.32, p=0.046 |  | NR | Low | NE |
| Colloca & Venturino 201787 | ORR | PFS | Ovarian | - 1st - Published 2003-2016 | Chemo | 16 | NR | Spearman (diff ORR vs. diff med PFS) | rs=0.58, p=0.019 | LR R2 (log RR ORR vs. log HR PFS) | R2=0.53, p=0.003 |  | NR | Medium | NE |
| Siddiqui 2017132 | ORR | PFS | Ovarian | 2nd + | Chemo | 39† | 9223† | Pearson, wtd (OR ORR vs. HR PFS) | r=0.42, p=NR |  |  |  | NR | Low | Poor |
| Colloca 2016a88 | ORR | PFS | Pancreatic | 1st | Gemcitabine + chemo or targeted | 33\* | NR | Spearman (diff ORR vs. diff med PFS) | rs=0.34, p=NR |  |  |  | NR | Low | NE |
| Colloca 2016a88 | ORR | PFS | Pancreatic | 1st | Gemcitabine + targeted | 14\* | NR | Spearman (diff ORR vs. diff med PFS) | rs=0.25, p=NR |  |  |  | NR | Low | NE |
| Ritchie 2018125 | ORR | PFS | Various solid tumours | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 20† | 10828† | Correlation (NR), wtd (OR ORR vs. HR PFS) | r=0.63 (0.35 to 0.89), p=NR |  |  |  | NR | Medium | Poor |
| Roviello 2017127 | ORR | PFS | Various solid tumours | Various | Immune checkpoint inhibitors | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.32 (0.02 to 0.76), p=0.01 | logHR PFS = -0.1281 - 0.2384 \* logOR ORR | NR | Low | NE |
| Roviello 2017127 | ORR | PFS | Various solid tumours | Various | Immune checkpoint inhibitors (CTLA-4) | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.67 (0.02 to 1.00), p=0.05 |  | NR | Medium | NE |
| Roviello 2017127 | ORR | PFS | Various solid tumours | Various | Immune checkpoint inhibitors (PD-(L)1) | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR PFS) | R2=0.25 (0.02 to 1.00), p=0.08 |  | NR | Low | NE |
| Tsujino 2010137 | ORR | PFS | Various solid tumours | NR | Targeted | 17 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR PFS) | R2=0.50, p=0.001 | Slope -0.022 | 15% | Medium | NE |
| Wilkerson+Fojo 2009139 | ORR | PFS | Various solid tumours | NR | NR | 66† | NR |  |  | LR (unwtd R2): a) diff ORR vs. HR PFS b) diff ORR vs. diff med PFS | a) R2=0.45, p<0.0001 b) R2=0.62, p<0.0001 |  | NR | Medium | NE |
| **ORR vs. OS** | | | | | | | | | | | | | | | |
| Moriwaki 2016116 | ORR | OS | Biliary tract | 1st | Chemo | 17† | 2040 |  |  | WLR R2 (ratio ORR vs. log ratio med OS) | R2=0.29 (0.01 to 0.65), p=0.021 | log ratio med OS = 0.013 + 0.282 \* ratio ORR | NR | Low | NE |
| Moriwaki 2016116 | ORR | OS | Biliary tract | 1st | Chemo (gemcitabine) | 14† | 1880 |  |  | WLR R2 (ratio ORR vs. log ratio med OS) | R2=0.39 (0.02 to 0.75), p=0.013 | log ratio med OS = 0.020 + 0.268 \* ratio ORR | NR | Low | NE |
| Moriwaki 2016116 | ORR | OS | Biliary tract | 1st | Targeted | 6† | 953 |  |  | WLR R2 (ratio ORR vs. log ratio med OS) | R2=0.43 (0.03 to 0.89), p=0.090 | log ratio med OS = 0.119 + 0.155 \* ratio ORR | NR | Low | NE |
| Bruzzi 200583 | ORR | OS | Breast | All | Chemo | 10 | 2126 |  |  | WLR R2: a) logOR ORR vs. logHR OS b) diff ORR vs. diff med OS | a) R2=0.10 (0.00 to 0.43), p=NR b) R2=0.20 (0 to 0.65), p=NR |  | NR | Low | NE |
| Burzykowski 200884 | ORR | OS | Breast | 1st | Chemo | 11 | 3953 | Spearman via LR with Plackett copula (logOR ORR vs. logHR OS) | rs=0.57 (-0.31 to 1.44), p=NR |  |  |  | NR | Medium | NE |
| Hackshaw 200596 | ORR | OS | Breast | 1st | Chemo | 42\* | 9163 |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.34, p<0.0001 | logHR OS = -0.0081 + 0.28 \* logOR ORR  Slope 0.28 | NR | Low | NE |
| Hackshaw 200596 | ORR | OS | Breast | - 1st - Recruited pre-1990 | Chemo | 26\* | 5244\* |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.26, p=0.004 | Slope 0.28 | NR | Low | NE |
| Hackshaw 200596 | ORR | OS | Breast | - 1st - Recruited 1990 or after | Chemo | 16\* | 3919\* |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.41, p=0.005 | Slope 0.24 | NR | Low | NE |
| Buyse 200085 | ORR | OS | Colorectal | 1st | Chemo | 25 | 3791 |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.38 (0.09 to 0.68), p=NR |  | NR | Low | NE |
| Ciani 201586 Elia 201893 | ORR | OS | Colorectal | All | Systemic | 32 | NR | Spearman (logOR ORR vs. logOR OS) | rs=0.53, p<0.01 | a) WLSR R2 (logOR ORR vs. logOR OS) (timepoint NR) b) Adj R2 (logOR ORR vs. logHR OS) | a) R2=0.06 (0.01 to 0.29), p=NR b) Adj R2=0.33 (0.00 to 0.91), p=NR | logHR OS = -0.03 - 0.05 \* logOR ORR | 0.28 | Low | NE |
| Ciani 201586 Elia 201893 | ORR | OS | Colorectal | - All - No crossover | Systemic | 7 | NR |  |  | LR: Adj R2 (logOR ORR vs. logHR OS) | Adj R2=0.40 (0.00 to 0.96), p=NR | logHR OS = -0.04 - 0.10 \* logOR ORR | NR | Low | NE |
| Colloca 2016b89 | ORR | OS | Colorectal | 1st | Bevacizumab + chemo | 11 | NR | Spearman (diff ORR vs. diff med OS) | rs=0.82, p<0.001 | LR R2 (diff ORR vs. diff med OS) | R2=0.58, p=0.002 |  | NR | Medium | NE |
| Cremolini 201791 | ORR | OS | Colorectal | 2nd | Targeted | 20\* | 7571 | Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | a) r=0.17, p=0.476 b) r=0.35, p=0.092 | WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | b) R2=0.03, p=0.476 b) R2=0.12, p=0.092 | a) Slope -0.029 b) Slope 0.071 | NR | Low | NE |
| Cremolini 201791 | ORR | OS | Colorectal | 2nd | Targeted, anti-angiogenic | 13\* | NR | Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | a) r=0.36, p=0.249 b) r=0.52, p=0.038 | WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | b) R2=0.13, p=0.249 b) R2=0.27, p=0.038 | a) Slope -0.113 b) Slope 0.133 | NR | Low | NE |
| Cremolini 201791 | ORR | OS | Colorectal | 2nd | Targeted, not anti-angiogenic | 7\* | NR | Pearson (via WLR): a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | a) r=0.44, p=0.274 b) r=0.63, p=0.068 | WLR R2: a) rr ORR vs. HR OS b) diff ORR vs. diff med OS | b) R2=0.20, p=0.274 b) R2=0.40, p=0.068 | a) Slope -0.064 b) Slope 0.143 | NR | Low | NE |
| Johnson 2006106 | ORR | OS | Colorectal | 1st | Chemo | 146† | 35337† |  |  | WLSR R2 (diff ORR vs. diff med OS) | R2=0.10, p<0.0001 | Diff med OS = 0.340 + 0.096 \* diff ORR | NR | Low | NE |
| Sidhu 2013133 | ORR | OS | Colorectal | 1st (most) | Chemo +/- targeted | 24† | 20438† | Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) r=0.62 (0.37 to 0.79), p=NR b) r=0.64 (0.39 to 0.79), p=NR c) r=0.52 (0.23 to 0.72), p=NR | LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) R2=0.39 (0.13 to 0.62), p=NR b) R2=0.41 (0.15 to 0.63), p=NR c) R2=0.27 (0.05 to 0.52), p=NR |  | NR | Medium | NE |
| Sidhu 2013133 | ORR | OS | Colorectal | 1st (most) | Targeted + chemo | 13 | 12060\* | Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) r=0.50 (0.05 to 0.75), p=NR b) r=0.58 (0.19 to 0.80), p=NR c) r=0.42 (0.00 to 0.71), p=NR | LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) R2=0.25 (0.00 to 0.57), p=NR b) R2=0.33 (0.04 to 0.64), p=NR c) R2=0.18 (0.00 to 0.51), p=NR |  | NR | Medium | NE |
| Sidhu 2013133 | ORR | OS | Colorectal | 1st (most) | Targeted (anti-EGFR) | 9 | 7792\* | Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) r=0.67 (0.27 to 0.86), p=NR b) r=0.72 (0.35 to 0.88), p=NR c) r=0.52 (0.00 to 0.79), p=NR | LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) R2=0.45 (0.07 to 0.74), p=NR b) R2=0.52 (0.12 to 0.78), p=NR c) R2=0.27 (0.00 to 0.62), p=NR |  | NR | Medium | NE |
| Sidhu 2013133 | ORR | OS | Colorectal | 1st (most) | Targeted (anti-EGFR), KRAS non-mutant | 6\* | 4916\* | Correlation (NR): a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) r=0.68 (0.07 to 0.89), p=NR b) r=0.81 (0.38 to 0.94), p=NR c) r=0.48 (0.00 to 0.82), p=NR | LR (unwtd) R2: a) OR ORR vs. HR OS b) Diff ORR vs. HR OS c) Ratio ORR vs. HR OS | a) R2=0.46 (0.01 to 0.80), p=NR b) R2=0.65 (0.15 to 0.88), p=NR c) R2=0.23 (0.00 to 0.67), p=NR |  | NR | Medium | NE |
| Tang 2007135 | ORR | OS | Colorectal | 1st | Chemo | 39 | 18668 | Spearman (diff ORR vs. diff med OS) | rs=0.39 (0.08 to 0.63), p=0.015 |  |  |  | NR | Low | Poor |
| Tsujino 2010137 | ORR | OS | Colorectal | NR | Targeted | 7 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR OS) | R2=0.51, p=0.072 | Slope -0.029 | NR | Medium | NE |
| Blumenthal 201782 | ORR | OS | Lung (NSCLC) | Various | Chemo, immune or targeted | 25 | 20013† |  |  | WLR R2: a) OR ORR vs. HR OS b) 6mo ratio ORR vs. HR OS | a) R2=0.04 (0.0002 to 0.28), p=NR b) R2=0.05 (0.0001 to 0.31), p=NR |  | NR | Low | NE |
| Blumenthal 201581 | ORR | OS | Lung (NSCLC) | Various | Chemo or targeted | 14 | 12567† |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.09 (0 to 0.33), p=NR |  | NR | Low | NE |
| Blumenthal 201581 | ORR | OS | Lung (NSCLC) | Various | Chemo | 11 | 11701† |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.44 (0.08 to 0.80), p=NR |  | NR | Low | NE |
| Hashim 201899 | ORR | OS | Lung (NSCLC) | 2nd + | Various | 140 | 41725 | Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS | a) r=0.17 (0.00 to 0.38), p=NR b) r=0.18 (0.02 to 0.34), p=0.032 |  |  |  | NA | Low | NE |
| Hashim 201899 | ORR | OS | Lung (NSCLC) | - 2nd + - Phase III | Various | 59 | 32348 | Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS | a) r=0.37 (0.09 to 0.60), p=NR b) r=0.13 (0.00 to 0.38), p=0.32 |  |  |  | NA | Low | NE |
| Hashim 201899 | ORR | OS | Lung (NSCLC) | - 2nd + - Phase III excl per-protocol crossover | Various | 54 | 30654 | Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS | a) r=0.40 (0.10 to 0.63), p=NR b) r=0.36 (0.10 to 0.57), p=0.0074 |  |  |  | NA | Low | NE |
| Hashim 201899 | ORR | OS | Lung (NSCLC) | - 2nd + - Phase III excl any crossover | Various | 38 | 22574 | Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS | a) r=0.52 (0.18 to 0.75), p=NR b) r=0.45 (0.15 to 0.67), p=0.0051 |  |  |  | a) 55% b) NA | Medium | NE |
| Hashim 201899 | ORR | OS | Lung (NSCLC) | - 2nd + - Phase III excl crossover or unbalanced post-progression treatments | Various | 18 | 13349 | Correlation (NR) via WLR: a) diff ORR vs. logHR OS b) diff ORR vs. diff med OS | a) r=0.16 (0.00 to 0.60), p=NR b) r=0.53 (0.08 to 0.80), p=0.024 |  |  |  | a) NA b) 41% | Low | NE |
| Hotta 2015100 | ORR | OS | Lung (NSCLC) | Various | Targeted | 18 | 7633† |  |  | WLR R2 (OR ORR vs. HR OS) | R2=0.10, p=NR |  | NR | Low | NE |
| Hotta 2015100 | ORR | OS | Lung (NSCLC) | - Various - Molecularly selected | Targeted | 8 | NR |  |  | WLR R2 (OR ORR vs. HR OS) | R2=0.04, p=NR |  | NR | Low | NE |
| Hotta 2015100 | ORR | OS | Lung (NSCLC) | - Various - Non-molecularly selected | Targeted | 10 | NR |  |  | WLR R2 (OR ORR vs. HR OS) | R2=0.43, p=NR |  | NR | Low | NE |
| Ito 2019105 | ORR | OS | Lung (NSCLC) | Various | Immune checkpoint inhibitors (PD-(L)1) | 6 | 3752† | a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR OS) | a) r= -0.75, p<0.0001 b) rs= -0.96, p<0.0001 | WLR R2 (OR ORR vs. HR OS) | R2=0.57, p=0.051 |  | NR | Medium | Poor |
| Ito 2019105 | ORR | OS | Lung (NSCLC) | - Various - High PD-L1 expression | Immune checkpoint inhibitors (PD-(L)1) | 7 | 1381 | a) Pearson, wtd b) Spearman, wtd (OR ORR vs. HR OS) | a) r= -0.50, p<0.0001 b) rs= -0.21, p<0.0001 | WLR R2 (OR ORR vs. HR OS) | R2=0.25, p=0.253 |  | NR | Low | Fair |
| Johnson 2006106 | ORR | OS | Lung (NSCLC) | 1st | Chemo | 191† | 44125† |  |  | WLSR R2 (diff ORR vs. diff med OS) | R2=0.16, p<0.0001 | Diff med OS = -0.048 + 0.090 \* diff ORR | NR | Low | NE |
| Nakashima 2016118 | ORR | OS | Lung (NSCLC) | 1st | Chemo | 44 | 22709 | Spearman, wtd (lnOR ORR vs. HR OS) | rs=0.57, p=NR | WLSR adj R2 (lnOR ORR vs. lnHR OS) | Adj R2=0.35, p=NR | lnHR OS = -0.023 -0.133 x lnOR ORR | NR | Low | NE |
| Ritchie 2018125 | ORR | OS | Lung (NSCLC) | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 8 | NR | Correlation (NR), wtd (OR ORR vs. HR OS) | r=0.68 (0.08 to 1.10), p=NR |  |  |  | NR | Low | Good |
| Roviello 2017127 | ORR | OS | Lung (NSCLC) | Various | Immune checkpoint inhibitors | 7\* | 3369\* |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.0007 (0.09 to 0.91), p=0.94 |  | NR | Low | NE |
| Tsujino 2010137 | ORR | OS | Lung (NSCLC) | NR | Targeted | 5 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR OS) | R2=0.84, p=0.030 | Slope -0.011 | NR | Medium+ | NE |
| Foster 201194 | ORR | OS | Lung (SCLC) | 1st | Chemo | 3 (32 centres) | 596† | Spearman (logOR ORR vs. logHR OS) | rs=0.52, p=NR | WLSR R2 (logOR ORR vs. logHR OS) | R2=0.21, p=NR |  | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | 1st | Chemo | 48 | 8779 |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.33, p=NR | Diff med OS = 0.00 + 0.06 \* rr ORR | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | - 1st - Clear criteria | Chemo | 43 comp | NR |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.19, p=NR |  | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | - 1st - WHO criteria | Chemo | 23 comp | NR |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.13, p=NR |  | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | - 1st - Non-WHO criteria | Chemo | 20 comp | NR |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.28, p=NR |  | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | - 1st - Published 1990-1996 | Chemo | 26 comp | NR |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.23, p=NR | Diff med OS = 0.00 + 0.04 \* rr ORR | NR | Low | NE |
| Hotta 2009101 | ORR | OS | Lung (SCLC) | - 1st - Published 1997-2008 | Chemo | 26 comp | NR |  |  | WLR R2 (rr ORR vs. diff med OS) | R2=0.47, p=NR | Diff med OS = 0.00 + 0.09 \* rr ORR | NR | Low | NE |
| Colloca & Venturino 201787 | ORR | OS | Ovarian | 1st | Chemo | 27 | NR | Spearman (diff ORR vs. diff med OS) | rs=0.41, p=0.035 | LR R2 (log RR ORR vs. log HR OS) | R2=0.12, p=0.073 |  | NR | Low | NE |
| Colloca & Venturino 201787 | ORR | OS | Ovarian | - 1st - Published 1990-2002 | Chemo | 13 | NR | Spearman (diff ORR vs. diff med OS) | rs=0.65, p=0.016 | LR R2 (log RR ORR vs. log HR OS) | R2=0.15, p=0.199 |  | NR | Low | NE |
| Colloca & Venturino 201787 | ORR | OS | Ovarian | - 1st - Published 2003-2016 | Chemo | 14 | NR | Spearman (diff ORR vs. diff med OS) | rs= -0.02, p=0.940 | LR R2 (log RR ORR vs. log HR OS) | R2=0.34, p=0.027 |  | NR | Low | NE |
| Siddiqui 2017132 | ORR | OS | Ovarian | 2nd + | Chemo | 31† | 9223† |  |  |  |  |  | NR | NE | NE |
| Colloca 2016a88 | ORR | OS | Pancreatic | 1st | Gemcitabine + chemo or targeted | 36\* | NR | Spearman (diff ORR vs. diff med OS) | rs=0.29, p=0.067 |  |  |  | NR | Low | NE |
| Colloca 2016a88 | ORR | OS | Pancreatic | 1st | Gemcitabine + chemo | 22\* | NR | Spearman (diff ORR vs. diff med OS) | rs=0.23, p=0.250 | LR R2 (logRR ORR vs. logHR OS) | R2=0.15, p=NR |  | NR | Low | NE |
| Colloca 2016a88 | ORR | OS | Pancreatic | 1st | Gemcitabine + targeted | 14\* | NR | Spearman (diff ORR vs. diff med OS) | rs=0.55, p=0.035 | LR R2 (logRR ORR vs. logHR OS) | R2=0.28, p=NR |  | NR | Low | NE |
| Hamada 201697 | ORR | OS | Pancreatic | 1st | Chemo | 36 | 15906† | Spearman via WLSR (logOR ORR vs. logHR OS) | rs= -0.16 (-0.27 to -0.05), p=0.007 | WLSR adj R2 (logOR ORR vs. logHR OS) | Adj R2=0.30, p=0.007 |  | NR | Low | Poor |
| Makris 2017113 | ORR | OS | Pancreatic (adenocarcinoma) | 1st | Chemo (gemcitabine) | 22\* | 10379\* | Pearson (log HR OS vs. log OR ORR): a) wtd by sample size b) fixed effect c) random effects | a) r=0.27 (-0.14 to 0.60), p=0.20 b) r=0.52 (0.16 to 0.76), p=0.007 c) r=0.45 (0.07 to 0.72), p=0.02 |  |  |  | NR | Low | NE |
| Colloca 2016c90 | ORR | OS | Prostate | 1st + 2nd | Chemo, hormonal + targeted | 17 | NR | Pearson (diff ORR vs. diff med OS) | r=0.38, p=0.132 | LR R2 (log RR ORR vs. log HR OS) | R2=0.007, p=0.789 |  | NR | Low | NE |
| Colloca 2016c90 | ORR | OS | Prostate | - 1st + 2nd - Published 1995-2004 | Chemo, hormonal + targeted | 5 | NR | Pearson (diff ORR vs. diff med OS) | r=0.35, p=0.560 | LR R2 (log RR ORR vs. log HR OS) | R2=0.53, p=0.275 |  | NR | Medium | NE |
| Colloca 2016c90 | ORR | OS | Prostate | - 1st + 2nd - Published 2005-2014 | Chemo, hormonal + targeted | 12 | NR | Pearson (diff ORR vs. diff med OS) | r=0.41, p=0.185 | LR R2 (log RR ORR vs. log HR OS) | R2=0.02, p=0.690 |  | NR | Low | NE |
| Delea 201292 | ORR | OS | Renal cell | NR | Cytokine or targeted | 25\* | 10943† | Pearson, wtd (ln(rr) ORR vs. -lnHR OS) | r=0.78, p<0.0001 | WLSR adj R2 (ln rr ORR vs. -lnHR OS) | Adj R2=0.59, p<0.0001 | -lnHR OS = -0.11 + 0.30 \* lnrr ORR | NR | Medium | NE |
| Petrelli 2013123 | ORR | OS | Renal cell | 1st | Targeted | 6† | 3188† | a) Pearson, wtd b) Spearman, wtd (diff med OS vs. diff ORR) | a) r =0.52, p<0.0001 b) rs = 0.49, p<0.0001 | LR | R2=0.27, p=NR |  | NR | Low | Fair |
| Tanaka 2019134 | ORR | OS | Soft tissue sarcoma | 1st | Chemo | 27† | 6156† | Kendall's Tau (logOR ORR vs. logHR OS) | τ=0.41, p=NR | Regression (NR) R2 (logOR ORR vs. logHR OS) | R2=0.28 (0.02 to 0.54), p=NR |  | NR | Low | NE |
| Zer 2016140 | ORR | OS | Soft tissue sarcoma | All | Systemic | 52† | 9762† | Correlation (NR) via WLR (OR ORR vs. HR OS) | r=0.51, p=NR |  |  |  | NR | Low | NE |
| Kaufman 2018107 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors + chemo | 27† | 10300† |  |  | WLR adj R2 (OR ORR vs. HR OS) | Adj R2= -0.07, p=0.866 |  | NR | NE | NE |
| Kaufman 2018107 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors alone | NR | NR |  |  | WLR adj R2 (OR ORR vs. HR OS) | Adj R2= -0.08, p=0.799 |  | NR | NE | NE |
| Mushti 2018117 | ORR | OS | Various solid tumours | NR | Immune checkpoint inhibitors (PD-(L)1) | 13 | 6722 |  |  | WLR R2 (OR ORR vs. HR OS) | R2=0.13, p=NR |  | NR | Low | NE |
| Nie 2019120 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors (PD-(L)1) | 43† | 15088† |  |  | WLR R2 (lnOR ORR vs. lnHR OS) | R2=0.10, p=0.053 |  | NR | Low | Poor |
| Ritchie 2018125 | ORR | OS | Various solid tumours | All | Immune checkpoint inhibitors (PD-(L)1 or CTLA4) | 20† | 10828† | Correlation (NR), wtd (OR ORR vs. HR OS) | r=0.57 (0.23 to 0.89), p=NR |  |  |  | NR | Low | Poor |
| Roviello 2017127 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.47 (0.03 to 0.77), p=0.001 | logHR OS = -0.1329 -0.2575 \* logOR ORR | NR | Low | NE |
| Roviello 2017127 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors (CTLA-4) | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.00 (0.00 to 0.97), p=0.96 |  | NR | Low | NE |
| Roviello 2017127 | ORR | OS | Various solid tumours | Various | Immune checkpoint inhibitors (PD-(L)1) | 17† | 8994† |  |  | WLR R2 (logOR ORR vs. logHR OS) | R2=0.18 (0.00 to 0.97), p=0.17 |  | NR | Low | NE |
| Tsujino 2010137 | ORR | OS | Various solid tumours | NR | Targeted | 18 | NR |  |  | LR (unwtd) R2 (diff ORR vs. HR OS) | R2=0.47, p=0.002 | Slope -0.016 | 21% | Low | NE |
| Wilkerson+Fojo 2009139 | ORR | OS | Various solid tumours | NR | NR | 66† | NR |  |  | LR (unwtd R2): a) diff ORR vs. HR OS b) diff ORR vs. diff med OS | a) R2=0.37, p<0.0001 b) R2=0.34, p<0.0001 |  | NR | Low | NE |
| **CR vs. PFS** | | | | | | | | | | | | | | | |
| Lee 2011108 | CR | PFS | NHL (aggressive) | 1st | Chemo | 12† | NR | Spearman (diff CR vs. diff 3yr PFS) | rs=0.63 (0.21 to 0.84), p=0.005 |  |  |  | NR | Medium | NE |
| Lee 2011108 | CR | PFS | NHL (indolent) | 1st | Chemo | 6† | NR | Spearman (diff CR vs. diff 3yr PFS) | rs=0.41 (-0.52 to 0.88), p=0.35 |  |  |  | NR | Medium | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Chemo or immuno (induction or maintenance) | 13 | 3837 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.88 (0.77 to 0.96), p=NR b) R2Copula=0.86 (0.72 to 1.00), p=NR | logHR PFS = -0.093 - 0.636 \* logOR CR 30mo | 1.56 | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Rituximab-based (induction or maintenance) | 9 | 2851 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.85 (0.62 to 0.97), p=NR b) R2Copula=0.80 (0.56 to 1.00), p=NR |  | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Non-rituximab-based (induction or maintenance) | 4 | 986 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.91 (0.05 to 1.00), p=NR b) R2Copula=0.96 (0.90 to 1.00), p=NR |  | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Induction | 8 | 2207 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.89 (0.75 to 0.98), p=NR b) R2Copula=0.89 (0.74 to 1.00), p=NR |  | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Maintenance | 5 | 1630 |  |  | wtd least squares (reported as R2WLS) and bivariate Plackett copula model (reported as R2copula), CR30 vs PFS | a) R2WLS=0.93 (0.84 to 1.00), p=NR b) R2Copula=0.89 (0.71 to 1.00), p=NR |  | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | - 1st - High FLIPI score | Chemo or immuno (induction or maintenance) | 9 | 1415 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.87 (0.68 to 0.98), p=NR b) R2Copula=0.73 (0.42 to 1.00), p=NR |  | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | - 1st - Low to intermediate FLIPI score | Chemo or immuno (induction or maintenance) | 10 | 1882 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.45 (0.02 to 0.93), p=NR b) R2Copula=0.57 (0.17 to 0.97), p=NR |  | NR | Low | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | 1st | Chemo or immuno (induction or maintenance) | 11 | 2728 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 24mo vs. logHR PFS) | a) R2WLS=0.84 (0.63 to 0.95), p=NR b) R2Copula=0.67 (0.35 to 0.99), p=NR | logHR PFS = 0.043 - 0.726 \* logOR CR24mo | NR | Medium+ | NE |
| Shi 2017129 | CR | PFS | NHL (indolent; follicular) | - 1st - Stage IV | Chemo or immuno (induction or maintenance) | NR | 2585 |  |  | a) WLSR R2 b) Bivariate Plackett copula model (logOR CR 30mo vs. logHR PFS) | a) R2WLS=0.92 (0.85 to 0.97), p=NR b) R2Copula=0.94 (0.87 to 1.00), p=NR |  | NR | Medium+ | NE |
| Colloca & Venturino 201787 | CR | PFS | Ovarian | 1st | Chemo | 12 | NR | Spearman (diff RR vs. diff med PFS) | rs=0.19, p=0.555 |  |  |  | NR | Low | NE |
| **CR vs. OS** | | | | | | | | | | | | | | | |
| Hackshaw 200596 | CR | OS | Breast | 1st | Chemo | 41\* | 9163† |  |  | WLR R2 (logOR CR vs. logHR OS) | R2=0.12, p=0.02 | logHR OS = -0.0097 + 0.13 \* logOR CR  Slope 0.13 | NR | Low | NE |
| Hackshaw 200596 | CR | OS | Breast | - 1st - Recruited pre-1990 | Chemo | 26\* | 5244† |  |  | WLR R2 (logOR CR vs. logHR OS) | R2=0.05, p=0.24 | Slope 0.09 | NR | Low | NE |
| Hackshaw 200596 | CR | OS | Breast | - 1st - Recruited 1990 or after | Chemo | 15\* | 3919† |  |  | WLR R2 (logOR CR vs. logHR OS) | R2=0.36, p=0.01 | Slope 0.16 | NR | Low | NE |
| Foster 201194 | CR | OS | Lung (SCLC) | 1st | Chemo | 3 (32 centres) | 596† | Spearman (logOR CR vs. logHR OS) | rs=0.50, p=NR | WLSR R2 (logOR CR vs. logHR OS) | R2=0.48, p=NR |  | NR | Low | NE |
| Lee 2011108 | CR | OS | NHL (aggressive) | 1st | Chemo | 36† | 16103† | Spearman: a) diff CR vs. diff 3yr OS b) diff CR vs. diff 5yr OS | a) rs=0.58 (0.29 to 0.77), p=0.004 b) rs=0.50 (0.23 to 0.74), p=0.01 |  |  |  | NR | Medium | NE |
| Lee 2011108 | CR | OS | NHL (indolent) | 1st | Chemo | 15† | 5128† | Spearman: a) diff CR vs. diff 3yr OS b) diff CR vs. diff 5yr OS | a) rs=0.41 (-0.10 to 0.74), p=0.098 b) rs=0.21 (-0.34 to 0.50), p=0.44 |  |  |  | NR | Medium | NE |
| Colloca & Venturino 201787 | CR | OS | Ovarian | 1st | Chemo | 12 | NR | Spearman (diff pCR vs. diff med OS) | rs=0.42, p=0.180 |  |  |  | NR | Low | NE |
| **DoR vs. OS** | | | | | | | | | | | | | | | |
| Colloca 2016b89 | DoR | OS | Colorectal | 1st | Bevacizumab + chemo | 5 | NR | Spearman (diff med DoR vs. diff med OS) | rs=0.70, p=0.188 |  |  |  | NR | Medium | NE |
| Colloca 2016a88 | DoR | OS | Pancreatic | 1st | Gemcitabine + chemo or targeted | 7† | NR | Spearman (diff med DoR vs. diff med OS) | rs=0.76, p=0.049 |  |  |  | NR | Medium | NE |
| Colloca 2016a88 | DoR | OS | Pancreatic | 1st | Gemcitabine + chemo | 3† | NR | Spearman (diff med DoR vs. diff med OS) | rs=0.50, p=0.667 |  |  |  | NR | Low | NE |
| Colloca 2016a88 | DoR | OS | Pancreatic | 1st | Gemcitabine + targeted | 4† | NR | Spearman (diff med DoR vs. diff med OS) | rs=0.40, p=0.600 |  |  |  | NR | Low | NE |
| \*Calculated from reported data. †Unclear for individual subgroups.  adj, adjusted; BSES2, Biomarker-Surrogate Evaluation Schema criteria 2; chemo, chemotherapy; CI, confidence interval; CR, complete response; diff, difference;; DoR, duration of response; FO, final outcome; HR, hazard ratio; IQWiG, Institute of Quality and Efficiency in Health Care; ln, natural logarithm; log, logarithm; LR, linear regression; med, median; mo, months; NE, not estimable; NHL, non-Hodgkin’s lymphoma; NR, not reported; NSCLC, non-small cell lung cancer; OR, odds ratio; ORR, overall response rate (ORR=PR+CR); OS, overall survival; PFS, progression-free survival; r, Pearson correlation; R2, regression coefficient of determination; rs, Spearman rank correlation; rr, relative risk; SCLC, small cell lung cancer; SO, surrogate outcome; STE, surrogate threshold effect; unwtd, unweighted; wtd, weighted; WLR, weighted linear regression; WLSR, weighted least squares regression. | | | | | | | | | | | | | | | |

Appendix 7: Studies excluded at full-text screening (n=135)

**Not clinical study (n=4)**

1. Ascierto PA, Long GV. Progression-free survival landmark analysis: a critical endpoint in melanoma clinical trials. Lancet Oncology 2016;17:1037-9.

2. Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. Biometrical Journal 2016;58:104-32.

3. Estey E, Othus M, Lee SJ, Appelbaum FR, Gale RP. New drug approvals in acute myeloid leukemia: what's the best end point? Leukemia 2016;30:521-5.

4. Xia Y, Cui L, Yang B. A note on breast cancer trials with pCR-based accelerated approval. Journal of Biopharmaceutical Statistics 2014;24:1102-14.

**Not meta-analysis of multiple studies (n=28)**

5. Choi SI, Yu A, Kim BH, et al. A model predicting survival of patients with recurrent or progressive hepatocellular carcinoma: the MORE score. Journal of Gastroenterology and Hepatology 2017;32:651-8.

6. Fiteni F, Bonnetain F. Surrogate end points for overall survival in breast cancer trials: A review. Breast 2016;29:44-8.

7. Fiteni F, Westeel V, Bonnetain F. Surrogate endpoints for overall survival in lung cancer trials: a review. Expert Review of Anticancer Therapy 2017;17:447-54.

8. Goorts B, van Nijnatten TJ, de Munck L, et al. Clinical tumour stage is the most important predictor of pathological complete response rate after neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Research & Treatment 2017;163:83-91.

9. Liu L, Zhao Y, Jia J, et al. The Prognostic Value of Alpha-Fetoprotein Response for Advanced-Stage Hepatocellular Carcinoma Treated with Sorafenib Combined with Transarterial Chemoembolization. Scientific Reports 2016;6.

10. Long J, Zheng JS, Sun B, Lu NN. Microwave ablation of hepatocellular carcinoma with portal vein tumour thrombosis after transarterial chemoembolization: a prospective study. Hepatology International 2016;10:175-84.

11. Mahringer-Kunz A, Weinmann A, Schmidtmann I, et al. Validation of the SNACOR clinical scoring system after transarterial chemoembolisation in patients with hepatocellular carcinoma. Bmc Cancer 2018;18.

12. Matsubara Y, Sakabayashi S, Nishimura T, et al. Surrogacy of tumour response and progression-free survival for overall survival in metastatic breast cancer resistant to both anthracyclines and taxanes. International Journal of Clinical Oncology 2011;16:623-9.

13. Mauguen A, Michiels S, Rondeau V. Joint model imputation to estimate the treatment effect on long-term survival using auxiliary events. Journal of Biopharmaceutical Statistics 2017;27:1043-53.

14. Meyer M, Hohenberger P, Apfaltrer P, et al. CT-based response assessment of advanced gastrointestinal stromal tumour: dual energy CT provides a more predictive imaging biomarker of clinical benefit than RECIST or Choi criteria. European Journal of Radiology 2013;82:923-8.

15. Michl M, Stintzing S, von Weikersthal LF, et al. CEA response is associated with tumour response and survival in patients with KRAS exon 2 wild-type and extended RAS wild-type metastatic colorectal cancer receiving first-line FOLFIRI plus cetuximab or bevacizumab (FIRE-3 trial)(aEuro). Annals of Oncology 2016;27:1565-72.

16. Mosconi C, Gramenzi A, Ascanio S, et al. Yttrium-90 radioembolization for unresectable/recurrent intrahepatic cholangiocarcinoma: a survival, efficacy and safety study. British Journal of Cancer 2016;115:297-302.

17. Motzer RJ, Bukowski RM, Figlin RA, et al. Prognostic nomogram for sunitinib in patients with metastatic renal cell carcinoma. Cancer 2008;113:1552-8.

18. Nault JC, Nkontchou G, Nahon P, et al. Percutaneous Treatment of Localized Infiltrative Hepatocellular Carcinoma Developing on Cirrhosis. Annals of Surgical Oncology 2016;23:1906-15.

19. Negrier S, Bushmakin AG, Cappelleri JC, et al. Assessment of progression-free survival as a surrogate end-point for overall survival in patients with metastatic renal cell carcinoma. European Journal of Cancer 2014;50:1766-71.

20. Ogasawara S, Chiba T, Ooka Y, et al. Analysis of Sorafenib Outcome: Focusing on the Clinical Course in Patients with Hepatocellular Carcinoma. Plos One 2016;11.

21. Park S, Kim HJ, Choi CM, et al. Predictive factors for a long-term response duration in non-squamous cell lung cancer patients treated with pemetrexed. Bmc Cancer 2016;16.

22. Plano Sánchez AI, Velasco Roces L, Zapico García I, et al. Value of αfetoprotein as an early biomarker for treatment response to sorafenib therapy in advanced hepatocellular carcinoma. Oncology letters 2018;15:8863-70.

23. Prados M, Cloughesy T, Samant M, et al. Response as a predictor of survival in patients with recurrent glioblastoma treated with bevacizumab. Neuro-Oncology 2011;13:143-51.

24. Qian Q, Zhan P, Yu L, et al. Baseline levels and decrease in serum soluble intercellular adhesion molecule-1 during chemotherapy predict objective response and survival in patients who have advanced non-small-cell lung cancer. Clinical Lung Cancer 2011;12:131-7.

25. Ramanathan RK, Goldstein D, Korn RL, et al. Positron emission tomography response evaluation from a randomized phase III trial of weekly nab-paclitaxel plus gemcitabine versus gemcitabine alone for patients with metastatic adenocarcinoma of the pancreas. Annals of Oncology 2016;27:648-53.

26. Riaz A, Gabr A, Abouchaleh N, et al. Radioembolization for hepatocellular carcinoma: Statistical confirmation of improved survival in responders by landmark analyses. Hepatology 2018;67:873-83.

27. Riedl CC, Pinker K, Ulaner GA, et al. Comparison of FDG-PET/CT and contrast-enhanced CT for monitoring therapy response in patients with metastatic breast cancer. European Journal of Nuclear Medicine & Molecular Imaging 2017;44:1428-37.

28. Salvador-Coloma C, Lorente D, Palanca S, et al. Early radiological response as predictor of overall survival in non-small cell lung cancer (NSCLC) patients with epidermal growth factor receptor mutations. Journal of Thoracic Disease 2018;10:1386-93.

29. Schwarz JK, Siegel BA, Dehdashti F, Grigsby PW. Association of posttherapy positron emission tomography with tumour response and survival in cervical carcinoma. JAMA 2007;298:2289-95.

30. Ter-Minassian M, Zhang S, Brooks NV, et al. Association Between Tumour Progression Endpoints and Overall Survival in Patients with Advanced Neuroendocrine Tumours. Oncologist 2017;22:165-72.

31. Yau T, Yao TJ, Chan P, et al. The Significance of Early Alpha-Fetoprotein Level Changes in Predicting Clinical and Survival Benefits in Advanced Hepatocellular Carcinoma Patients Receiving Sorafenib. Oncologist 2011;16:1270-9.

32. Zabor EC, Heller G, Schwartz LH, Chapman PB. Correlating Surrogate Endpoints with Overall Survival at the Individual Patient Level in BRAFV600E-Mutated Metastatic Melanoma Patients Treated with Vemurafenib. Clinical Cancer Research 2016;22:1341-7.

**Neoadjuvant or adjuvant treatment (n=10)**

33. Berruti A, Amoroso V, Gallo F, et al. Pathologic complete response as a potential surrogate for the clinical outcome in patients with breast cancer after neoadjuvant therapy: a meta-regression of 29 randomized prospective studies. Journal of Clinical Oncology 2014;32:3883-91.

34. Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: surrogacy in question? European Journal of Cancer 2012;48:1781-90.

35. Broglio KR, Quintana M, Foster M, et al. Association of Pathologic Complete Response to Neoadjuvant Therapy in HER2-Positive Breast Cancer With Long-Term Outcomes: A Meta-Analysis. JAMA Oncol 2016;2:751-60.

36. Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 2014;384:164-72.

37. Korn EL, Sachs MC, McShane LM. Statistical controversies in clinical research: assessing pathologic complete response as a trial-level surrogate end point for early-stage breast cancer. Annals of Oncology 2016;27:10-5.

38. Nekljudova V, Loibl S, von Minckwitz G, et al. Trial-level prediction of long-term outcome based on pathologic complete response (pCR) after neoadjuvant chemotherapy for early-stage breast cancer (EBC). Contemporary Clinical Trials 2018;71:194-8.

39. Petrelli F, Borgonovo K, Cabiddu M, Ghilardi M, Lonati V, Barni S. Pathologic complete response and disease-free survival are not surrogate endpoints for 5-year survival in rectal cancer: an analysis of 22 randomized trials. Journal of Gastrointestinal Oncology 2017;8:39-48.

40. Petrelli F, Coinu A, Cabiddu M, Ghilardi M, Vavassori I, Barni S. Correlation of pathologic complete response with survival after neoadjuvant chemotherapy in bladder cancer treated with cystectomy: a meta-analysis. European Urology 2014;65:350-7.

41. Petrelli F, Tomasello G, Barni S. Surrogate end-points for overall survival in 22 neoadjuvant trials of gastro-oesophageal cancers. European Journal of Cancer 2017;76:8-16.

42. Petrelli F, Tomasello G, Ghidini M, Lonati V, Passalacqua R, Barni S. Disease-free survival is not a surrogate endpoint for overall survival in adjuvant trials of pancreatic cancer: a systematic review of randomized trials. Hpb 2017;19:944-50.

**No relevant outcomes (n=67)**

43. Aggarwal C, Borghaei H. Treatment Paradigms for Advanced Non-Small Cell Lung Cancer at Academic Medical Centers: Involvement in Clinical Trial Endpoint Design. Oncologist 2017;22:700-8.

44. Araujo RLC, Herman P, Riechelmann RP. Recurrence-free survival as a putative surrogate for overall survival in phase. trials of curative-intent treatment of colorectal liver metastases: Systematic review. World Journal of Clinical Oncology 2017;8:266-72.

45. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. Journal of Clinical Oncology 2007;25:5218-24.

46. Cai J, Ma H, Huang F, et al. Correlation of bevacizumab-induced hypertension and outcomes of metastatic colorectal cancer patients treated with bevacizumab: a systematic review and meta-analysis. World Journal of Surgical Oncology 2013;11:306.

47. Chirila C, Odom D, Devercelli G, et al. Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer. International Journal of Colorectal Disease 2012;27:623-34.

48. Colloca GA, Venturino A, Guarneri D. Early tumour shrinkage after first-line medical treatment of metastatic colorectal cancer: a meta-analysis. International journal of clinical oncology 2019:1-10.

49. Forsythe A, Chandiwana D, Barth J, et al. Is progression-free survival a more relevant endpoint than overall survival in first-line HR+/ HER2- Metastatic breast cancer? Cancer Management and Research 2018;10:1015-25.

50. Forsythe A, Chandiwana D, Barth J, Thabane M, Baeck J, Tremblay G. Progression-free survival/time to progression as a potential surrogate for overall survival in HR+, HER2- metastatic breast cancer. Breast Cancer-Targets and Therapy 2018;10:69-78.

51. Foster NR, Renfro LA, Schild SE, et al. Multitrial Evaluation of Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Extensive-Stage Small-Cell Lung Cancer. Journal of Thoracic Oncology 2015;10:1099-106.

52. Francini E, Petrioli R, Rossi G, Laera L, Roviello G. PSA response rate as a surrogate marker for median overall survival in docetaxel-based first-line treatments for patients with metastatic castration-resistant prostate cancer: an analysis of 22 trials. Tumour Biology 2014;35:10601-7.

53. Glynne-Jones R, Mawdsley S, Pearce T, Buyse M. Alternative clinical end points in rectal cancer--are we getting closer? Annals of Oncology 2006;17:1239-48.

54. Grunwald V, Lin X, Kalanovic D, Simantov R. Early Tumour Shrinkage: A Tool for the Detection of Early Clinical Activity in Metastatic Renal Cell Carcinoma. European Urology 2016;70:1006-15.

55. Grunwald V, McKay RR, Krajewski KM, et al. Depth of remission is a prognostic factor for survival in patients with metastatic renal cell carcinoma. European Urology 2015;67:952-8.

56. Gyawali B, Hey SP, Kesselheim AS. A Comparison of Response Patterns for Progression-Free Survival and Overall Survival Following Treatment for Cancer With PD-1 Inhibitors A Meta-analysis of Correlation and Differences in Effect Sizes. Jama Network Open 2018;1.

57. Harshman LC, Xie WL, Moreira RB, et al. Evaluation of Disease-Free Survival as an Intermediate Metric of Overall Survival in Patients With Localized Renal Cell Carcinoma: A Trial-Level Meta-Analysis. Cancer 2018;124:925-33.

58. Hasan B, Greillier L, Pallis A, et al. Progression free survival rate at 9 and 18 weeks predict overall survival in patients with malignant pleural mesothelioma: an individual patient pooled analysis of 10 European Organisation for Research and Treatment of Cancer Lung Cancer Group studies and an independent study validation. European Journal of Cancer 2014;50:2771-82.

59. Hotta K, Kiura K, Fujiwara Y, et al. Role of Survival Post-Progression in Phase III Trials of Systemic Chemotherapy in Advanced Non-Small-Cell Lung Cancer: A Systematic Review. Plos One 2011;6.

60. Huan HB, Wu LL, Lau WY, et al. Surrogate endpoint for overall survival in assessment of adjuvant therapies after curative treatment for hepatocellular carcinoma: a re-analysis of meta-analyses of individual patients' data. Oncotarget 2017;8:90291-300.

61. Jiang ZC, Ding P, Geng Z. Principal causal effect identification and surrogate end point evaluation by multiple trials. Journal of the Royal Statistical Society Series B-Statistical Methodology 2016;78:829-48.

62. Kasuga A, Hamamoto Y, Takeuchi A, et al. Positive relationship between subsequent chemotherapy and overall survival in pancreatic cancer: meta-analysis of postprogression survival for first-line chemotherapy. Cancer Chemotherapy and Pharmacology 2017;79:595-602.

63. Kasuga A, Hamamoto Y, Takeuchi A, et al. Post-progression survival following second-line chemotherapy in patients with advanced pancreatic cancer previously treated with gemcitabine: a meta-analysis. Investigational New Drugs 2018;36:939-48.

64. Kataoka K, Nakamura K, Mizusawa J, et al. Surrogacy of progression-free survival (PFS) for overall survival (OS) in esophageal cancer trials with preoperative therapy: Literature-based meta-analysis. Ejso 2017;43:1956-61.

65. Kudo M, Izumi N, Sakamoto M, et al. Survival Analysis over 28 Years of 173,378 Patients with Hepatocellular Carcinoma in Japan. Liver Cancer 2016;5:190-7.

66. Kundu MG, Acharyya S. Surrogacy of progression free survival for overall survival in metastatic breast cancer studies: Meta-analyses of published studies. Contemporary Clinical Trials 2017;53:20-8.

67. Kurokawa Y, Shibata T, Sasako M, et al. Validity of response assessment criteria in neoadjuvant chemotherapy for gastric cancer (JCOG0507-A). Gastric Cancer 2014;17:514-21.

68. Lee DW, Jang MJ, Lee KH, et al. TTP as a surrogate endpoint in advanced hepatocellular carcinoma treated with molecular targeted therapy: meta-analysis of randomised controlled trials. British Journal of Cancer 2016;115:1201-5.

69. Li CF, Xiang AZ, Chen XZ, Yin K, Lu JS, Yin WJ. Optimizing the treatment of bevacizumab as first-line therapy for human epidermal growth factor receptor 2 (HER2)-negative advanced breast cancer: an updated meta-analysis of published randomized trials. Oncotargets and Therapy 2017;10:3155-68.

70. Li DY, Lv H, Hao XY, Dong YD, Dai HX, Song YS. Prognostic value of bone scan index as an imaging biomarker in metastatic prostate cancer: a meta-analysis. Oncotarget 2017;8:84449-58.

71. Li M, Dave N, Salem AH, Freise KJ. Model-based meta-analysis of progression-free survival in non-Hodgkin lymphoma patients. Medicine 2017;96.

72. Lien K, Georgsdottir S, Sivanathan L, Chan K, Emmenegger U. Low-dose metronomic chemotherapy: a systematic literature analysis. European Journal of Cancer 2013;49:3387-95.

73. Liu Y, Waterton JC. Imaging Biomarkers in Clinical Trials. Imaging Biomarkers: Springer; 2017:295-306.

74. Loupakis F, Bria E, Vaccaro V, et al. Magnitude of benefit of the addition of bevacizumab to first-line chemotherapy for metastatic colorectal cancer: meta-analysis of randomized clinical trials. Journal of Experimental & Clinical Cancer Research 2010;29:58.

75. Martens A, Wistuba-Hamprecht K, Geukes Foppen M, et al. Baseline Peripheral Blood Biomarkers Associated with Clinical Outcome of Advanced Melanoma Patients Treated with Ipilimumab. Clinical Cancer Research 2016;22:2908-18.

76. Mauguen A, Pignon JP, Burdett S, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. Lancet Oncology 2013;14:619-26.

77. Meeks JJ, Bellmunt J, Bochner BH, et al. A Systematic Review of Neoadjuvant and Adjuvant Chemotherapy for Muscle-invasive Bladder Cancer. European Urology 2012;62:523-33.

78. Mehra N, Dolling D, Sumanasuriya S, et al. Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). European Urology 2018;74:283-91.

79. Michiels S, Pugliano L, Marguet S, et al. Progression-free survival as surrogate end point for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer. Annals of Oncology 2016;27:1029-34.

80. Miura Y, Imai H, Sakurai R, et al. The effect of post-progression survival on overall survival among patients with sensitive relapse of small cell lung cancer. Medical Oncology 2018;35.

81. Miyake H, Harada K-i, Ozono S, Fujisawa M. Prognostic Significance of Early Tumour Shrinkage Under Second-Line Targeted Therapy for Metastatic Renal Cell Carcinoma: A Retrospective Multi-Institutional Study in Japan. Molecular Diagnosis & Therapy 2016;20:385-92.

82. Miyake H, Miyazaki A, Imai S, Harada K-i, Fujisawa M. Early Tumour Shrinkage Under Treatment with First-line Tyrosine Kinase Inhibitors as a Predictor of Overall Survival in Patients with Metastatic Renal Cell Carcinoma: a Retrospective Multi-Institutional Study in Japan. Targeted Oncology 2016;11:175-82.

83. Mocellin S, Baretta Z, Roque IFM, et al. Second-line systemic therapy for metastatic colorectal cancer. Cochrane Database of Systematic Reviews 2017;1:CD006875.

84. Montagnani F, Di Leonardo G, Pino MS, et al. Progression-free Survival as a Surrogate End-point in Advanced Colorectal Cancer Treated with Antiangiogenic Therapies. Anticancer Research 2016;36:4259-65.

85. Morris PD, Laurence JM, Yeo D, et al. Can Response to Locoregional Therapy Help Predict Longterm Survival After Liver Transplantation for Hepatocellular Carcinoma? A Systematic Review. Liver Transplantation 2017;23:375-85.

86. Munshi NC, Avet-Loiseau H, Rawstron AC, et al. Association of Minimal Residual Disease With Superior Survival Outcomes in Patients With Multiple Myeloma: A Meta-analysis. JAMA Oncology 2017;3:28-35.

87. Othus M, van Putten W, Lowenberg B, et al. Relationship between event-free survival and overall survival in acute myeloid leukemia: a report from SWOG, HOVON/SAKK, and MRC/NCRI. Haematologica 2016;101:284-6.

88. Petrelli F, Coinu A, Cabiddu M, et al. Early analysis of surrogate endpoints for metastatic melanoma in immune checkpoint inhibitor trials. Medicine 2016;95.

89. Peveling-Oberhag J, Arcaini L, Bankov K, Zeuzem S, Herrmann E. The anti-lymphoma activity of antiviral therapy in HCV-associated B-cell non-Hodgkin lymphomas: a meta-analysis. Journal of Viral Hepatitis 2016;23:536-44.

90. Polley MYC, Lamborn KR, Chang SM, Butowski N, Clarke JL, Prados M. Six-month progression-free survival as an alternative primary efficacy endpoint to overall survival in newly diagnosed glioblastoma patients receiving temozolomide. Neuro-Oncology 2010;12:274-82.

91. Pujol JL, Pirker R, Lynch TJ, et al. Meta-analysis of individual patient data from randomized trials of chemotherapy plus cetuximab as first-line treatment for advanced non-small cell lung cancer. Lung Cancer 2014;83:211-8.

92. Rotolo F, Pignon JP, Bourhis J, et al. Surrogate End Points for Overall Survival in Loco-Regionally Advanced Nasopharyngeal Carcinoma: An Individual Patient Data Meta-analysis. Jnci-Journal of the National Cancer Institute 2017;109.

93. Shafrin J, Brookmeyer R, Peneva D, et al. The value of surrogate endpoints for predicting real-world survival across five cancer types. Current Medical Research and Opinion 2016;32:731-9.

94. Shi Q, Schmitz N, Ou FS, et al. Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Diffuse Large B-Cell Lymphoma: An Individual Patient-Level Analysis of Multiple Randomized Trials (SEAL). Journal of Clinical Oncology 2018;36:2593-+.

95. Shih T, Lindley C. Bevacizumab: an angiogenesis inhibitor for the treatment of solid malignancies. Clinical Therapeutics 2006;28:1779-802.

96. Shimokawa M, Kogawa T, Shimada T, et al. Overall survival and post-progression survival are potent endpoint in phase III trials of second/third-line chemotherapy for advanced or recurrent epithelial ovarian cancer. Journal of Cancer 2018;9:872-9.

97. Sjoquist KM, Lord SJ, Friedlander ML, Simes RJ, Marschner IC, Lee CK. Progression-free survival as a surrogate endpoint for overall survival in modern ovarian cancer trials: a meta-analysis. Therapeutic Advances in Medical Oncology 2018;10:1-16.

98. Song SY, Seo H, Kim G, Kim AR, Kim EY. Trends in endpoint selection in clinical trials of advanced breast cancer. Journal of Cancer Research and Clinical Oncology 2016;142:2403-13.

99. Sonpavde G, Pond GR, Rosenberg JE, et al. Complete response as an intermediate end point in patients receiving salvage systemic therapy for urothelial carcinoma. Clinical Genitourinary Cancer 2015;13:185-92.

100. Suciu S, Eggermont AM, Lorigan P, et al. Relapse-free survival as a surrogate for overall survival in the evaluation of stage II–III melanoma adjuvant therapy. JNCI: Journal of the National Cancer Institute 2017;110:87-96.

101. Tan A, Porcher R, Crequit P, Ravaud P, Dechartres A. Differences in Treatment Effect Size Between Overall Survival and Progression-Free Survival in Immunotherapy Trials: A Meta-Epidemiologic Study of Trials With Results Posted at ClinicalTrials.gov. Journal of Clinical Oncology 2017;35:1686-+.

102. Terashima T, Yamashita T, Takata N, et al. Post-progression survival and progression-free survival in patients with advanced hepatocellular carcinoma treated by sorafenib. Hepatology Research 2016;46:650-6.

103. Terashima T, Yamashita T, Toyama T, et al. Surrogacy of Time to Progression for Overall Survival in Advanced Hepatocellular Carcinoma Treated with Systemic Therapy: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. Liver Cancer 2019;8:130-9.

104. Trotman J, Luminari S, Boussetta S, et al. Prognostic value of PET-CT after first-line therapy in patients with follicular lymphoma: a pooled analysis of central scan review in three multicentre studies. Lancet Haematology 2014;1:E17-E27.

105. Valentini V, van Stiphout RG, Lammering G, et al. Selection of appropriate end-points (pCR vs 2yDFS) for tailoring treatments with prediction models in locally advanced rectal cancer. Radiotherapy & Oncology 2015;114:302-9.

106. van Mackelenbergh MT, Denkert C, Nekljudova V, et al. Outcome after neoadjuvant chemotherapy in estrogen receptor-positive and progesterone receptor-negative breast cancer patients: a pooled analysis of individual patient data from ten prospectively randomized controlled neoadjuvant trials. Breast Cancer Research & Treatment 2018;167:59-71.

107. Xie WL, Regan MM, Buyse M, et al. Metastasis-Free Survival Is a Strong Surrogate of Overall Survival in Localized Prostate Cancer. Journal of Clinical Oncology 2017;35:3097-104.

108. Xu XS, Ryan CJ, Stuyckens K, et al. Correlation between Prostate-Specific Antigen Kinetics and Overall Survival in Abiraterone Acetate-Treated Castration-Resistant Prostate Cancer Patients. Clinical Cancer Research 2015;21:3170-7.

109. Zhao S, Zhang Z, Zhang Y, et al. Progression‐free survival and one‐year milestone survival as surrogates for overall survival in previously treated advanced non‐small cell lung cancer. International journal of cancer 2018.

**No correlation coefficient or R2 (n=13)**

110. Lee CK, Lord S, Marschner I, et al. The Value of Early Depth of Response in Predicting Long-Term Outcome in EGFR-Mutant Lung Cancer. Journal of Thoracic Oncology 2018;13:792-800.

111. Loupakis F, Cremolini C, Salvatore L, et al. Clinical impact of anti-epidermal growth factor receptor monoclonal antibodies in first-line treatment of metastatic colorectal cancer. Cancer 2012;118:1523-32.

112. Mandrekar SJ, Qi Y, Hillman SL, et al. Endpoints in Phase II trials for advanced non-small cell lung cancer. Journal of Thoracic Oncology 2010;5:3-9.

113. Mardis M, Davis J, Benningfield B, et al. Shift-to-Shift Handoff Effects on Patient Safety and Outcomes: A Systematic Review. American Journal of Medical Quality 2017;32:34-42.

114. Meyer T, Palmer DH, Cheng AL, Hocke J, Loembe AB, Yen CJ. mRECIST to predict survival in advanced hepatocellular carcinoma: Analysis of two randomised phase II trials comparing nintedanib vs sorafenib. Liver International 2017;37:1047-55.

115. Ming-Wen A, Xinxin D, Meyers J, et al. Evaluating Continuous Tumour Measurement-Based Metrics as Phase II Endpoints for Predicting Overall Survival. JNCI: Journal of the National Cancer Institute 2015;107:1-7.

116. Okines AF, Gonzalez de Castro D, Cunningham D, et al. Biomarker analysis in oesophagogastric cancer: Results from the REAL3 and TransMAGIC trials. European Journal of Cancer 2013;49:2116-25.

117. Park WH, Shim JH, Han SB, et al. Clinical utility of des-gamma-carboxyprothrombin kinetics as a complement to radiologic response in patients with hepatocellular carcinoma undergoing transarterial chemoembolization. Journal of Vascular & Interventional Radiology 2012;23:927-36.

118. Pinato DJ, Arizumi T, Jang JW, et al. Combined sequential use of HAP and ART scores to predict survival outcome and treatment failure following chemoembolization in hepatocellular carcinoma: a multi-center comparative study. Oncotarget 2016;7:44705-18.

119. Schwaederle M, Zhao M, Lee JJ, et al. Association of Biomarker-Based Treatment Strategies With Response Rates and Progression-Free Survival in Refractory Malignant Neoplasms: A Meta-analysis. JAMA Oncology 2016;2:1452-9.

120. Shao YY, Lin ZZ, Hsu C, Shen YC, Hsu CH, Cheng AL. Early alpha-fetoprotein response predicts treatment efficacy of antiangiogenic systemic therapy in patients with advanced hepatocellular carcinoma. Cancer 2010;116:4590-6.

121. Wong D, Ko AH, Hwang J, Venook AP, Bergsland EK, Tempero MA. Serum CA19-9 Decline Compared to Radiographic Response as a Surrogate for Clinical Outcomes in Patients With Metastatic Pancreatic Cancer Receiving Chemotherapy. Pancreas 2008;37:269-74.

122. Zhang YJ, Zhang MP, Chen MS, et al. Association of Sustained Response Duration With Survival After Conventional Transarterial Chemoembolization in Patients With Hepatocellular Carcinoma. Jama Network Open 2018;1.

**Secondary publication, no additional data (n=10)**

123. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. J Roy Stat Soc a Sta 2004;167:103-24.

124. Lee LM, Wang L, Crump M. Identification of Potential Surrogate Endpoints in Randomized Clinical Trials of Aggressive Non-Hodgkin Lymphoma: Correlation of Complete Response, Time-to-Event and Overall Survival Data. Blood 2009;114:1424-, Abstract 3699.

125. Narkhede MS, Cheson BD. Surrogate Endpoints and Risk Adaptive Strategies in Previously Untreated Follicular Lymphoma. Clinical Lymphoma Myeloma & Leukemia 2018;18:447-51.

126. Rotolo F, Paoletti X, Burzykowski T, Buyse M, Michiels S. A Poisson approach to the validation of failure time surrogate endpoints in individual patient data meta-analyses. Statistical Methods in Medical Research 2019;28:170-83.

127. Rotolo F, Paoletti X, Michiels S. surrosurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. Computer Methods and Programs in Biomedicine 2018;155:189-98.

128. Savina M, Gourgou S, Italiano A, et al. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: A critical review. Critical Reviews in Oncology Hematology 2018;123:21-41.

129. Sherrill B, Amonkar M, Wu Y, et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. British Journal of Cancer 2008;99:1572-8.

130. Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. Oncotargets and Therapy 2012;5:287-96.

131. Van der Elst W, Molenberghs G, Alonso A. Exploring the relationship between the causal-inference and meta-analytic paradigms for the evaluation of surrogate endpoints. Statistics in Medicine 2016;35:1281-98.

132. Xie W, Halabi S, Tierney JF, et al. A Systematic Review and Recommendation for Reporting of Surrogate Endpoint Evaluation using Meta-analyses (ReSEEM). JNCI Cancer Spectrum 2019.

**Insufficient data reported (n=1)**

133. Zierhut ML, Puchalski T, Rizo A, et al. Is Complete Remission Rate Predictive of Median Overall Survival? A Model-Based Meta-Analysis in Patients with Acute Myeloid Leukemia. Journal of Pharmacokinetics and Pharmacodynamics 2017;44:S59-S.

**Non-English and insufficient detail (n=1)**

134. Paoletti X, Rotolo F, Michiels S. How a biomarker can become an acceptable substitution criteria ? Bulletin Du Cancer 2016;103:S63-S70.

**Not available (n=1)**

135. Woo KM, Gonen M, Schnorr G, Silvestri GA, Bach PB. Surrogate Markers and the Association of Low-Dose CT Lung Cancer Screening With Mortality. Jama Oncology 2018;4:1006-8.

Appendix 8: A targeted review of published NICE technology appraisals where initial marketing authorisation was based on response outcomes from single-arm studies

For each TA, a summary of the main clinical studies supporting the intervention being assessed along with the number of patients, the study location, the magnitude of the ORR and whether PFS and OS data were available were extracted. Full results can be seen in Table 33.

Table 33: Summary of clinical studies supporting the NICE TAs

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Intervention** | **NICE TA** | **Clinical Study** | **Sample Size (n)** | **Number of Countries** | **ORR Outcomes (response rate, 95% CI)** | **PFS Outcomes Median, months (95% CI)** | **OS outcomes Median, months (95% CI)** |
| Ceritinib | TA395 | ASCEND-1 | 163 | Multiple | 92% (48.5, 64.2) | 6.9 (5.6, 8.7) | 16.7 (14.78, NE) |
| ASCEND-2 | 140 | Multiple | 54% (30.5, 47.2) | 5.7 (5.4, 7.6) | 14.9 (13.5, NE) |
| Osimertinib | TA416 | AURA ext | 201 | Multiple | 61.3% (54.2, 68.1) | NC (8.1, NC) | NC |
| AURA2 | 210 | Multiple | 70.9% (64.0, 77.1) | 8.6 (8.3, 9.7) | NC |
| Pooled | 411 | Multiple | 66.1% (61.2, 70.7) | 9.7 (8.3, NC) | NC |
| Nivolumab | TA462 | CheckMate 205 cohort B | 80 | Multiple | 67.5% (57.2, 77.8) | 14.78 (11.33, NA) | NR |
| CheckMate 205 cohort C | 98 | Multiple | 73.0% (64.3, 81.7) | 11.17 (8.51, NA) | NR |
| CA209-039 | 15 | Multiple | 60% (NR) | 12.65 (5.91, NA) | NR |
| Venetoclax | TA487 | M12-175 | 67 | Multiple | 82.1% (70.8, 90.4) | 41.4 (17.7-41.5) | NA |
| M13-982 | 158 | Multiple | 77.2% (66.9, 83.5) | 27.2 (21.9, NA) | NR |
| M14-032 (Cohort A) | 43 | Single | 67.4% (51.5, 80.9) | NR | NR |
| M14-032 (Cohort B) | 21 | Single | 57.1% (34.0, 78.2) | NR | NR |
| Atezolizumab | TA492 | IMvigor 210 Cohort 1 | 119 | Multiple | 19.3% (12.66, 27.58) | 2.7 (2.1, 4.2) | 15.9 (10.4, NE) |
| IMvigor 210 Cohort 2 | 310 | Multiple | 15.1% (11.3, 19.6) | 2.1 (2.1, 2.1) | 7.9 (6.7, 9.3) |
| Daratumumab | TA510 | MMY2002 | 106 | Multiple | 31 (20.8, 38.9%) | 3.7 (2.8, 4.6) | 18.6 (13.7, NR) |
| GEN501 | 42 | Multiple | 15 (21.6, 52.0%) | 6.2 (4.2, 11.6) | NR (18.7, NR) |
| Pooled | 148 | Multiple | 46 (23.7, 39.2%) | 4.0 (2.8, 5.6) | 20.1 (16.6, NR) |
| Avelumab | TA517 | JAVELIN Part A | 88 | Multiple | 33.0 (23.3-43.8) | 2.7 (0.03, 28.9) | 12.9 (7.5, NE)\*\*\*\* |
| JAVELIN Part B | 39 | Multiple | 62.1 (42.3- 79.3)\*\* | 9.1 (1.9, NR) | NR (9.1,NR)\*\*\* |
| Crizotininb | TA529 | PROFILE 1001 | 53 | Multiple | 69.8%; (55.7, 81.7) | 19.3 (14.8, NR) | NR |
| Pembrolizumab | TA540 | KEYNOTE-087 Cohort 1 | 69 | Multiple | 75.4% (63.5, 84.9) | 16.7 (11.2, NR) | NA |
| KEYNOTE-087 Cohort 2 | 81 | Multiple | 66.7% (55.3, 76.8) | 11.1 (7.6, 13.7) | NA |
| Brigatinib | ID1328  (Ongoing) | ALTA (Arm A) | 110 | Multiple | 56.4 (45.2, 67.0)\* | 15.6 (11.1, 21.0) | 34.1 (27.7, NR) |
| ALTA (Arm A) | 112 | Multiple | 45.5 (34.8, 56.5)\* | 9.2 (7.4, 11.1) | 29.5 (18.2, NR) |
| Study 101 | 25 | Multiple | 76 (54.9-90.6)\* | 16.3 (9.2, NE)  Range: 0.5-27.8) | NR (1.4, 24.3) |

\*97.5% CI for ALTA ORR (investigator), \*\* 3 month follow up n=29, \*\*\* Full analysis n=39, \*\*\*\* 18 month follow up not reported. Note, no IQR was provided except where specified. NE, not estimable; NC, not calculable; NR, not reached; NA, not available

**Issue 1: ORR as a primary endpoint and possible surrogate relationship assumed for PFS/OS** Given the different maturity in PFS and OS data evident in Table 11, the review further subdivides the TAs based on whether median PFS and OS were both reached and those where only median PFS was reached. This allowed consideration of whether the approaches to dealing with different levels of maturity in these endpoints differed.

All 10 TAs used a model structure based on a partitioned survival model (PSM) or “area under the curve” analysis comprising of three mutually exclusive health states: (i) PFS (progression free), (ii) progressive disease (PD; progression), and (iii) death. Importantly, despite ORR being the primary endpoint supporting marketing authorisation, none of the TAs made use of the ORR or DoR data. Instead, the proposed approaches relied on extrapolations of the available PFS and OS data or used external evidence and/or assumptions.

*Median PFS and OS reached*

The 7 TAs for which median PFS and OS had been reached are:

* TA395: Ceritinib for ALK+ NSCLC;
* TA492: Atezolizumab for PD-L1+ urothelial cancer;
* TA510: Daratumumab for relapsed and refractory multiple myeloma;
* TA462: Nivolumab for relapsed or refractory classical Hodgkin lymphoma (RRcHL);
* TA540: Pembrolizumab for RRcHL;
* TA487: Venetoclax for chronic lymphocytic leukaemia;
* ID1328: Brigatinib for ALK+ NSCLC.

The observed survival data in each TA was extrapolated over a lifetime horizon using conventional parametric survival modelling. In accordance with the NICE DSU Technical Support Document 14,181 each company approached the data limitation by fitting various candidate distributions to the observed data, assessing statistical goodness of fit and clinical plausibility. Although median PFS and OS were reached in the clinical studies, the ERGs consistently highlighted concerns with the immaturity of the OS data, relative to the long extrapolation periods applied within the economic models.

There appeared to be no obvious trend in terms of the committee’s final decision based solely on median OS having been reached. Atezolizumab and daratumumab both received recommendations for use in the CDF due to uncertainty in their respective incremental cost-effectiveness ratio (ICER) estimates attributed to survival data immaturity.212, 213 Venetoclax also received a recommendation for use in the CDF. However, the uncertainty raised by the committee for this TA centred on the trial population and whether their disease severity reflected those in the NHS and not specifically uncertainty in the survival data.214 Nivolumab and brigatinib both received recommendations for routine use in the NHS despite immaturity of the survival data being highlighted by the committee and the ERGs.215, 216

*Median PFS reached, median OS not reached*

The TAs based on studies that had not reached median OS were:

* TA416: Osimertinib for EGFR T790M mutation-positive NSCLC;217
* TA517: Avelumab for merkel cell carcinoma (MCC);218
* TA529: Crizotinib for ROS1-positive NSCLC.219

For these appraisals there was a clearer trend in the final NICE decision given the greater uncertainty surrounding the OS data. Recommendations for all three products were restricted to use in the CDF. There was also greater variation in the modelling approaches used to extrapolate OS data.

The TA of osimertinib217 included data from two studies, neither of which reached median OS at the time of the NICE appraisal. Despite the immaturity of the OS evidence, the extrapolation of OS was still undertaken using conventional parametric survival modelling. The committee concluded that these extrapolations were highly uncertain based on the very immature OS data, making it difficult to determine a robust cost-effectiveness estimate. Osimertinib was subsequently approved for use within the CDF despite the lack of robustness of the ICER estimates. The most critical factor appeared to be the committee’s view that there was plausible potential in the ICER estimates and that the uncertainties in OS would be addressed by an ongoing Phase III RCT.

For the TA of avelumab,218 the evidence base was derived from two cohorts: (i) treatment-experienced (second-line and further [2L+]) metastatic patients, (JAVELIN Part A) and (ii) treatment-naïve (first-line [1L]) metastatic patients (JAVELIN Part B). Each cohort was considered separately reflecting a potentially different position of avelumab in the pathway. However, important differences in data maturity were evident in the 2L+ (median PFS and OS reached) and 1L (mean PFS but not OS reached) positions. Recruitment of 1L patients was also reported to be ongoing, such that more mature survival data were expected over time.

For the 1L+ population, the company considered that the data were too immature to be extrapolated. As an alternative to extrapolating based on the immature evidence, the company proposed to estimate the relative improvement with avelumab that might be seen in treatment-naïve patients versus those in the treatment-experienced group. The company elicited a hypothetical hazard ratio (HR) for PFS and OS from clinical experts. The elicited HRs were then applied to the treatment-experienced avelumab PFS and OS curves (based on more mature evidence) to estimate equivalent estimates for treatment-naïve patients receiving avelumab.

The ERG expressed significant concerns regarding the approach employed to adjust treatment effectiveness between treatment lines. Despite the immaturity in the survival data, the ERG expressed a preference to use independent survival functions fitted to the available PFS and OS data rather than using elicited HRs. However, the ERG also noted that using the observed survival data did not solve the fundamental issue of data immaturity in the treatment-naïve population.

For 1L treatment of MCC, NICE recommended avelumab for use in the CDF. This reflected the committee’s concerns regarding the immaturity of the PFS and OS data and the proposed use of clinical assumptions rather than direct evidence. The committee acknowledged that ongoing data collection in JAVELIN part B would reduce the uncertainty about the progression-free and overall survival benefit and that there was plausible potential for first-line use of avelumab to be cost-effective, if further trial data proved favourable.

Crizotinib for treating ROS1-positive advanced non-small-cell lung cancer was the only TA in which the evidence of clinical benefit was based on a single clinical study. The clinical-effectiveness evidence was based on a single-arm study (n=53), with a median follow-up of 25.4 months. Median OS had not been reached at the time of the appraisal. Due to the small study and immature survival data from this study, the company proposed the use of more mature PFS and OS data from previous RCTs of crizotinib for ALK-positive NSCLC as a proxy for ROS1-positive patients.

The committee considered the use of proxy data for ROS1-positive patients from an RCT to be more robust than using the available immature ROS1 positive PFS and OS curves from a single-arm study. However, using data from a proxy population was concluded to be far from ideal, making the assessment of clinical and cost-effectiveness highly uncertain. Although the committee agreed to explore the proxy data in its decision-making, the committee also stated that the approach was very unusual and should not set a precedent for such an approach in future appraisals.

Given these uncertainties, crizotinib was not recommended for routine use in the NHS for patients. However, crizotinib was recommended for use in the CDF. The committee considered that further data on the use of crizotinib within the CDF would help address uncertainties in existing survival data estimates, particularly the comparability of ROS1-positive and ALK-positive advanced NSCLC.

**Issue 2: Challenges of heterogeneous populations**

Due to the nature of basket trials, significant heterogeneity may be present in the study populations enrolled in the trials. The potential importance of accounting for heterogeneity and exploring the cost-effectiveness in subgroups of the target population is acknowledged in the current NICE methods guide.4 Differences in the cost-effectiveness and decision uncertainty across these separate subgroups may lead to an optimised recommendation that is more restrictive than the marketing authorisation.

Although the review of the 10 TAs provided examples of appraisals where heterogeneity had been accounted for within an overall target population, the evidence for the separate subgroups was commonly derived from separate studies relevant to specific subgroups or from studies where there were relatively large numbers of patients to undertake meaningful subgroup analysis. These appraisals also typically considered only a small number of subgroups, most commonly based on alternative positions of a new treatment in an existing pathway (i.e. 1st or 2nd line). While these findings are helpful in demonstrating the potential importance of accounting for heterogeneity within a target population, important differences are also expected for histology-independent appraisals given the potential for a much larger number of potential subsets and smaller sample sizes. The review was subsequently broadened to consider select additional TAs where issues related to heterogeneity were considered more similar to those expected for histology-independent appraisals.

The NICE appraisals for neuroendocrine tumours (NET), considered in TA449 (everolimus and sunitinib) and TA539 (lutetium), appear particularly relevant. NETs affect different organs, namely the pancreas, gastrointestinal tissue, and lungs. The broad population covered by the marketing authorisation was acknowledged in the NICE scoping documents, which stated that the relevant population was people with progressed unresectable or metastatic neuroendocrine tumours according to the specific locations covered by the existing and anticipated marketing authorisations. However, heterogeneity with the licensed population was recognised and the NICE scopes also stated that the location of tumour should be considered as a basis for identifying possible subgroups.

In both TAs, the NICE committee considered each organ separately and issued optimised recommendations based on tumour site. For example, everolimus and sunitinib were both recommended for pancreatic NET, while everolimus only was recommended for gastrointestinal and lung NET. The optimised recommendations were possible because the companies either submitted separate evidence for the different sites or provided subgroup analysis related to specific organs. In these appraisals the Committee acknowledged the importance of considering each organ separately, noting that prognosis, quality of life and cost of comparator therapies were likely to differ, impacting the cost-effectiveness estimates.

A similar example is found in the appraisal of denosumab for the prevention of skeletal-related events in adults with bone metastases from solid tumours (TA265). The scope of this appraisal covered a broad population characterised by a wide range of histologies, as almost any form of solid tumour can metastasise to the bone. Again, the NICE scope acknowledged possible heterogeneity within the licensed population and suggested that the appraisal should also consider patient subgroups based on location or type of primary cancer.

Separate studies were available for TA265 for different tumour types, thus allowing for separate clinical and cost-effectiveness analyses to be performed. For example, the company submitted a model assessing the cost-effectiveness of denosumab in the three different patient groups; breast, prostate, and other solid tumours. Different risks, such as skeletal-related adverse events and mortality, and utility values were assigned to reflect differences between cancer types. The separate analyses led again to separate recommendations. Denosumab was approved for routine use for adults with bone metastases from breast cancer and other solid tumours, but not from prostate cancer.

**Issue 3: Challenges of developing a counterfactual**

Company submissions supporting histology indications will frequently, if not always, present data collected as part of single-arm studies. The lack of a direct comparator creates challenges as estimates of comparative effectiveness are essential to perform robust cost-effectiveness assessments. A previous review comparing the results from single-arm studies and randomised designs, led the authors to conclude that single-arm studies can only be considered to provide reliable indicators of treatment benefit when the disease natural history is very well known, the patient population is homogenous, and the control (standard care) treatment has little impact on outcomes.190 Current guidance on the selection of a counterfactual and methods to deal with possible biases has also been reported to be limited.220

Hatswell *et al*.220 previously performed a review and developed a taxonomy of approaches used in economic modelling for drugs, which were previously licensed by FDA or EMA without RCT data. The most commonly identified approach used a historical control, though there was variation in the sources of comparison data (i.e. single trial, meta-analysis of multiple trials, registry data or expert opinion). Importantly, the review highlighted that most submissions did not try to control for differences between trials, thus performing a “naïve” comparison.

Naïve comparisons are prone to bias in the presence of systematic differences between patients across clinical studies. Several approaches have been proposed to control for observable (and unobservable) differences between non-randomised comparisons by balancing baseline covariates or matching patients. These methods are outlined in a series of NICE Decision Support Unit (DSU) Technical Support Documents (TSDs)221,158 and a related report.222 TSD17221 provides practical guidance on methods used to analyse treatment effect data from non-randomised studies, including an algorithm for method selection. The methods reviewed are separated according to the assumption of selection on observables (such as regression adjustment, propensity score matching), or selection on unobservables (instrumental variable and panel data methods). Natural Experiment designs are also considered, utilising difference in differences and regression discontinuity approaches. A subsequent DSU report builds on TSD17 by assessing current guidance by NICE on the use of Real World Data (RWD),222 another situation where the analyses are particularly prone to selection bias. Finally, TSD18158 considers the use of novel methodologies for improving indirect comparisons whilst controlling for imbalances in baseline characteristics across different studies, namely “Matching-adjusted Indirect Comparisons” (MAICs), and “Simulated Treatment Comparisons” (STCs).

The approach used to estimate the counterfactual in the 10 NICE TAs included in the review was classified according to the taxonomy developed by Hatswell *et al*.220 This taxonomy distinguishes between the approach taken to developing a comparison group and the source of this data (see Table 12). One appraisal was excluded (TA529) as the submission was based on proxy data from a randomised trial.

Table 34: Classification of approaches taken to construct a counterfactual

|  |  |  |  |
| --- | --- | --- | --- |
| **Intervention** | **NICE TA** | **Classification** | |
| **Approach Taken** | **Source of Comparison Data** |
| Ceritinib | TA395 | Historical control | Clinical trial |
| Osimertinib | TA416 | Historical control | Clinical trial |
| Nivolumab | TA462 | Historical control | Case series |
| Venetoclax | TA487 | Historical control | Clinical trial |
| Atezolizumab | TA492 | Historical control | Meta-analysis |
| Daratumumab | TA510 | Historical control | Mixed sources |
| Avelumab | TA517 | Historical control | Mixed sources |
| Pembrolizumab | TA540 | Historical control | Case series |
| Brigatinib | ID1328 (Ongoing) | Historical control | Clinical trial |

All 9 TAs generated a counterfactual by using a historic control. However, there was variation in the source of comparison data.

*Single Clinical Trial*

The 4 TAs generating a counterfactual using data from a single external trial arm are TA395 (ceritinib for ALK+ NSCLC), TA416 (osimertinib for EGFR T790M+ NSCLC), TA487 (venetoclax for chronic lymphocytic leukaemia) and ID1328 (brigatinib for ALK+ NSCLC). Each of these TAs only considered a single source of external evidence, or, when multiple sources were available, no attempt was made to pool the data and instead indirect comparisons were conducted using one source at a time.

In these appraisals, the committee expressed concerns that the single-arm design of the trials made it difficult to assess the efficacy of the new treatment due to the lack of a comparator arm and that these difficulties were compounded by the small numbers of patients in the trials. Although the use of a historic control from a single external trial was generally accepted as an approach to inform the counterfactual, the committees clearly closely scrutinised the source of external data and the adjustment approaches applied. This was particularly evident when only naïve comparisons were presented, as was the case for the appraisals of ceritinib and venetoclax.

The evidence used in the ceritinib submission was critiqued by the committee due to the lack of an appropriate match between patient characteristics, and limited information about the treatments received by the historical control. This led the committee to conclude that the naïve approach presented by the company was inappropriate. However, in the absence of any suitable alternative estimates or approaches, the committee concluded that the results presented by the company were highly uncertain but represented the best evidence available for their decision-making.

Similarly, in the venetoclax appraisal the committee highlighted the lack of any attempt to match for difference in baseline characteristics and considered the approach to be biased in favour of venetoclax. Again, in the absence of any alternative approaches, the naïve approach was concluded to provide an acceptable basis for decision-making but the results were highly uncertain.

Both the osimertinib and brigatinib appraisals used adjusted comparisons. The osimertinib appraisal used a subgroup of patients with EGFR T790M+ in the control arm of an external prospective, randomised Phase III study and undertook comparative analyses using propensity score matching. Although the committee and the ERG acknowledged the company’s approach to adjusting for possible confounding, concerns remained regarding the immaturity of OS data and the small number of patients.

In the brigatinib appraisal, data for the comparator, ceritinib, came from two separate trials which were assessed separately. The company performed both a naïve comparison and an unanchored MAIC. The committee acknowledged the consistency of the results across both naïve and adjusted analysis. Despite limitations identified relating to the assumptions of the MAIC, the consistency across the different sets of results appeared to provide reassurance to the committee who considered that the comparator evidence was acceptable.

*Case Series*

The TAs supporting the submission with the use of case series data were TA462 (nivolumab for classical Hodgkin lymphoma)215 and TA540 (pembrolizumab for classical Hodgkin lymphoma).223 Both submissions compared their respective product with standard of care data collected from a US database of patients who had been treated with brentuximab vedotin between 2007 and 2015. In both submissions, naïve indirect comparisons and MAICs were performed. The main issue raised by the committee was related to the relevance of the US database to UK practice. For both TAs, the committee acknowledged that the US database might not fully represent UK practice. However, they also deemed it to be the best available evidence while acknowledging the comparative effectiveness results were highly uncertain.

*Meta-analysis*

TA492 (atezlolizumab for PD-L1+ urothelial cancer) derived comparator data from historical trial sources using a range of approaches, including an STC, a MAIC and network meta-analysis.

The initial company submission presented an STC. An STC is a statistical model describing the outcomes in terms of the covariates fitted to the individual patient data for the treatment of interest. This model is used to predict the outcomes that would have been observed in a population with the same characteristics as the historical comparator data source(s). The company then performed a network meta-analysis by linking the outcomes of the various STCs for separate comparators. The ERG highlighted several concerns, particularly the limited number of covariates used in the STC prediction model and the lack of justification for the covariate selection. In response to consultation, the company also provided results from a MAIC to validate the results from the STC.

The committee acknowledged the ERG’s concerns and concluded that the STC analysis was not robust. The committee agreed that the MAIC provided useful validation, but that did not alter its view that the adjustment approaches were not robust. Although the committee acknowledged that atezolizumab was likely to be clinically effective, they had concerns about the magnitude of the effect size given the lack of robust adjustment. Atelzolizumab was subsequently approved for use within the CDF based on the committee’s view that there was plausible potential the treatment was cost-effective and that the key uncertainties surrounding comparative efficacy would be addressed by an ongoing RCT.

*Mixed Sources*

Two TAs, TA510 (daratumumab for multiple myeloma)213, and TA517 (avelumab for MCC),218 considered different approaches and sources. The main submission for daratumumab was based on a MAIC between the daratumumab trials and other comparator trials. However, the ERG and committee expressed concerns about the unreliability of the estimates because of the number of variables which could be controlled for. The company subsequently performed an additional regression analysis of IPD from the pooled daratumumab cohort and the International Myeloma Foundation (IMF) registry. The ERG considered that multivariate regression and MAIC were very different methods, and so it was inappropriate to use the multivariate regression to validate the results of the MAIC. Accordingly, the committee concluded that it was not possible to establish the relative effectiveness of daratumumab due to the high level of uncertainty in the relative effectiveness estimates, issues with the number of variables controlled for in the MAIC, and lack of cross validation of the MAIC with other estimates. Despite these concerns, the committee approved daratumumab for use in the CDF. This was justified by the committee based on the plausible potential that daratumumab could be cost-effective and the view that additional data being collected within the Early Access Programme (EAP) would provide more robust evidence on the clinical effectiveness of daratumumab.

In the avelumab appraisal, the company performed a naïve comparison with a retrospective observational study of patients with metastatic MCC. The company supplemented this with regression analysis but the ERG had concerns due to data immaturity and small patient numbers. Again, problems were mostly around identification of subgroups and variables that might influence the final estimates and lack of suitable head-to-head data.

**Issue 4: Validation of a new test and biomarker**

Three technology appraisals were selected for the case studies in this section: these were of targeted technologies that were “first in class” or positioned at a new point in the treatment pathway where diagnostic testing was not presently commonplace for the relevant genomic alteration. These appraisals contained a discussion of the specific issue of identifying patients with the genomic alteration for which the technology was licensed. We explored how evidence related to the diagnostic accuracy of the test and the predictive and/or prognostic performance of the biomarker was considered.

***Diagnostic accuracy***

To ensure that individuals are able to access targeted treatments, diagnostic tests are required to identify eligible patients. These tests are not specifically appraised during the STA process; however, it is important to consider the diagnostic accuracy of available testing and the appropriateness of proposed testing strategies, as these have implications on the population that is identified and the costs incurred. Implementation of diagnostic strategies with a low sensitivity (high rates of false negative patients) would mean that a proportion of patients are likely to be missed, while strategies with low specificity (high rates of false positive patients) may result in additional resources allocated to unnecessary procedures.

*Crizotinib for untreated ALK-positive NSCLC (TA406)*

In order to identify an ALK-positive patient, it was assumed that the implemented testing strategy would consist of patients first tested with IHC, with positive cases confirmed by FISH. Little detail was provided into the diagnostic accuracy of these types of test to identify ALK mutations; however, it was stated that studies have indicated that IHC is sensitive and specific for determining ALK status and is a viable alternative to FISH. While the validation of a companion diagnostic test is not required in the context of a NICE submission, the company’s IHC test for detecting ALK status did have FDA approval as a companion diagnostic for crizotinib and also received CE marketing for use in Europe.

In order to calculate the cost per ALK-positive patient of testing for ALK status, the company submission described the expected distribution of NSCLC patients according to IHC and FISH tests using data pooled from two sources to estimate the total testing costs. These used assumptions regarding the positivity rate of the IHC test using a specific antibody for ALK-testing, to estimate the number of confirmatory FISH tests that would be required. The company noted that two antibodies were available for use in ALK testing; however, only the antibody considered to be more accurate was used in the analysis.

The ERG provided further detail on the accuracy of the IHC test in comparison to the FISH test. It was acknowledged that there was a possibility that some patients would be incorrectly treated with crizotinib if this testing strategy is adopted, but the exact number was unknown. The ERG considered that the proposed test strategy of IHC followed by confirmatory FISH to be reasonable, in context of their diagnostic accuracy.

*Osimertinib for T790M NSCLC (TA416)*

The testing strategy in the analysis of osimertinib consisted of either tissue biopsy, or circulating tumour DNA (ctDNA) followed by biopsy in those who were negative for the T790M mutation.

The clinical effectiveness of osimertinib was evaluated in the AURA clinical trial programme. Patients were screened using a tissue biopsy and centrally assessed. The tissue biopsy was shown to have a high accuracy rate, with the large majority of patients who were screened as eligible for osimertinib being confirmed as T790M-positive (three patients were later found to not have the T790M mutation and one was of unknown status with insufficient tissue to perform the test).

The sensitivity and specificity of tissue biopsy and ctDNA was described, within the context of estimating the expected testing costs to identify a patient with the T790M mutation. The sensitivity and specificity of tissue biopsy was obtained from a single study, and on unpublished results for ctDNA. The company estimated the overall positive detection rate of 60.1% (i.e. for every 1.66 patients tested, one patient is identified as T790M mutation positive and eligible for osimertinib treatment). Due to limitations of data, the company made assumptions regarding the diagnostic accuracy, assuming that it would be equal in patients who would be eligible for osimertinib as a second-line and as a third-line treatment. The impact on varying the diagnostic accuracy of these tests on the cost-effectiveness of osimertinib was not explored.

*Crizotinib for ROS1-positive NSCLC (TA529)*

The primary testing strategy for identification of patients with the ROS1 oncogene was expected to consist of IHC screening followed by confirmatory FISH. However, in the pivotal clinical trial for crizotinib in ROS1-positive NSCLC, the majority of patients were identified at the screening stage through either central or local testing using FISH, and a small number using the PCR technique. Retrospective testing using NGS showed that two patients in the trial were actually ROS1-negative.

The cost of identifying ROS1-positive patients in the economic analysis consisted of IHC followed up confirmatory FISH. The number of confirmatory FISH tests required was based on the expected sensitivity and specificity of IHC. The company cited an 83% specificity and 100% sensitivity for IHC, as suggested by a validation study for the use of ROS1 IHC staining in screening for ROS1 translocations in lung cancer. FISH was assumed to have sensitivity and specificity of 100%, as FISH was the reference test in the diagnostic accuracy study providing the specificity of IHC in ROS1 testing.

**Predictive validity**

Predictive biomarkers provide an estimate of the expected response to treatment, and these are often targets for treatment. If a biomarker is not predictive, then the targeted treatment will not work in patients without the biomarker present. It is important to understand the predictive nature of the biomarker, it will be difficult, without evidence to support this, to estimate and adjust for the degree of error in estimates of relative effectiveness.

*Crizotinib for untreated ALK-positive NSCLC (TA406)*

ALK was identified as a key oncogenic driver in a number of cancers, including NSCLC. The role of the ALK oncogene in cancer development and the clinical basis for the underlying mechanism of crizotinib in relation to the ALK oncogene is described in the company submission. Crizotinib is an inhibitor of ALK, and is alleged to block the activity of the abnormal ALK protein, which slows the growth and spread of the cancer in ALK-positive NSCLC, and may cause the cancer to shrink.

Since there was no evidence presented for the impact of crizotinib in ALK-negative patients, it was not possible to compare outcomes to those who are ALK-positive, and as such the predictive validity of the ALK oncogene cannot be commented on in this respect

*Osimertinib for T790M NSCLC (TA416)*

The submission provided a description for the clinical basis of the predictive validity of the T790M mutation. Osimertinib was positioned as a second-line option for those who have failed an EGFR tyrosine kinase inhibitor (TKI) for NSCLC. EGFR mutation status had been established as a key predictive biomarker in NSCLC, correlating with sensitivity to an EGFR TKI. However, patients subsequently develop resistance to therapy, which can be either due to secondary mutations or via activation of bypass signalling pathways. The T790M mutations account for 50–60% of all cases of acquired resistance, and secondary T790M mutations were believed to provide a resistance to EGFR TKIs by two potential mechanisms.

As with crizotinib for ALK-positive NSCLC, no mutation-negative patients received treatment with osimertinib, so it was not possible to compare outcomes to those who are mutation positive, and subsequently evaluate empirical evidence of the predictive validity of the T790M mutation.

*Crizotinib for ROS1-positive NSCLC (TA529)*

Since there was no comparative evidence for patients in a ROS1-positive population, the company assumed equivalent efficacy of crizotinib as was observed in ALK-positive patients. The similarities between these two groups of patients was recognised by the EMA, who considered the generalisability of data from ALK-positive patients to the ROS1-positive patients to be sufficient in their approval of crizotinib. The rationale for the similarities was described both in terms of their biological basis and similarities in observed clinical behaviour such as response to crizotinib, patient characteristics (such as age and smoking status) and histology.

The Appraisal Committee considered that relative effectiveness remained uncertain, but agreed to explore the proxy data in its decision-making. However, it regarded this approach as very unusual and stated that this should not set a precedent for the use of data from proxy populations in future appraisals. The lack of knowledge on the implications of the ROS1 oncogene was a factor in the decision to recommend crizotinib through the CDF, in order to collect data about its use in ROS1-positive advanced NSCLC. The committee concluded that collecting data on disease progression in people with ROS1-positive NSCLC treated with crizotinib would help to address the uncertainties around the survival benefit and the comparability of ROS1-positive and ALK-positive NSCLC populations.

Retrospective testing using NGS showed that two patients enrolled in the pivotal trial crizotinib were actually ROS1-negative. While the ROS1-negative patients were included in the ITT analysis of the trial data, the company presented a scenario where these patients were excluded from the analyses. These two patients were described as having a worse or comparable response compared to most other ROS1-positive patients, suggestive of a predictive impact associated with the ROS1 oncogene.

**Prognostic validity**

Prognostic biomarkers indicate the likelihood that a patient will have a particular disease course or natural history independent of treatment, such as the risk of disease progression or the mean survival time. Those with the biomarker present might be expected to experience a different course of disease to someone without the biomarker. The prognostic implications of the target mutation is important when considering the outcomes of patients receiving standard care. This is especially the case since trials of new targeted therapies often do not contain a control arm. In many cases where the relevance of the biomarker is a new discovery, evidence for the natural history of patients with the target mutation will be limited. When only the clinical outcomes of the targeted therapy are known, estimating the relative clinical effectiveness, and subsequently the cost-effectiveness, will be challenging.

*Crizotinib for untreated ALK-positive NSCLC (TA406)*

Evidence on the prognosis of ALK-positive patients on standard chemotherapy was limited at the time of the appraisal, with research into ALK-positive patients only having been studied in the context of investigations of crizotinib.

The life expectancy for patients with ALK-positive NSCLC on standard care could not be established with any certainty. Four estimates of median OS for chemotherapy were presented, but the applicability of these trials to the decision problem was limited. Two trials were identified that enrolled a population that was not specifically ALK-positive, that is, it is possible that both ALK-positive and ALK-negative patients were enrolled, and a further trial of ALK-positive patients of which the majority were not a first-line population having received previous treatments for advanced disease.

The prognosis of ALK-positive patients was compared to that of general NSCLC patients, but it was considered that differences in survival could be due to differences in the patient populations. It had been established that ALK-positive patients tended to be younger with a median age in the early 50s for ALK-positive patients as opposed to mid–late 60s for ALK-negative NSCLC, and are more likely to be non-smokers.

However, no information was presented regarding the disease burden of ALK-positive patients relative to ALK-negative NSCLC patients, and so it was not possible to draw any conclusions as to the impact of ALK status to the prognosis of these patients.

*Osimertinib for T790M NSCLC (TA416)*

The role of the T790M mutation in patient prognosis was not understood at the time of this appraisal, and the discussion regarding a plausible biological basis for any differences between groups of patients was not presented. However, the company presented a number of analyses comparing outcomes for T790M+ve and T790M-ve patients, to demonstrate empirically the extent to which this biomarker may influence prognosis.

The company presented results of a subgroup analysis by presence of T790M status for patients receiving chemotherapy. However, it was acknowledged that the trial used in the example was not designed to explore differences between T790M+ve and T790M-ve patients, and the patients were identified retrospectively as having the EGFR T790M mutation, and so the conclusions that could be drawn from this analysis were limited.

The median TTP for patients on untargeted chemotherapy was demonstrated as being similar between T790M mutation positive and T790M mutation negative patients, although there was some limited evidence to show that there may be some long-term differences, with a Kaplan Meier plot for OS illustrating some divergence between the two groups after 12 months, with the T790M mutation-positive group having marginally poorer survival.

Clinical advice given to the ERG, however, contradicted this evidence of poorer prognosis and suggested that patients with EGFR mutation-positive NSCLC have a better prognosis than patients in an unselected advanced NSCLC population. This is because they tend to be younger and have fewer co-morbidities. This difference in opinion demonstrates that the role of T790M mutations in the prognosis of NSCLC was yet to be established.

*Crizotinib for ROS1-positive NSCLC (TA529)*

The ROS1 oncogene was a relatively new discovery at the time of the appraisal, and ROS1-positive advanced NSCLC is an ultra-orphan indication. As such, little was known about the natural history, patient characteristics and the clinical effectiveness of untargeted chemotherapy for tumours that are ROS1-positive.

As such, the majority of discussion regarding the prognostic validity of the ROS1 mutation was limited to its biological basis, with very limited clinical evidence yet available to demonstrate any differences between ROS1-positive and ROS1-negative groups empirically.

At the time of the appraisal, differences between the characteristics of ROS1-positive patients and those with unselected NSCLC had been established to only a limited degree, with ROS1-positivity showing some associations with non-smoker status and a younger age of diagnosis, both of which are established prognostic factors. NSCLC associated with an underlying ROS1 gene-rearrangement is fundamentally different from unselected NSCLC, as disease progression in ROS1-positive NSCLC patients is dependent on the activated ROS1 receptor tyrosine kinase (RTK) protein.

A systematic review conducted by the company found that the limited studies that reported long-term outcomes for ROS1 patients on chemotherapy were based on very small patient numbers and were not considered to provide reliable estimates of OS. As a result, the prognosis of ROS1 patients on chemotherapy was assumed to be equivalent to that of ALK-positive patients, and data from patients with ALK-positive NSCLC were used as a proxy for the life expectancy of ROS1-positive NSCLC patients treated with current standard of care. The similarities between ROS1-positive and ALK-positive NSCLC allowed for the use of the better quality data available in the latter indication. Evidence in the ALK-positive population was more established, with a large Phase III trial of previously-treated patients and two previous NICE appraisals in this indication, and there was greater clinician experience. Clinical experts predicted that ROS1-positive advanced NSCLC patients will be comparable to overall ALK-positive patients, due to the similar patient characteristics and homology (discussed previously in Section 6.4.2). Similarly to ALK-positive NSCLC, ROS1-positive NSCLC was not considered to be a favourable prognostic factor.

As a result of the uncertainty regarding the comparability of the ROS1 and ALK populations, the most plausible ICERs were considered highly uncertain and crizotinib was recommended for use in the CDF for this indication. This enabled evidence to be collected on patient characteristics and natural history, in order to further understand the ROS1 population and similarities to the ALK-positive population.

**Issue 5: Implementation challenges of incorporating a new diagnostic approach/pathway**

To ensure that individuals are able to access targeted treatments, such as those that are histology-independent, the infrastructure to identify such patients is needed. The introduction or alteration of such infrastructure is associated with a number of challenges. Capacity constraints have been identified as a key barrier to the introduction of precision medicines onto the NHS.224 An increase in service provision may result in investment in NHS genomics services, to increase staffing capacity, laboratory infrastructure, and a need for education and training to ensure that clinicians are aware of where targeted medicines could fit within the treatment pathway. Not only will the requirement of diagnostic tests for patient identification result in additional costs to the NHS, the manner in which patients are identified may also have implications on the type of patients that receive treatment and how similar they are to the patients enrolled in the trials. There may be a variety of testing strategies that could be used in clinical practice, including the diagnostic tests that are used and in which sequence they are used. The time at which patients are identified, whether tested at diagnosis or after treatment failure may influence which is the relevant comparator treatment, which differs by treatment line. In this section, we discuss the extent to which these issues are explored in a number of technology appraisals of targeted therapies.

**Crizotinib for ALK-positive NSCLC (TA406)**

Crizotinib for ALK-positive NSCLC was evaluated initially as a second-line therapy (TA296), with a first-line indication evaluated subsequently in this appraisal. At the time of the appraisal, infrastructure was already in place for the service provision and management of molecular testing to confirm ALK status, with several providers set up with this testing facility. A number of issues regarding the implementation of ALK testing were discussed, including the testing strategy, the timing of testing, the unit costs of testing, and the impact of testing to the number of eligible patients who receive treatment.

*Testing strategy*

The company provided details of a two-tiered testing approach to identify ALK-positive in clinical practice: “testing is performed initially with IHC and positive results are then validated by FISH”, a strategy that was endorsed by two professional bodies (the European Society for Medical Oncology [ESMO] and the Royal College of Pathologists [RCP]) and also implemented in the economic analysis. No specific tests are detailed in the Summary of Product Characteristics (SmPC) for crizotinib, and so the company provided a description of the specific IHC and FISH assays that are endorsed and validated by other clinical bodies such as ESMO, RCP and FDA. This approach to diagnosis appears to differ to the strategy that was used in the pivotal trial of crizotinib, where the identification of ALK patients was based on a FISH test only. However, limited discussion was given as to the implications of the differing testing strategies regarding the patient population identified, although the ERG noted the potential for a two-tiered approach resulting in delays to treatment and patients having a reduced capacity to benefit from treatment if the disease is allowed to progress.

The clinical and cost-effectiveness implications of testing strategies using alternative diagnostic tests, such as NGS or reverse transcription polymerase chain reaction (RT-PCR), were not explored by the company in this appraisal. The company justified their approach by stating that IHC and FISH represent the significant majority of tests used in the NHS, and provided supporting information on the number of IHC tests used in practice. However, the ERG noted that the possibility of using NGS would make the cost of ALK testing less predictable in the near future.

*Timing of testing*

In this appraisal, crizotinib was evaluated as a first-line treatment for NSCLC. At the time, there existed first-line treatments for NSCLC that targeted the EGFR mutation. The company assumed that testing would be done upfront at diagnosis, alongside EGFR testing, based on feedback from an advisory board. Upfront testing alongside EGFR tests means no significant increase in the number of tests required and no potential capacity issues. Sequential testing of ALK status, i.e. after EGFR testing, was not acknowledged as an option.

*Unit costs of testing*

At the time of the appraisal, there was some uncertainty regarding the unit costs of testing, as it is unclear whether laboratory and overhead costs were included in the cost supplied by the company. The impact to the cost-effectiveness of crizotinib by using alternative unit costs of treatment that were estimated by other sources was explored, with a higher cost of testing being associated with a modest increase to the ICER. However, the committee considered that the true cost remained uncertain, and that it was likely to lie between the ranges identified.

*Impact on the number of patients identified*

The challenges of a new diagnostic process were described as having an impact on the number of patients expected to be eligible for crizotinib treatment, noting that the number who received the treatment while it was available on the CDF for a later line of treatment was lower than the expected number of eligible patients, since not all ALK-positive patients were being identified in practice.

**Osimertinib**

Osimertinib, appraised for NSCLC patients with a T790M mutation (TA416), was positioned as a second-line treatment option following treatment with an EGFR TKI, given the low prevalence of T790M mutations at diagnosis. The challenges in the diagnostic pathway with a second-line therapy was discussed, including the increase in service provision and the testing strategy required.

*Increase in service provision*

The identification of patients eligible for osimertinib was discussed as the main additional resource use to the NHS. The appraisal discusses how not all centres routinely test for the EGFR T790M mutation either at diagnosis or after treatment failure with a first-line EGFR-TKI, and its introduction will therefore necessitate a change in service provision.

The expansion of testing was not considered to be problematic, as the pathway for acquisition, handling and testing of tissue, in addition to mechanisms for reporting of results, was described as being well-established, and so no additional costs were associated with assessment of tumour specimens beyond the increase in testing volumes. Details of the laboratories enrolled to conduct EGFR testing and their current ability to detect T790M mutations using existing platforms were provided to support this assumption.

However, tissue biopsy at disease progression following resistance to EGFR TKI therapy is not routine, and the company provided a detailed description of how the change of pathway to acquire tumour specimens would be implemented. There were a number of challenges highlighted, including the optimal selection of lesions for biopsy due to tumour heterogeneity, and reduced willingness to undergo tissue biopsy. Feasibility studies to validate the pre-analytical steps of the plasma processing pathway were expected to commence shortly after the appraisal.

*Testing strategies*

Four possible testing strategies to detect T790M mutations were described: (i) tissue biopsy; (ii) ctDNA (plasma) test followed by tissue biopsy in patients identified as T790M negative by ctDNA; (iii) ctDNA alone, and (iv) tissue biopsy followed by ctDNA. The company considered that only the first two testing strategies were relevant and in line with the SmPC for osimertinib, and included a weighted average of these strategies in their base case analysis based on the proportion expected to be identified in each way. A number of clinical benefits with the use of ctDNA were described, with it being a less expensive alternative and offering more rapid results, and mitigating the complications associated with the acquisition of lung tissue samples which may be of particular concern for later-stage disease.

**Crizotinib for ROS1-ve NSCLC**

The appraisal of crizotinib for treating ROS1-positive advanced NSCLC (TA529) also highlighted some uncertainty with the introduction of diagnosis into the patient pathway. Diagnostic testing was not routinely undertaken in England and Wales to identify ROS1-positive patients at the time of the appraisal; however, there were pre-existing targeted treatments available for NSCLC, where patients were tested for the associated biomarker for EGFR and ALK, at diagnosis of NSCLC. The discussions around the challenges of identifying ROS1-positive patients were centred on the point in the pathway where ROS1 would be detected, and the implications of testing to the time of crizotinib treatment.

*Timing of testing*

The company presented different scenarios to illustrate the impact of introducing testing at different points in the treatment pathway: one where testing could either be done upfront upon diagnosis alongside testing for other targets associated with treatment for NSCLC (EGFR and ALK), or where it would be done sequentially after confirmed EGFR-negativity and ALK-negativity. Upfront testing minimises tissue wastage and avoids delays in access to therapy by waiting for the patient to complete testing for the targets with existing therapies.

Other parties, including NHS England and the committee also considered that upfront testing was more appropriate than sequential testing. The ERG also considered the impact of the timing of testing on the cost that it incurs: there may be a discount available for upfront testing when testing for more than one mutation at the same time, and that patients treated in the subsequent-line would already have been tested for ALK and/or other mutations, so the cost of testing these (ALK-positive) patients need not be taken into account.

*Implications of testing to the timing of treatment*

The issue was also raised as to the positioning of crizotinib in the pathway. While the economic analysis considered crizotinib against a comparator that is commonly used as first-line in NSCLC, it was expected that patients treated by crizotinib in clinical practice may be either treatment-naïve or treatment-experienced. If access to diagnostic testing causes delays in diagnosis of ROS1-positivity, or ROS1-testing had not been performed prior to initiating first-line therapy, patients would be treatment-experienced upon starting crizotinib; however, over time it was expected for patients to become predominately treatment-naïve as testing becomes more established and diagnosis occurs at an earlier stage in the treatment pathway.

Appendix 9: OpenBUGS code

Bayesian Hierarchical Model: Uniform(0,5) prior distribution for the between-tumour standard deviation

# CODE ADAPTED FROM: Thall et al (2003)

# Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes.

# Statist. Med., 22: 763-780. doi:10.1002/sim.1399

#

# Uniform prior distribution for between-group SD, as recommended by Cunanan et al. (Clinical Trials, 2019)

#

model{

for (i in 1:numGroups){ # numGroups is k, the number of different probabilities

x[i] ~ dbin(p[i],n[i]) # In each group, x is the number of responses and n is the number of patients

# set up deviance code with correction for zero cells

x1[i] <- max(x[i],0.1) # zero cell correction

xhat[i] <- p[i] \* n[i] # expected value of the numerators

xhat1[i] <- max(xhat[i], 0.1) # zero cell correction

# Deviance contribution with zero cell correction

dev1[i] <- 2 \* (x1[i] \* (log(x1[i])-log(xhat1[i]))

+ (n[i]-x1[i]) \* (log(n[i]-x1[i]) - log(n[i]-xhat1[i])))

# deviance contribution for for zero cells

dev0[i] <- 2 \* n[i] \* log(n[i]/(n[i]-xhat[i]))

# deviance contribution

dev[i] <- dev1[i] \* (1-equals(x[i],0)) + dev0[i] \* equals(x[i],0)

# logit model for p

logit(p[i]) <- rho[i]

rho[i] ~ dnorm(mu,tau) # RE for log-odds

# Probability that the response rate for each group is > than targetResp (given as data)

pg[i] <- step(p[i] - targetResp)

pg2[i] <- step(p[i] - targetResp2)

}

totresdev <- sum(dev[]) # total residual deviance

# Priors

mu ~ dnorm(mean.Mu, perc.Mu) # pooled mean of log-odds

#tau ~ dgamma(tau.alpha, tau.beta) # used in Thall (2003)

#sd <- 1/sqrt(tau) # between-group sd (log-odds scale)

sd ~ dunif(0,5) # recommended by Cunanan (2019)

tau <- pow(sd,-2)

# predictive distribution

rho.new ~ dnorm(mu,tau) # log-odds response across groups

# convert to probabilities

logit(p.pooled) <- mu # mean probability of response across groups

logit(p.new) <- rho.new # probability response across groups

# predictive probabilities of response rates > targetResp (given as data)

pg.new <- step(p.new - targetResp)

pg2.new <- step(p.new - targetResp2)

}

**Data**

list(x=c(2,2,9,2,6,5,1,6,0,0,0,0), n=c(11,12,7,5,4,4,4,3,2,1,1,1), numGroups=13, mean.Mu=-0.847298, perc.Mu=0.1, targetResp=0.3, targetResp2=0.1)

Appendix 10: Model Fit Statistics and Residual Deviance Base Case

For all analyses, 55,000 iterations were run on 2 parallel chains and the first 5,000 iterations discarded as “burn-in”. Convergence was assessed by visual inspection of the Brooks-Gelman-Rubin plots and assessment of the  statistic.152, 153

The model fit statistics for the base-case and the sensitivity analyses can be seen in Table 35. The results show that all models fit the data well. Inspection of box-plots of individual groups’ contributions to the residual deviance support this.

Table 35: Model fit statistics for the base-case and sensitivity analyses

|  |  |  |
| --- | --- | --- |
|  | Posterior mean of the residual deviance\* | DIC |
| Base-case: Uniform (0,5), 0.3 mean response probability | 11.6 | 30.7 |
| Uniform (0, 5), 0.5 mean response probability | 11.9 | 30.9 |
| Inverse Gamma (2, 20) | 10.4 | 28.9 |
| Half-normal (0,0.01)T(0,) | 10.9 | 30.5 |
| Half-normal (0,0.1)T(0,) | 12.1 | 31.4 |
| Half-normal (0,0.5)T(0,) | 14.8 | 33.1 |
| \*Compare to 12 groups | | |

Appendix 11: Bayesian Hierarchical Model Sensitivity Analyses

The sensitivity of the BHM results to the prior distribution were assessed. The results showed the BHM estimated substantial heterogeneity between tumour types irrespective of the prior distribution or the response probability. This can be seen in the estimates of the posterior distributions for the between-group heterogeneity standard deviations in Table 36. The 95% credible intervals around all of the results are wide indicating considerable uncertainty in these estimates.

Table 36: Sensitivity analyses of the Bayesian hierarchical model to alternative prior distributions

|  |  |
| --- | --- |
| **Prior distribution** | **Posterior distributions for the between-group heterogeneity standard deviations [95% CrI]** |
| Base-case: Uniform (0,5), 0.3 mean response probability | 2.863 [0.922; 4.826] |
| Uniform (0, 5), 0.5 mean response probability | 2.828 [0.865; 4.828] |
| Inverse Gamma (2, 20) | 3.273 [1.879; 5.901] |
| Half-normal (0,0.01)T(0,) | 3.738 [0.970; 9.544] |
| Half-normal (0,0.1)T(0,) | 2.740 [0.812; 5.814] |
| Half-normal (0,0.5)T(0,) | 1.820 [0.455; 3.466] |

Appendix 12: Estimating annual eligible population

**Diagnostic Accuracy Illustration**

The diagnostic accuracy of a test will vary depending on the prevalence of the genetic alteration within each tumour type even when the sensitivity and specificity are held constant. This can be expressed by looking at the positive predictive value (PPV) and negative predictive value (NPV) of a test. PPV is defined as the likelihood that an individual with a positive test truly has the condition. Alternatively, the NPV is the likelihood that the individual with a negative test truly does not have the condition. The predictive value of a test will differ depending on the prevalence of a genetic alteration.

For example, the sensitivity and specificity of an IHC test for detecting NTRK fusions is 88% and 81%, respectively. If IHC was used to detect NTRK fusions in 2000 patients, 1000 with gastrointestinal stromal tumour (NTRK fusion prevalence = 0.1%) and 1000 with papillary thyroid tumour (NTRK fusion prevalence = 13.30%), the PPV and NPV of a test will differ. Table 37 demonstrates how the prevalence of NTRK fusion changes the PPV and NPV of a test.

Table 37: Illustrative example of how the diagnostic accuracy of IHC (sensitivity = 88%, specificity = 81%) is different for two tumour types with differing NTRK fusion prevalence.

|  |  |  |
| --- | --- | --- |
|  | Papillary Thyroid Cancer | Gastrointestinal Stromal Tumour |
| NTRK fusion prevalence | 13% | 1% |
| Total Population with NTRK fusion | 130 | 10 |
| True Positives (Test Positive, NTRK Positive) | 114 | 9 |
| False Positives (Test Positive, NTRK Negative) | 165 | 1 |
| True Negatives (Test Negative, NTRK Negative) | 705 | 871 |
| False Negatives (Test Negative, NTRK Positive) | 165 | 119 |
| Positive Predictive Value | 88% | 99% |
| Negative Predictive Value | 7% | 98.9% |

**Estimation of eligible population**

Table 38 details the sources used to estimate the annual eligible population.

The prevalence of NTRK fusions for each tumour type were determined from a pragmatic review of the literature. For the majority of tumour types, the prevalence of NTRK fusions were acquired from Foundation Medicine Data in the FDA review for larotrectinib, which assessed 34,476 tumour samples.173 Frequency of NTRK fusions of the remaining tumour types were taken from published evidence. 171, 172, 174 Even though NTRK fusions have been identified in renal cell carcinoma, 225 ovarian225 and prostate cancer226, the exact frequencies are not recorded. Therefore, it was assumed that the NTRK fusion frequency in these tumours was the same as the average frequency of NTRK fusions across all tumours, which is estimated to be 0.26%. 176

The annual incidence of tumours in an anatomical site (e.g. pancreatic cancer) in England were obtained from the Cancer Registration Statistics and Rare and Less Common Cancer databases. 227, 228 Where the NTRKfusion has been reported in a specific tumour type (e.g. pancreatic adenocarcinoma) rather than an anatomical site, we used an estimate of the proportion of patients with that cancer type, based on published evidence. Incidence estimates for NSCLC neuroendocrine tumours, soft tissue sarcoma were not available and were obtained from other published sources.229-231 As there was no crude incidence available for appendiceal adenocarcinoma, the annual incidence was estimated using an incidence per 100,000 and the annual population within the UK in 2017.232, 233

The proportion of individuals diagnosed with stage III/IV cancer was used as a proxy measure for the proportion of the patients with locally advanced or metastatic cancer. Values for this were primarily obtained from Cancer Research UK. For the tumour types where data was not available, values were taken from published sources.225, 231, 233-239 For the tumour types in which a known proportion of the patient population had an unknown stage at diagnosis, the unidentified proportion of stage III/IV cancer at diagnosis was assumed to follow the same distribution as the known proportion.

Table 39 presents the calculations of the annual eligible population for each testing strategy.

Table 38: Literature Sources of NTRK fusion prevalence, annual cancer incidence and proportion of patients diagnosed with stage 3/4 cancer in the tumour known to harbour an NTRK fusion

|  |  |  |  |
| --- | --- | --- | --- |
| **Tumour Type** | **NTRK Fusion Prevalence** | **Annual Cancer Incidence (England)** | **Proportion Stage 3/4 at diagnosis** |
| **Represented Tumour Types** | | | |
| Appendix Cancer | Amatu, Sartore-Bianchi et al. 2016174 | Based on incidence of 0.97/100,000233 and the population in England in 2017232 | Marmor, Portschy et al. 2015233 |
| Breast cancer (NOS) | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227 | Cancer Research UK: Breast cancer incidence by stage at diagnosis240 |
| Cholangiocarcinoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Rare and Less Common Cancers Statistics228 | Tsuchiya, Sawada et al. 2015. (Assumed to be the same as Heptatocellular Carinoma)234 |
| Colorectal cancer | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227, 241 | Cancer Research UK: Bowel cancer incidence by stage at diagnosis242 |
| GIST | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Amelio, Ruzafa et al. 2014 243 | PDQ Adult Treatment Editorial |
| Infantile Fibrosarcoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Based on incidence of 0.5/100,000173and the population in England in 2017232 | Orbach, Rey et al. 2009236 |
| MASC | Skálová, Vanecek et al. 2010244 | Cancer Registration Statistics227 and Luk, Selinger et al. 2015 245 | Sethi, Kozin et al. 2014235 |
| Melanoma | Okamura, Boichard et al. 2018172 | Cancer Registration Statistics227 | Cancer Research UK: Melanoma skin cancer incidence by stage at diagnosis246 |
| NSCLC (NOS) | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | National Lung Cancer Audit231 | National Lung Cancer Audit231 |
| Pancreatic Cancer | Okamura, Boichard et al. 2018172 | Cancer Registration Statistics227, 247 | Cancer Research (UK): Pancreatic cancer incidence by stage at diagnosis248 |
| Soft tissue sarcoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Research UK: Soft Tissue Sarcoma Incidence Statistics249 | American Cancer Society 2017250 |
| Thyroid | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227 | Deen, Burke et al. 2016238 |
| **Unrepresented Tumour Types** | | | |
| Cervical cancer | Okamura, Boichard et al. 2018172 | Cancer Registration Statistics227 | Cancer Reasearch UK: Cervical cancer incidence by stage at diagnosis251 |
| Congenital mesoblastic nephroma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Rare and Less Common Cancers Statistics228 and Gooskens, Houwing et al. 2017237 | Gooskens, Houwing et al. 2017237 |
| Gastro-oesphoegeal junction adenocarcinoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | National Institute for Health and Care Excellence (NICE) 252 | Cancer Research UK: Oesophageal cancer incidence by stage at diagnosis253 |
| Head and neck squamous cell carcinoma (NOS) | Okamura, Boichard et al. 2018172 | Cancer Registration Statistics227 | Cancer Research UK: Head and Neck cancer incidence by stage at diagnosis254 |
| High grade glioma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Public Health England255 | All high-grade glioma cancers are advanced or metastatic225 |
| Neuroendocrine tumour | Sigal, Bhangoo et al. 2018171 | UK and Ireland Neuroendocrine Tumour Society229 | UK and Ireland Neuroendocrine Tumour Society229 |
| Ovarian cancer (NOS) | Assumption based on average prevelence (Solomon, Linkov et al. 2019)176 | Cancer Registration Statistics227 | Cancer Research (UK): Ovarian cancer incidence by stage at diagnosis256 |
| Paediatric high grade glioma | Okamura, Boichard et al. 2018172 | Farrimond 2010257 | Lee, Kim et al 2019225 |
| Paediatric melanoma | Okamura, Boichard et al. 2018172 | Cancer Research UK258 | Austin, Xing et al. 2013239 |
| Papillary thyroid tumour | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics (9) and Brzeziańska, Karbownik et al. 2006 259 | Deen, Burke et al. 2016238 |
| Prostate cancer (NOS) | Assumption based on average prevelence (Solomon, Linkov et al. 2019)176 | Cancer Registration Statistics227 | Cancer Research (UK): Prostate cancer incidence by stage at diagnosis260 |
| Renal cell carcinoma | Assumption based on average prevelence (Solomon, Linkov et al. 2019)176 | Cancer Registration Statistics227 and Cancer Research UK261 | Cancer Research (UK): Kidney cancer incidence by stage at diagnosis261 |
| Salivary gland (non MASC) | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227 | Assumed to be the same as Head and Neck Squamous Cell Carcinoma254 |
| Secretory breast carcinoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227 and Horowitz, Sharma et al. 2012262 | Jacob, Hodge et al. 2016263 |
| Sinonasal adenocarcinoma | Assumption based on average prevelence (Solomon, Linkov et al. 2019)176 | Cancer Registration Statistics227 and Rushton 2012264 | Assumed to be the same as Head and Neck Squamous Cell Carcinoma254 |
| Uterine carcinoma | NDA Multidisciplinary Review and Evaluation: Larotrectinib173 | Cancer Registration Statistics227 | Cancer Research UK: Uterine cancer incidence by stage at diagnosis.265 |

1 Prevalence of NTRK fusions in papillary thyroid tumour is based on the average of two estimates provided in the NDA multidisciplinary review

Table 39: calculations of the annual eligible population for three testing strategies based on the prealence of NTRK fusion, the annual incidence of cancer in England and the proportion of patients with advanced or metastatic cancer at diagnosis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tumour Type** | **Prevalence of NTRK fusion** | **Cancer Incidence (England)** | **% with Stage III/IV Cancer** | **Annual TRK-inhibitor eligible population** | | |
| **Hierarchical** | **RNA-based NGS** | **Exhaustive** |
| **Tumours represented in trial** | | | | |  |  |
| **Appendix** | 4.00% | 540 | 74% | 14.04 | 15.97 | 12.95 |
| **Breast** | 0.07% | 46102 | 15% | 4.25 | 4.84 | 3.93 |
| **Cholangiocarcinoma** | 0.10% | 556 | 60% | 0.29 | 0.33 | 0.27 |
| **Colorectal** | 0.12% | 34825 | 55% | 20.20 | 22.98 | 18.64 |
| **GIST** | 1.28% | 734 | 40% | 3.30 | 3.76 | 3.05 |
| **Infantile Fibrosarcoma** | 90.90% | 59 | 51% | 24.04 | 27.35 | 22.18 |
| **MASC** | 92.90% | 11 | 22% | 1.80 | 2.25 | 1.82 |
| **Melanoma** | 0.21% | 13740 | 10% | 2.28 | 2.60 | 2.11 |
| **NSCLC** | 0.09% | 32576 | 57% | 14.69 | 16.71 | 13.55 |
| **Pancreatic** | 0.26% | 8388 | 78% | 14.95 | 17.01 | 13.80 |
| **Soft tissue sarcoma** | 0.56% | 2740 | 32% | 4.32 | 4.91 | 3.98 |
| **Thyroid** | 0.92% | 2195 | 31% | 5.50 | 6.26 | 5.08 |
| **Tumours not represented in trial** | | | | |  |  |
| **Cervix** | 0.33% | 2591 | 24% | 1.80 | 2.05 | 1.66 |
| **Congenital Mesoblastic Nephroma** | 60.70% | 2 | 17% | 0.20 | 0.23 | 0.18 |
| **Gastro-oesophageal junction** | 0.10% | 7569 | 73% | 3.40 | 3.87 | 3.14 |
| **High Grade Glioma** | 0.05% | 2781 | 100% | 1.18 | 1.34 | 1.09 |
| **HNSCC** | 0.38% | 9946 | 63% | 20.93 | 23.81 | 19.31 |
| **Neuroendocrine** | 0.30% | 4363 | 53% | 6.10 | 6.94 | 5.63 |
| **Ovarian** | 0.25% | 2724 | 55% | 3.29 | 3.75 | 3.04 |
| **Paediatric High Grade Glioma** | 5.30% | 67 | 100% | 3.11 | 3.54 | 2.87 |
| **Paediatric melanoma** | 11.11% | 56 | 34% | 1.83 | 2.08 | 1.68 |
| **Papillary thyroid tumour** | 13.30% | 1057 | 31% | 38.30 | 43.57 | 35.34 |
| **Prostate** | 0.25% | 41201 | 43% | 38.93 | 44.29 | 35.92 |
| **Renal cell carcinoma** | 0.25% | 7438 | 43% | 7.19 | 8.18 | 6.64 |
| **Salivary gland** | 1.72% | 517 | 63% | 4.92 | 5.60 | 4.54 |
| **Secretory Breast Carcinoma** | 91.70% | 7 | 9% | 0.46 | 0.58 | 0.47 |
| **Sinonasal adenocarcinoma** | 0.25% | 5 | 63% | 0.01 | 0.01 | 0.01 |
| **Uterine** | 0.10% | 7862 | 18% | 1.24 | 1.42 | 1.15 |

**Modelled testing strategies**

Tables 40-42 present the testing that would be used to identify NTRK fusions for each tumour type across the three testing strategies. The appropriate test is dependent on NTRK fusion frequency and current testing availability.

Table 40: Testing strategy for each tumour type under the hierarchical approach

| **Testing Strategy** | **Costs** | **Tumour Type** | |
| --- | --- | --- | --- |
| FISH | No incremental costs | * MASC | * Secretory breast carcinoma |
| WGS & Confirmatory RNA-based NGS | Cost of confirmatory RNA-based NGS only. | * Congenital Mesoblastic Nephroma * Infantile fibrosarcoma | * Paediatric high grade glioma * Paediatric melanoma * Soft tissue sarcoma |
| IHC + RNA-based NGS | Total cost of IHC and RNA-based NGS | * Appendiceal adenocarcinoma * Breast cancer (NOS) * Cervical cancer (NOS) * Cholangiocarcinoma * Colorectal adenocarcinoma * Gastrointestinal stromal tumour * GEJ adenocarcinoma * HNSCC (NOS) * High grade glioma * Melanoma (NOS) | * Neuroendocrine (NOS) * NSCLC (adenocarcinoma) * Ovarian cancer (NOS) * Pancreatic adenocarcinoma * Papillary thyroid tumour * Prostate cancer (NOS) * Renal cell carcinoma * Salivary gland carcinoma * Sinonasal adenocarcinoma * Thyroid tumour (NOS) * Cancer of Unknown Primary * Uterine carcinoma |

Table 41: Testing strategy for each tumour type under the first-line RNA based NGS approach

| **Testing Strategy** | **Costs** | **Tumour Type** | |
| --- | --- | --- | --- |
| First-line RNA-based NGS | Incremental costs of displacing FISH | * MASC | * Secretory breast carcinoma |
| WGS & Confirmatory RNA-based NGS | Cost of confirmatory RNA-based NGS only. | * Congenital Mesoblastic Nephroma * Infantile fibrosarcoma | * Paediatric high grade glioma * Paediatric melanoma * Soft tissue sarcoma |
| First-line RNA-based NGS | Total cost of RNA-based NGS | * Appendiceal adenocarcinoma * Breast cancer (NOS) * Cervical cancer (NOS) * Cholangiocarcinoma * Colorectal adenocarcinoma * Gastrointestinal stromal tumour * GEJ adenocarcinoma * HNSCC (NOS) * High grade glioma * Melanoma (NOS) | * Neuroendocrine (NOS) * NSCLC (adenocarcinoma) * Ovarian cancer (NOS) * Pancreatic adenocarcinoma * Papillary thyroid tumour * Prostate cancer (NOS) * Renal cell carcinoma * Salivary gland carcinoma * Sinonasal adenocarcinoma * Thyroid tumour (NOS) * Cancer of Unknown Primary * Uterine carcinoma |

Table 42: Testing strategy for each tumour type under the exhaustive approach

| **Testing Strategy** | **Costs** | **Tumour Type** | |
| --- | --- | --- | --- |
| DNA-based NGS & Confirmatory RNA-based NGS | Incremental costs of displacing FISH | * MASC | * Secretory breast carcinoma |
| WGS & Confirmatory RNA-based NGS | Cost of confirmatory RNA-based NGS only. | * Congenital Mesoblastic Nephroma * Infantile fibrosarcoma | * Paediatric high grade glioma * Paediatric melanoma * Soft tissue sarcoma |
| DNA-based NGS & Confirmatory RNA-based NGS | Total cost of DNA-based NGS and RNA-based NGS | * Appendiceal adenocarcinoma * Cervical cancer * Cholangiocarcinoma * Gastrointestinal stromal tumour * GEJ adenocarcinoma * HNSCC (NOS) * High grade glioma | * Neuroendocrine tumour * Pancreatic adenocarcinoma * Prostate cancer (NOS) * Renal cell carcinoma * Salivary gland carcinoma * Sinonasal adenocarcinoma * Cancer of Unknown Primary * Uterine carcinoma |
| DNA-based NGS & Confirmatory RNA-based NGS | Cost of confirmatory RNA-based NGS only | * Breast cancer (NOS) * Colorectal adenocarcinoma * Melanoma (NOS) * NSCLC | * Ovarian cancer (NOS) * Papillary thyroid tumour * Thyroid tumour (NOS) |

Appendix 13 : Case Study: Economic model

***Survival***

The distribution of patients in each health state is determined by using observed PFS and OS. Traditionally, observed time-to-event data for PFS and OS is utilised for both treatment arms and depending on the maturity of the data, direct extrapolation required. However, time-to-event data for PFS and OS were not available in the literature for either of the approved Trk-inhibitors: larotrectinib and entrectinib.

The literature did report median PFS and OS for both larotrectinib and entrectinib, however there were significant differences between the median survival estimates of the two. Median PFS and OS were 28.3 months and 44.4 months for patients in the larotrectinib study, and 11.2 months and 20.9 months for patients in the entrectinib study. 266, 267 Furthermore, the reported OS and PFS were deemed highly uncertainty due to the significant data immaturity and uncertainty about the extent to which OS is driven by the efficacy of subsequent therapies.

Due to these uncertainties, hypothetical estimates of PFS and OS were used in the economic model and can be seen in Table 43. Standard errors were assumed to be 10% of the mean.

Table 43: Survival Estimates

|  |  |  |
| --- | --- | --- |
|  | **Median PFS (months) [95% Confidence Interval]** | **Median OS (months) [95% Confidence Interval]** |
| Responders | 24 [21.6;26.4] | 36 [32.4;39.6] |
| Non Responders | 6 [5.4;6.6] | 12 [10.8;13.2] |

It is assumed the survival function of responders and non-responders follows an exponential distribution. Exponential parametric survival curves were therefore generated based on median OS and PFS values.

The resulting OS and PFS survival curves for responders and non-responders can be seen in Figure 16.

Figure 16: Plots showing the stylised progression free survival (a) and the overall survival (b) curves for responders and non-responders

(a)

(b)

**Utilities**

Stylised health state utilities were used in the economic model and can be seen in Table 44. The utility values used for progression-free disease for Drug X and progressed disease were based on the mean values reported in the NICE technology appraisal of Brigatinib for the treatment of patients with ALK+ advanced NSCLC previously treated with crizotinib.71, 102, 268 Given the cytotoxic nature of chemotherapy and the targeted nature of Drug X, the utility value of SoC was assumed to be lower than that of Drug X. As a result, a utility value of 0.72 was used for SoC. This value was based on the utility reported in the NICE technology appraisal of Crizotinib for treating ROS1 positive advanced NSCLC.219 The progressed disease health state utilities for Drug X and SoC were assumed to be equivalent as active treatment was assumed to be discontinued upon progression.

It was assumed the health state utilities were unchanged across tumour types. To reflect uncertainty in the utilities, standard errors were assumed to be 10% of the mean.

Table 44: Health state utilities

|  |  |  |
| --- | --- | --- |
|  | **Drug X [95% Confidence Interval]** | **SoC [95% Confidence Interval]** |
| Progression-free disease | 0.79 [0.71;0.87] | 0.72 [0.65;0.79] |
| Progressed disease | 0.64 [0.57;0.71] | 0.64 [0.57;0.71] |

**Resource Use and Costs**

In the absence of the acquisition cost of any currently available Trk-inhibitors inclusive of pricing discounts for the NHS, it is assumed the manufacturer of Drug X would employ a value-based approach to pricing. This assumes the drug acquisition cost would be set at a level that results in a histology independent ICER (inclusive of testing costs and weighted according to the eligible population) at approximately NICE’s decision threshold. For the purpose of the case study, Drug X is assumed to meet NICE’s end-of-life criteria, allowing a maximum willingness to pay threshold of £50,000 per QALY.

To ensure generalisability of the results, the preferred approach to generate a weighted average is to use the eligible population rather than the trial population. This should also include the unrepresented tumour types. The threshold analysis used to generate the value-based price of Drug X is conducted using the eligible population including the unrepresented tumour types. The acquisition cost of the SoC is assumed to be £20 per month.

For simplicity it is assumed there are no costs associated with administering Drug X or SoC and that there are no adverse event costs. It is also assumed patients discontinue Drug X and SoC upon disease progression.

Health state costs were assumed to be £350 per month per patient in the progression-free disease state and £500 per month per patient in the progressed disease health state. These values were informed by the health state costs reported in the NICE technology appraisals of brigatinib and crizotinib.216, 268 219

To reflect uncertainty in the health state costs, standard errors were assumed to be 10% of the mean.

A one-off terminal care cost of £6,878 is applied upon transition from the progressed disease state to the death state. The terminal care cost was obtained from Georghiou and Bardsley. 269

The cost parameters used in the economic model can be seen in Table 45.

Table 45: Drug acquisition costs and health state costs

|  |  |
| --- | --- |
|  | **Costs** **[95% Confidence Interval]** |
| **Drug Acquisition Costs** | |
| Drug X | £1250 |
| SoC | £20 |
| **Health State Costs** | |
| PFS | £350 [£315; £385] |
| PPS | £500 [£450; 550] |
| End of Life | £6878 |

To assess the uncertainty surrounding the variables included in the cost-effectiveness model, a probabilistic sensitivity analysis (PSA) was undertaken using 10,000 samples. All results reported are the mean averages of the 10,000 iterations.

**Testing Costs for Unrepresented Tumours**

Table 46: Summary of NNS, Annual Eligible Population and testing costs for the tumours unrepresented in the trial

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Annual Eligible Population** | **Number Needed to Screen** | **Cost to Identify One Patient Eligible for NTRK treatment** |
| Cervix | 2 | 303.0 | £15,152 |
| Congenital Mesoblastic Nephroma | 0 | 2.0 | £102 |
| Gastro-oesophageal junction | 4 | 1000.0 | £50,000 |
| Head and Neck Squamous Cell Carcinoma | 24 | 263.2 | £13,158 |
| High Grade Glioma | 1 | 2000.0 | £100,000 |
| Neuroendocrine | 7 | 333.3 | £16,667 |
| Ovarian | 4 | 400.0 | £20,000 |
| Papillary Thyroid Tumour | 3 | 23.3 | £1,163 |
| Paediatric High Grade Glioma | 2 | 9.0 | £450 |
| Paediatric Melanoma | 44 | 461.4 | £23,070 |
| Prostate cancer (NOS) | 8 | 400.0 | £20,000 |
| Renal cell carcinoma | 6 | 58.1 | £2,907 |
| Salivary gland | 1 | 1.1 | £55 |
| Secretory breast carcinoma | 0 | 400.0 | £20,000 |
| Sinonasal Adenocarcinoma | 44 | 7.5 | £376 |
| Uterine | 1 | 1000.0 | £50,000 |
| Total | **151** |  |  |

This illustrative example assumes that the testing costs equal £50 with a 100% sensitivity and specificity. The average cost to identify one individual eligible for treatment is estimated to be £14,322.