Improving economic evaluation and decision-making for oncology drugs using real-world data

Jiyeon Kang

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

November 2022

Department of Health Service Research and Policy
Faculty of Public Health and Policy
LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Centre for Cancer Biomarkers, Bergen Norway.
CCBIO is established by Research Council of Norway in Collaboration with University of Bergen

Declaration

I, Jiyeon Kang, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed

Jiyeon Kan 

November 2022

Abstract

Health Technology Assessment (HTA) faces challenges such as the absence of direct treatment comparisons to evaluate technologies since health technologies get more complex. There is a growing interest that real-world data (RWD) could help fill evidence gaps in appraising new drugs, particularly cancer drugs. This thesis explored the use of RWD in appraisals of oncology medicines by the National Institute for Health and Care Excellence (NICE) to investigate how RWD have been used, how the use varies by cancer areas and over time, what factors are associated with greater or lesser use of RWD and to what extent RWD can help reduce uncertainty in the economic evaluation of cancer medicines. In this thesis, data were systematically extracted from 229 NICE Single Technology Appraisals of cancer drugs following a specially-developed data extraction protocol. Beyond simple counting, patterns and intensity of use of RWD were identified to review the use of RWD in appraisals. Patterns were then categorised based on the number of uses of RWD in three major components (overall survival (OS), volume of treatment, choice of comparators) for an analysis of the intensity of use of RWD. Seven factors (time, internal/external validity, availability of direct treatment comparison, incidence rate, maturity of the data on OS, previous technology appraisal recommendation by NICE) were identified, which could be possible incentives for greater use of RWD. Regression models were estimated to find the associations between the use of RWD and a set of the factors. An association between the use of RWD in estimating OS and the maturity of OS data was reviewed in depth. By highlighting the challenges, such as uncertain generalisability of a clinical trial and immature survival data, sources of uncertainty in appraisals and patterns of use of RWD were investigated. This was reviewed by comparing appraisals of targeted cancer therapy with appraisals of non-targeted cancer therapy. Lastly, appraisals of the cancer drugs exiting the 2016 Cancer Drugs Fund (CDF) were reviewed to understand how RWD on provision through the CDF helped address uncertainties in the review process. The research showed that NICE has incorporated RWD substantially in diverse parts of the economic evaluation. However, the use of RWD to reduce uncertainty in appraisals appears to be quite limited.

Acknowledgements

I would like to express my deepest gratitude to my supervisor John Cairns for the enormous amount of time and support that he has given me during this PhD. It has been indeed a privilege to learn from him. His endless support and faith guide me through this PhD and nurture me in academic and personal growth. This PhD under his supervision is undoubtedly the most precious experience. I would also like to thank my second supervisor, Alec Miners, who provided me with gentle support and warmth. I especially thank him for his unwavering support even after he left the school.

I am sincerely grateful for the generous funding from the Centre for Biomarkers. This thesis would not have been possible without their continued support. I especially thank Roger Strand for always taking the time to take care of my progress and providing valuable advice on research and future careers. I have benefited immensely from Anne Bremer's encouragement, discussions about my chapters and future opportunity. Their endless support and care have been my strength in this PhD.

I acknowledge David Lugo-Palacio and Keiko Miyagawa for validating the data extraction protocol. I am greatly indebted to David and Keiko, who made this thesis research more reliable through their dedication to validating the protocol. David has provided constructive feedback and support from start to finish (not to mention many pints of beer!). I am incredibly grateful to my friends and colleagues who provided advice, laughter, encouragement, coffee breaks (and beer breaks), proofreading, etc. In particular, I would like to thank Henry, Lea, Denise, Silvia, Kelly, Manuela, Kay, Julia, Nikita, Sirin and you girls, my besties.

Finally, I dedicate this work to my parents, who support me in doing this PhD in the UK and send endless love and encouragement no matter what day or night. I also thank my partner Young, for always being next to me and showing his confidence in me. Thank my two bros for encouraging me to keep going. They are my strengths. Last, I would like to commemorate people who left us prematurely and celebrate the new lives coming to us. As your family, I will always pray for you.

Table of Contents

List of tables

List of figures

Abbreviations

AD	Availability of direct treatment comparison
BCRi	B-cell receptor pathway inhibitor
CDF	Cancer drugs fund
CUP	Compassionate use programme
DSU	Decision support unit
ERG	Evidence review group
EunetHTA	European network for health technology assessment
EV	External validity of a randomised controlled trial
FAD	Final appraisal determination
FDA	US Food and drug administration
GIDEON	Global investigation of therapeutic decisions in hepatocellular carcinoma and of its treatment with sorafenib
GP	General practitioner
HMRN	Haematological malignancy research network
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
InV	Internal validity of a randomised controlled trial
IR	Incidence rate
ISPOR	The Professional Society for Health Economics and Outcomes Research
MAIC	Matching adjusted indirect comparison
MS	Maturity of survival data
MTA	Multiple technology appraisal
NICE	National institute for health and care excellence
NHS	National health service

NMA	Network meta-analysis
NPP	Named patient programme
NSCLC	Non-small cell lung cancer
ONS	Office for National Statistics
OR	Odds ratio
OS	Overall survival
PFS	Progression free survival
PO	Proportional odds
PR	Previous technology recommendation by NICE
QALYs	Quality adjusted life years
RCTs	Randomised controlled trials
RWD	Real-world data
SACT	Systemic anti-cancer therapy
SC	Stage of cancer
SEER	Surveillance, epidemiology, and end results program
STA	Single technology appraisal
STC	Simulated treatment comparison
TA	Technology appraisals
TCT	Targeted cancer therapy
TKI	Tyrosine kinase inhibitors
TSD	Technical support documents

Preface

This thesis has been written as a research paper style thesis. Each chapter in this thesis, apart from the Chapter 1 introduction and Chapter 7 discussion, is developed as stand-alone work that can be read independently. These individual research papers are integrated into a single document for this thesis. This research paper style has brought three consequences.

- Chapters may be shorter than a book-style thesis.
- Repetition can be more often observed in comments on the research background and the information about the data preparation and methods.
- Flow and linking can be less tightly bound than a book-style thesis, although this thesis is carefully structured.

This thesis consists of seven chapters. As of the submission, three chapters are peer-reviewed. Two chapters are published, and one is accepted by open-access, peer-reviewed journals. It is noted that Chapter 1 was partly based on a book chapter, *Real-World Data in Health Technology Assessment: Do We Know It Well Enough?* in a book, *Precision Oncology and Cancer Biomarkers: Issues at Stake and Matters of Concern* (May 2022, doi.org/10.1007/978-3-030-92612-0).

- Published
  - Chapter 2 *Protocol for data extraction: how real-world data have been used in the National Institute for Health and Care Excellence appraisals of cancer therapy* is available on *BMJ Open* (published in January 2022, doi.org/10.1136/bmjopen-2021-055985).
  - Chapter 6 *Don't Think Twice, It's All Right": Using Additional Data to Reduce Uncertainty Regarding Oncologic Drugs Provided Through Managed Access Agreements in England* is available on *Pharmacoeconomics Open* (published in September 2022, doi.org/10.1007/s41669-022-00369-9).
- Accepted
  - Chapter 5 *Exploring uncertainty and use of real-world data in the National Institute for Health and Care Excellence single technology appraisals of targeted cancer therapy* is accepted by *BMC cancer* in November 2022.

Chapter 1. Introduction

Oncology research has been accompanied by important health care innovations in cutting-edge technology over the last decades (1). Although innovative cancer treatments have improved the care for cancer patients, these treatments have generated complicated decision-making problems. Randomised controlled trials (RCTs) are regarded as the gold standard for clinical evidence but are often not enough to provide information for economic evaluation concerning outcomes and resource costs of health care. Recently, there is an increasing interest in the application of real-world data (RWD). These data are expected to help provide scientific and systematic evidence to support policymakers in health technology assessment (HTA). Although RWD appears to be new, such data have been already used in many areas, even in the context of HTA. This thesis investigates to what extent RWD have been used and in which circumstances RWD have been used in the context of HTA, especially focusing on cancer appraisals of the National Institute for Health and Care Excellence (NICE).

1.1 Health Technology Assessment

Evidence-based practice is an "integration of best research evidence with clinical expertise and patient values" (2). When integrating the evidence, it is necessary to consider all available data in an unbiased, transparent and scientific manner. As an example of evidence-based practice, HTA is a systematic evaluation of short- and long term safety, clinical effects, and cost-effectiveness of a health technology and technology-related social, economic, and ethical issues in terms of health care resource use (3–5). The assessment aims to provide systematic and structured evidence for policymakers in order to formulate safe and effective health policies and achieve the best value (6). Over recent decades, HTA has become critical as evidence-based decision making has become more prominent in the health system.

1.1.1 The National Institute for Health and Care Excellence

NICE has a responsibility for advising the National Health Service (NHS) on balancing the best care with value for money across the NHS and social care in England and Wales (7). NICE uses a deliberative process to appraise new technologies (8). Its guidance is often considered as a benchmark by other health systems. NICE produces evidence-based guidance, including technology appraisal (TA) guidance and advice for health, public health and social care practitioners (9). TA guidance allows NHS patients to have the most clinically and economically effective treatments by providing information on clinical and cost-effectiveness of health technology.

1.1.2 Single technology appraisal

In the NICE health technology evaluations manual, five processes are available for assessing technologies: single technology appraisal (STA), multiple technology appraisals (MTA), cost comparison, rapid review and update after loss of market exclusivity of a technology (10). Each appraisal process differs substantially in terms of a format of appraisals and a principal responsibility for producing the main evidence. In rapid review of guidance and update after loss of market exclusivity of a technology, new evidence on clinical or cost outcomes is not necessarily provided. Rapid review of guidance is a form of price negotiation only used in approved patient access schemes or commercial access agreements. If new evidence is submitted, NICE considers whether it is acceptable or new appraisal assessment is required. Update after loss of market exclusivity of a technology focuses on the update on the economic model which is caused by biosimilars or generics. Cost comparison analysis is only acceptable when the clinical evidence supports the similarity between the intervention and comparators (11). These three processes are for the technology already fully scrutinised and require only a few additional adjustments. As these processes were explicitly outlined in the recent NICE methods and processes manual, there are too few examples of such appraisals to date. Hence, this thesis will not include them in the following chapters.

Many more assessments are conducted using STA and MTA. An MTA assesses several drugs or treatments used for one or more conditions and focuses on the evidence produced by the independent Assessment Group (12). On the other hand, an STA is an appraisal to assess a drug for one condition and the main evidence is produced by the manufacturer. As the way of gathering the information is distinctive, the appraisal timeline also differs. MTAs usually have a longer appraisal timeline than STAs. As patients' access becomes an important issue, MTAs are much less common than STAs in appraisals of cancer drugs. Given that STAs are the predominant form of appraisals for cancer drugs, this study only focuses on STAs which assess a single treatment.

In STAs, there are four important steps – 1) scoping, 2) phase 1: initiation of the STAs and evidence submission, 3) phase 2: evidence review by NICE and Evidence Review Groups (ERGs), and 4) phase 3: appraisal (13). The first step is to develop the remit and scope. The scope defines research questions including population, clinical settings, comparators, main health outcomes, costs and other special considerations. Phase 1 is initiated after the scoping stage. The company provides an *evidence submission* following a submission template. During phase 1, NICE ensures that the company prepares the best possible evidence for the appraisal by clarifying the issues. If the submitted evidence is adequate, the ERG independently critically reviews it and prepares an *ERG report*. Finally, the Appraisal Committee reviews the evidence from the manufacturer and the ERG and makes the final decision. The appraisal committee's decision and discussion can be found in a *final appraisal determination (FAD)*. Although several other documents can be found during appraisal such as clarification letters, patient and professional group submissions and expert statements, the major evidence for decision making can be found in these four documents (final scope, company submission, ERG report, FAD). Hence, these documents were used as primary sources to extract the information about use of RWD.

1.1.3 Challenges in HTA

In the hierarchy of evidence, RCTs are regarded as the highest level of evidence to show efficacy and safety (14). Although RCTs are the gold standard of evidence, there can be some cases where RCTs are not feasible. For example, a medicine for treating a rare cancer has issues with having good quality of RCTs due to the lack of appropriate trial designs, proper measurements to complement the trial design, the selection of the correct sample and ethical recruitment to participation (15). Also, health economic models used in drug appraisals require a range of data, not all of which are available from RCTs. Figure 1.1 presents the challenges often found in drug appraisals. There are three main challenges – generalisability of the outcome of a clinical trial, understanding long-term effects and the absence of direct treatment comparisons (16–18).

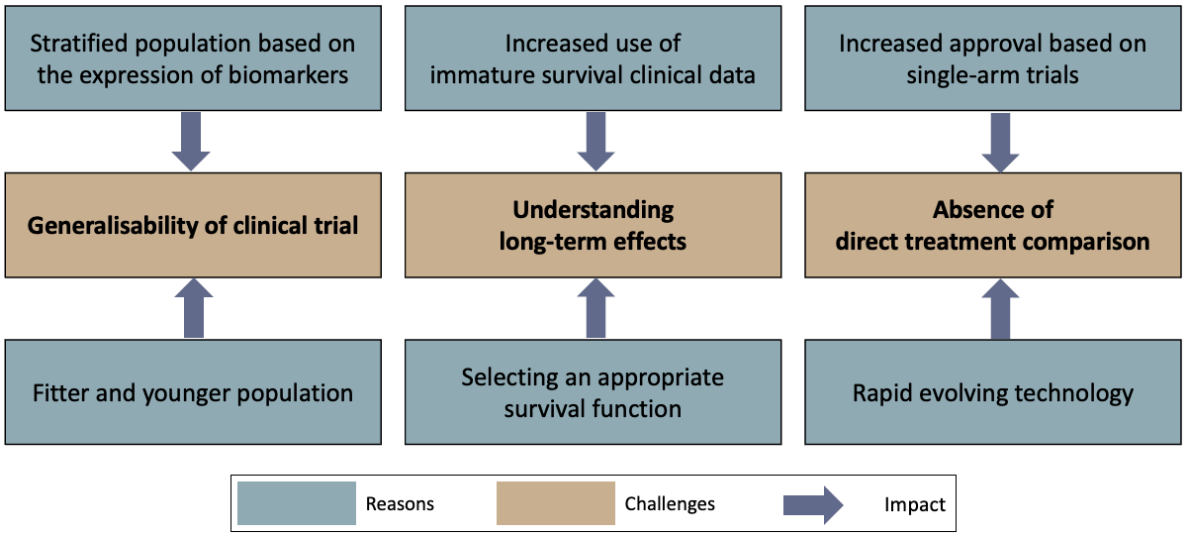


Figure 1.1 Current challenges in health technology assessment for cancer drugs

The generalisability of RCT outcomes is often discussed in drug appraisals. As RCTs tend to recruit a fitter and younger population, the representativeness of the trial data is often uncertain (19). Also, it is recognised that different subgroups defined by the expression level of biomarkers or previous treatment sequence will have different impacts on the efficacy or safety of a medicine. How well a trial population represents the population in routine practice is one of the important questions in appraisals.

Another challenge is to understand long-term effects of the drug of interest. It is important to understand how the drug will work in the future. However, it is practically impossible in RCTs to observe all the death events, with important implications for assessing the long-term impact of the intervention. As a result, the survival rate needs to be extrapolated based on the survival data observed in the trial. When extrapolating survival curves, the survival estimates can be different depending on which survival distributions are used due to varying functional forms (20). It has become more challenging to understand long-term effects of treatments due to increased use of immature survival data in evidence submissions (21). If the observed clinical survival data are for too short a period, a substantial amount of extrapolation is required. This introduces uncertainty around the long-term effect of the drug. It also impacts on obtaining appropriate estimates for both clinical- and cost-effectiveness. Consequently, immature survival data generate uncertainty around the results of economic models.

Increased use of indirect treatment comparisons is another challenge in drug appraisals. It is common that information about relevant comparators for decision-making might not be available in one single RCT. Indirect comparison methodologies like network meta-analysis (NMA) are used when head-to-head comparison is not available. Although NMA cannot fully provide the direct treatment comparison effects, it helps to understand the comparative effects indirectly. However, there are several cases where NMA is not feasible. Recently, more oncologic medicines have been approved based on single-arm trials (22). Indirect treatment comparisons are described as unanchored when there is a disconnected treatment network or single arm studies (23). A statistical method such as matching adjusted indirect comparison (MAIC) can be used to compare treatments across separate trials indirectly by balancing the trial differences regardless of whether anchored or not (24). Although indirect comparisons have been used in several appraisals, there are some concerns about the methodology including unclear understanding of underlying assumptions, inappropriate search and selection of relevant trials, and lack of objective and validated methods to assess (25). Most

importantly, it is unlikely to be possible to adjust for all the differences between trials in the distribution of variables which potentially influence outcomes.

1.1.4 Cancer Drugs Fund (CDF)

NICE has a provision through a separate fund for cancer drugs called the Cancer Drugs Fund (CDF). This is a source of funding for a cancer drug of which uncertainties are too great for it to be recommended for routine commissioning. While the CDF was introduced in 2010 – 2011, this original model had some problems due to the absence of entry and exit criteria for drugs (26). In 2016, the CDF was revised to fund promising new drugs more sustainably (27). There are clear entry and exit criteria in the revised CDF (from here, 2016 CDF). The 2016 CDF offers a mechanism for conditional approval, *recommended for use in the CDF*. A drug not recommended for routine commissioning due to clinical uncertainty, such as unclear long-term effects because of immature survival data, can be recommended within the 2016 CDF if the drug meets the eligibility criteria (28). These criteria include whether the drug has the potential to be cost-effective and whether its model is structurally robust. Also, another criterion is whether further data collection can reduce uncertainty. This is one of the essential features of the 2016 CDF. The drug provided through 2016 CDF should collect additional data following a data collection arrangement (29). During the provisional period, more evidence is collected, mainly focusing on the clinical effectiveness of the drug, to reduce uncertainty. The evidence includes RWD as well as trial data. Among diverse sources of RWD, Systemic Anti-Cancer Therapy (SACT) dataset is highlighted in the 2016 CDF. This dataset is routinely collected for patients who receive anti-cancer treatments from NHS England providers under the responsibility of NHS Digital. The additional data collection, including RWD, is expected to address issues of uncertainty and help reduce them (30).

1.1.5 Increased interest in real-world data (RWD)

There has been increasing interest in making greater use of RWD in HTA. As part of this general trend, NICE has also expressed its interest in RWD. While NICE was already committed to embracing

all available evidence to appraise innovative health technologies, they have set out their ambitions to increase and extend the use of data, including RWD in the development and evaluation of NICE guidance. NICE mentioned that "This type of evidence (real-world evidence) is an important topic, and NICE health technology evaluations are ambitious in ensuring that we make the most of this valuable resource (p.10)" (31). In June 2022, NICE published the *Real-world Evidence Framework* (32). In this framework, NICE aims to provide guidance to improve the quality of RWD so that such data can be used to resolve knowledge gaps and to improve access to innovative technologies. This framework refers to the definition of RWD and how to set up the research design to use RWD. It is notable that the use of RWD to form external control arms for comparative effects is highlighted in this framework, and several ways to minimise the bias when RWD are used for external control arms are included. As a study design, the target trial approach is recommended for mimicking the randomised trial. Also, the identification of confounders and the use of statistical methods to address confounders when analysing RWD is emphasised. The framework provides detailed guidance to assist proper use of RWD in NICE decision-making.

1.2 Definition of real-world data

1.2.1 Various definitions of real-world data

Despite strong interest in RWD, there is no consensus over its definition. Different health organisations have their own definitions (Table 1.1). As an umbrella term, RWD cover broad categories of data. The most cited definition of RWD is probably that of Makady et al., who proposed four broad categories to define RWD (33). While the definitions have some similarities, there are relatively large differences between them. These differences allow for RWD to be more flexible and evolve with advances in technology (34).

Table 1.1 Definitions of real-world data

Organisation	Definition
NICE Real-world evidence framework (32)	Data relating to patient health or experience, or care delivery collected outside the context of a highly controlled clinical trial
NICE Evidence Standards Framework for Digital Health Technologies (35)	Data not collected in the context of RCTs, but either primary research data collected in a manner which reflects how interventions would be used in routine clinical practice or secondary research data derived from routinely collected data * This definition adopts the definition of Innovative Medicines Initiative project *GETREAL*
US Food and Drug Administration (FDA) (36)	Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources, including electronic health records, medical claims and billing, product and disease registries, as well as health-related data from mobile devices
The Professional Society for Health and Economics and Outcomes Research (ISPOR) task force report (37)	Data used for decision making that are not collected in conventional RCTs
Association of the British Pharmaceutical Industry (38)	Data obtained by any non-interventional methodology
Makady et al. (33)	Category 1: Data collected in a non-RCT setting Category 2: Data collected in a non-interventional/non-controlled setting Category 3: Data collected in a non-experimental setting Category 4: Other

1.2.2 Definition of real-world data in this study

The various definitions of RWD cover different concepts or types of information. The lack of agreed guidance for the operational definitions of RWD could be a limitation to apply research findings and roles in HTA (39,40). A review of previous studies on the use of RWD in decision-making, shows that most studies defined RWD broadly in line with the FDA definition. However, the definition of RWD is not clearly identified in many studies about RWD. For example, Grimberg et al. reviewed the challenges of RWD in a literature review without applying a clear definition of RWD (41). Although the FDA definition and that explored by Makady et al. are commonly used, each definition involves challenges for research purposes due to their flexibility. For instance, an observational study which collected patient data *once* from routine clinical practice can be classified differently depending on

which definitions are used. Under the FDA definition, it is not RWD as the data are not routinely collected. On the other hand, it is RWD according to Makady et al., data collected in a non-experimental setting. As RWD often refer to data which reflect routine practice, requiring data to meet both definitions can help to reduce the discretionary interpretation of RWD. Hence, this study combines the FDA's definition of RWD and one from Makady et al. In this research, RWD is defined as *data relating to patient health status and/or the delivery of health care routinely collected from non-experimental settings*.

1.2.3 Grey areas

Although this combined definition can reduce the flexibility, grey areas remain. Under this definition, data from Compassionate Use Programmes (CUPs) are excluded from RWD. CUPs are schemes, which allow patients who cannot enter a clinical trial to use the unauthorised medicine under strict conditions (42). These are also called Named Patient Programmes (NPPs). Although these data are collected outside of clinical trials, the data are not routinely collected. Also, CUPs often have purposes of stimulating interest and familiarity with the technology amongst clinicians rather than reflecting the routine practice. Therefore, data from CUPs are excluded in this study. However, CUP data could be defined as RWD under other definitions as it is not collected within an RCT. A few appraisals used CUP data for survival outcomes for the intervention or the comparators. For example, venetoclax for treating chronic lymphocytic leukaemia (NICE TA487) used data from the ibrutinib NPP to inform the modelling of the palliative care comparator in the base-case analysis (43). Differently from registry or electronic health records, CUPs/NPPs collect data in interventional settings from patients who have a disease with no satisfactory authorised therapies or cannot enter a clinical trial (44). As these data are not routinely collected, they are not regarded as RWD in this study.

Expert opinion and consultee statements are further examples. In an appraisal, consultee statements describing the patient's experience of having the condition or of receiving NHS care can give the

Committee a good insight regarding patients' concerns, priorities or needs. Also, experts' opinions can help justify the company's modelling assumptions. However, since the data are not routinely collected in a systematic manner, unstructured forms of data (e.g., physician notes, clinical experts' opinions, consultee statements) are not viewed as RWD. Although these statements or expert opinions are not structured data, such information could be obtained from structured sources of data such as registry or medical charts. For example, in NICE TA802, the committee was informed by NHS England experts about whether the baseline characteristics of the UK population eligible for the intervention are aligned with the trial patients (45). Clinical experts obtained information from the Systemic Anti-Cancer Therapy (SACT) dataset that 75% of people eligible for the decision problem were aged 70 and above, and 36% were aged 80 and above. While model assumptions were supported by a form of the expert opinion in the appraisal, the rationale of this opinion came from the SACT data, which is a form of RWD. As it is difficult to check if the expert opinion comes from a certain database, all expert opinion was regarded as non-RWD. However, it is noted that this was an operational decision for this research and expert opinion might be regarded as RWD in other situations.

Another example is data collected from routine settings such as hospitals as a one-off collection. These one-time collected data are not RWD according to the definition as the data are not routinely collected. For example, a study by Lloyd and colleagues on health state utility values is frequently used in the NICE appraisals. In this study, they interviewed the general public to get elicit the societal preferences about treatment of metastatic breast cancer. The study was designed to include 100 people in order to try and represent the preference of the general public once in the study period (46). Whilst the health utility values were collected outside clinical data, data about health status were collected within the study protocol and not routinely collected. Hence, data from this study were not defined as RWD.

Survey data present similar problems. One-off survey data that are collected from clinical experts are not defined as RWD in this study, as survey data are not routinely collected. However, there are

some survey data which are routinely collected from the general population. The Health Survey for England is annually collected to monitor the change in the health and lifestyles of people (47). Although the survey is conducted annually, the data only from a specific year can be used. It is debatable whether these annual data should be defined as RWD.

1.3 Opportunities to use real-world data in health technology assessment

Makady et al. have explored how RWD can supplement and enrich the evidence in the arena of health decision-making (48). Information about comparators, extrapolating survival distribution, appraising treatment for rare cancer and increasing generalisability are common areas where RWD can be used (Table 1.2). Some studies have reviewed the use of RWD, for instance, Allen et al. reviewed drug regulatory and HTA submissions for gene therapies and found that NICE included RWD as sources of input for cost-effectiveness analyses (49). The identified studies are summarised below with some examples of the use of RWD identified from NICE appraisals.

Table 1.2 Key opportunities of use of real-world data in health technology assessment

· RWD can provide information about the comparators such as the choice of relevant comparators reflecting clinical practice and comparative treatment effects.
· RWD can help appraise treatments for rare diseases or conditions.
· RWD can supplement the information when extrapolating the long-term survival curve after the trial period for economic evaluation
· RWD can help supplement the information on a generalisation of evidence which is hardly captured in clinical trials.

1.3.1 Use of RWD for information about comparators

RWD can provide information about comparators such as the choice of relevant comparators and comparative treatment effects. Anderson et al. reviewed 489 NICE technology appraisals issued from 2000 to 2016 and found that non-RCT data were used for comparative effectiveness in 4% of included appraisals (50). Rizzo et al. more specifically reviewed 29 NICE appraisals of oncology medicines using single-arm trials as the main evidence and found that 55% of included appraisals used multiple sources including RWD for external control arms (51). The challenge of using RWD for

an external control arm was also reviewed by Jaksa et al. who found that the confounding and selection bias were common critiques in seven cases of drug regulatory and HTA submissions (52). An example where RWD were used for comparator effectiveness is an appraisal of axicabtagene ciloleucel (NICE TA559) (53). In this appraisal, an observational cohort study was used to provide data for the comparators. As axicabtagene ciloleucel has been approved based on the ZUMA-1 trial, a single-arm study, comparator data needed to be taken from an alternative source, SCHOLAR-1. This database is a retrospective patient-level study with pooled data from two observational cohorts and follow-up of two large phase 3 RCTs.

1.3.2 Use of real-world data for appraisals of treatment for rare diseases

RWD can be used in HTA to appraise the treatments for rare diseases or conditions, the so-called orphan medicines (54). It can be difficult to populate economic evaluation models for orphan diseases owing to the typically small patient populations. This makes it challenging to conduct good quality RCTs. In most cases, modelling assumptions in appraisals of these medicines are based on clinical experts' opinions. RWD may be the best source for the data required by the economic model to inform health care decision-making for drugs treating rare disease (55). Mickle et al. reviewed use of RWD in appraisals of rare diseases and found that 37% of NICE submissions included RWD (56). When evaluation for orphan oncology drugs was reviewed, RWD were often used within HTAs to support comparative effectiveness (57) and its value seemed to be established in appraisals relying on single-arm trials (58). For example, among 1,930 people diagnosed with follicular lymphoma annually in the UK, only 52 double refractory patients are eligible for the idelalisib (NICE TA604) (59). The manufacturer of idelalisib submitted DELTA, a single-arm trial, as primary clinical evidence along with a comparator cohort created from registry data (HMRN; haematological malignancy research network). The committee acknowledged that it was likely that the HMRN was the only source of comparative data available for the UK population and agreed to accept the estimate of progression-free survival from HMRN even though HMRN data had limitations.

1.3.3 Use of real-world data for extrapolation

RWD can aid the extrapolation of long-term survival for economic evaluations. NICE makes appraisal recommendations based on the estimated costs of interventions in relation to expected health benefits over the lifetime of patients (60). The health benefit, usually in oncology, in terms of improved survival, is extrapolated from clinical trials as they only show the health outcome over limited periods. The extrapolation often has issues with plausibility and mortality risk assumptions, which leads to a rejection of the overall survival (OS) extrapolation by NICE. When survival outcomes in clinical trials are too immature to observe enough events, the extrapolation of survival curves is more likely to be biased. Immature survival data are a key concern in HTA. RWD such as electronic health record derived data may have the potential to reduce uncertainty (61). RWD can provide useful information such as change of disease hazard over a longer observation period. For instance, in an appraisal of pembrolizumab with carboplatin and paclitaxel or nab-paclitaxel for untreated metastatic squamous non-small-cell lung cancer (NSCLC) (NICE TA600) (62), the manufacturer used registry data, a form of RWD, to extrapolate OS. Although the committee preferred the model which did not use the registry data in NICE TA600 because of the absence of second-line treatment in the database and too optimistic assumptions in the model, it shows how RWD can be used to estimate long-term survival in HTA.

1.3.4 Generalisability

RWD can support the case regarding whether results of clinical trials are generalisable. The efficacy-effectiveness gap between tightly controlled RCTs and the real-world is often criticised as a limitation of RCTs (63). In NICE TA310, the ERG highlighted the value of a long-term clinical registry of all UK patients treated with tyrosine kinase inhibitors (TKIs), including afatinib (64). Such data sources could provide a basis for research and audit to inform future assessments of TKIs in a UK specific population. This implies that RWD such as clinical registries can help generalise the result of RCTs by including additional information that characterise the efficacy of a treatment in a real-world setting.

In appraisals of oncology medicines, the choice of comparators and the identification of subsequent treatments are important to populate the cost-effectiveness model as they impact not only survival outcomes but also the costs. Usually, the clinical guideline indicates the treatment line, which clearly informs which drugs are available in each treatment line. However, the pathways of anti-cancer therapy are complex (65). Available treatments are not always equally used in clinical practice. Some treatments are more frequently used than others due to better compliance or clinical prognosis. Also, there can be a lack of an established standard of care for later lines of treatment. In these cases, RWD can provide a snapshot of drug usage. In NICE TA491, a Pan-European chart review, a form of RWD, was used to reflect physicians' choice, which was a comparator in the economic model (66). RWD could help to maintain the validity and generalisability of the evidence by capturing the current clinical practice.

1.4 Challenges using real-world data in health technology assessment

Although RWD have the potential to provide useful information for drug appraisals, there are several challenges to the use of RWD in HTA (Table 1.3). RWD are prone to be biased due to confounders. Also, the quality of the data is often doubted. Different perspectives on RWD collection and lack of practical experience can make the use of RWD difficult (67). Although the limitations of RWD are widely discussed, Murphy et al. found that literature about the use of RWD often included blanket statements about RWD being or not being valuable, which misled and was not useful for HTA decision makers (34). In this section, the challenges of using RWD are discussed more specifically focusing on the HTA context with some examples from NICE appraisals.

Table 1.3 Key challenges in the use of real-world data for health technology assessment

· RWD are subject to bias and confounding factors.
· Unanchored comparison is unavoidable when using RWD.
· Data quality such as incompleteness is often questioned.
· RWD are not necessarily generalisable as they do not always reflect the entire patients or up-to-date practices.

1.4.1 Confounding factors

A major concern with using RWD in appraisals involves confounding factors. A confounding variable is a variable other than the independent variables of interest that may affect the dependent variable. It can lead to erroneous conclusions about the relationship between the independent and dependent variables (68). RWD are prone to generate biased results since it is hard to control for all the confounding factors, including explicit factors as well as underlying factors (69). It is inadequate to distinguish between the effect of the treatment, a placebo effect and the effect of natural history (70). For example, a patient's health status such as cancer stage and underlying health conditions are highly likely to influence clinical outcomes. As the response rate to second-line treatment differs from first-line treatment, it is critical to understand the patient characteristics for precise assessment. NICE TA502 included HMRN audit data for the comparator as the main clinical evidence was a single-arm trial (71). The HMRN data consisted of evidence from a unified clinical network operating across 14 hospitals in Northern England (Yorkshire). The company used data on the benefit of the comparator (R-chemo; rituximab + chemotherapy) from the HMRN audit. However, the ERG had a concern about the evidence that the HMRN audit did not specifically relate to patients with relapsed or refractory mantle cell lymphoma. The ERG also highlighted that the differences in outcome between different patient groups might be subject to confounding.

1.4.2 Unanchored treatment comparisons

Unanchored comparisons provide another set of challenges to the use of RWD in HTA. Unanchored treatment comparisons result from the network of studies being disconnected or single-arm studies (72). Unanchored comparisons are highly likely to bias the result as the comparison is confounded by the differences between the two populations. Since the number of technologies in which single-arm trials are the primary clinical evidence has increased for drug approval and reimbursement assessment, population adjustment methods such as MAIC and simulated treatment comparison (STC) have been highlighted (72). These methods assume that it is possible to take account of all

effect modifiers and prognostic factors and control them. If the assumption fails, it will lead to a biased conclusion. In NICE TA592, only two single-arm trials were available. The comparator data were very limited (73). Therefore, a non-UK retrospective chart review study was included in company's base case analysis. The study evaluated the outcome of patients who took systemic therapy by reviewing patient hospital records (74). The company tried to use STC and MAIC in their indirect treatment comparison. However, it finally chose a naïve comparison due to the uncertainty around missing unmeasured prognostic factors and the validity of the survival curve, which came from a significantly reduced effective sample size (65% of the original sample size). The committee noted that it was not methodologically recommended because outcomes were likely to be confounded by differences between the populations of the studies.

1.4.3 Questions on the quality of real-world data

Moreover, the quality of RWD questions the reliability of the outcome as evidence. To evaluate the quality of RWD, we need to know precisely how the data have been collected and how they have been used in HTA. Due to the characteristics of observational studies, RWD have limitations with respect to the quantity and quality of information. Also, each dataset has different characteristics. It is necessary to understand each dataset separately with the caveat that individual data categorised as RWD have different characteristics. In NICE TA487, the quality of data was one of the key issues for decision making (43). The target population for the decision problem was stratified by 17p deletion/TP53 mutation group and failure of the B-cell receptor pathway inhibitor (BCRi). Therefore, information on chromosomal abnormality and disease staging was essential. While the registry data had information on time from BCRi treatment failure to death, staging information was not complete. The lack of staging information introduced a significant mismatch between the comparator and intervention groups. In the company submission, it was noted that absence of stratification due to the incomplete staging information might contribute to overestimating the survival of palliative care.

1.4.4 Generalisability

Another challenge concerns generalisability. RCTs provide efficacy and safety data with relatively high internal validity, but their results may not be readily generalisable to a broader, more heterogeneous population (75). RWD are often expected to provide information that better reflects clinical practice. While RWD will often be more representative of patients in routine clinical practice this will not always be the case. For instance, the GIDEON study predominantly included Asian patients treated with sorafenib for advanced hepatocellular carcinoma. Since the treatment effect of sorafenib differed by global region, the use of GIDEON data to predict the treatment effect in the UK population was questionable (NICE TA474) (76).

Another example is the appraisal of ceritinib for previously treated anaplastic lymphoma kinase positive non-small-cell lung cancer (NICE TA395) (77). The manufacturer of ceritinib submitted additional real-world evidence, which were medical records reviewed to determine OS and progression free survival (PFS) in patients who were treated with sequential crizotinib and ceritinib between 2008 and 2014. The ERG commented that this retrospective non-randomised study did not show how similar these participants were to those in the ceritinib studies.

In an appraisal of nivolumab for treating relapsed or refractory classical Hodgkin lymphoma (NICE TA462), generalisability of the RWD to UK practice was questioned (78). The company used data from a study drawing on an American hospital database (79) for evidence on the clinical outcome estimates of comparator, OS and PFS. The committee considered whether the population and composition of treatments reflected clinical practice in the UK. The committee concluded that the study population only partially matched the population of interest. Furthermore, it was deemed that the study may not reflect UK practice, notably regarding subsequent treatment rates of allogeneic stem cell transplant. As routine clinical practice varies by country or region, RWD collected from a specific routine practice may not be always transferable to other clinical practice settings. As

observational studies do not always guarantee the generalisability of the evidence, the clinical and social context should be carefully considered when using RWD.

1.5 Research rationale and objectives

As interest in RWD has grown, the use of RWD has been reviewed in several studies. However, the samples included in these studies were too limited to have an in-depth understanding of use of RWD in NICE appraisals. A recent study by Bullement et al. reviewed the use of RWE in NICE appraisals of oncologic medicine (80) by counting the number of appraisals using RWD. This approach does not provide an understanding of how and why RWD were used. A more systematic and comprehensive study is required to understand the past use of RWD in NICE appraisals. The aim of this research is to contribute to having in-depth understanding of previous use of RWD. More specifically, this research aims to meet the following objectives:

- Investigate patterns of use of RWD and characterise the intensity of its use in economic models
- Analyse statistical associations between increased use of RWD/higher level of intensity of use and a set of factors
- Examine the sources of uncertainty and how RWD have been used to supplement the information required in the economic model in NICE appraisals
- Explore common sources of uncertainty which led drugs to being provided through the CDF and the extent to which RWD reduced uncertainty in CDF review appraisals.

This study includes all Single Technology Appraisals (STAs) of oncologic medicines for which NICE issued guidance between January 2011 and December 2021. As this study focuses on the NICE STAs of cancer drugs, the findings from this study might not be applicable to other diseases or other HTA settings. However, a focus on cancer drugs helps to review the appraisals more comprehensively and systematically since cancer drugs share some similar decision problems in appraisals.

1.6 Data sources

NICE STAs of oncologic medicines are the primary data source for this study. This study limits analysis to appraisals published between January 2011 and December 2021 in order to have long enough time period to capture potential changes over time in how RWD has been used but also recognising that STAs before 2011 might be of less interest because enthusiasm for RWD was largely absent. Here, the date when guidance was published refers to the date when the final appraisal determination (FAD) was published which can be regarded as an end point of the decision-making process. Forty oncology STAs appearing before 2011 are omitted as a consequence of the choice of years. This study focuses on oncology because it is one of the dominant treatment areas where STAs were issued over the period. The number of appraisals in cancer therapy was large enough to understand the use of RWD in economic models and compare its uses in diverse ways such as over time or by type of cancer. The relatively small number of appraisals in other therapeutic areas limits the value of any comparison with cancer appraisals. Figure 1.2 shows the number of included STAs per year that are analysed in this thesis. The data used for analysis were extracted following the protocol developed for this study. This data extraction protocol is explained fully in Chapter 2. The data extraction protocol was validated by two independent researchers to check the clarity and replicability of the protocol.

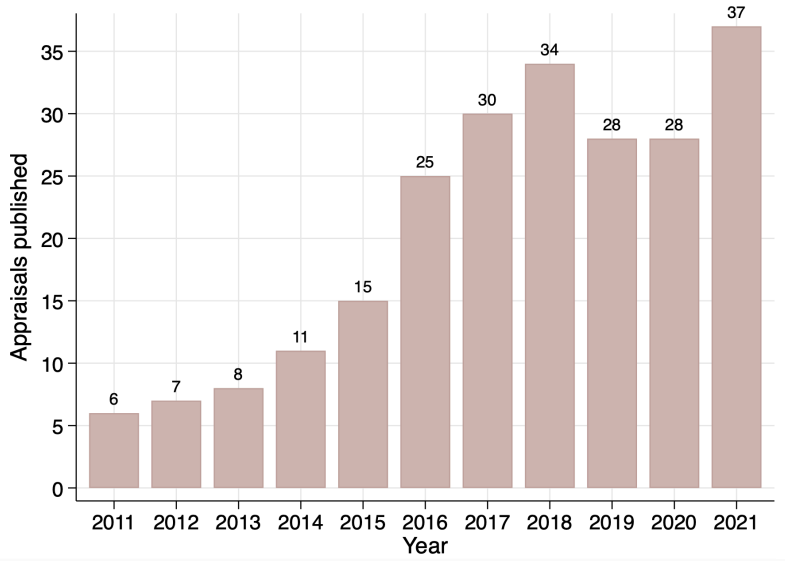


Figure 1.2 Included single technology appraisals for this thesis

1.7 Structure of the thesis

Each chapter of this thesis has been written as individual research papers that can be read as stand-alone pieces of work. A series of research papers are integrated into a single document. As each chapter shares the data and backgrounds of the research to some extent, several recurrent themes could be observed throughout the chapters. There will be as a consequence some repetition of contextual material. This thesis features chapters devoted to providing the background of using RWD in NICE appraisals (Chapter 1), data preparation with an extraction protocol (Chapter 2), patterns/intensity of use of RWD and factors associated with use of RWD (Chapter 3), maturity of data on overall survival (OS) in economic models and use of RWD for estimating OS (Chapter 4), different sources of uncertainty and use of RWD in appraisals of targeted cancer therapy (TCT) (Chapter 5), common sources of uncertainty and role of RWD in CDF review appraisals (Chapter 6) and a discussion of the contributions and limitations of the thesis (Chapter 7). Figure 1.3 provides an overview of what each chapter contributes to the understanding of use of RWD and to the overall thesis.

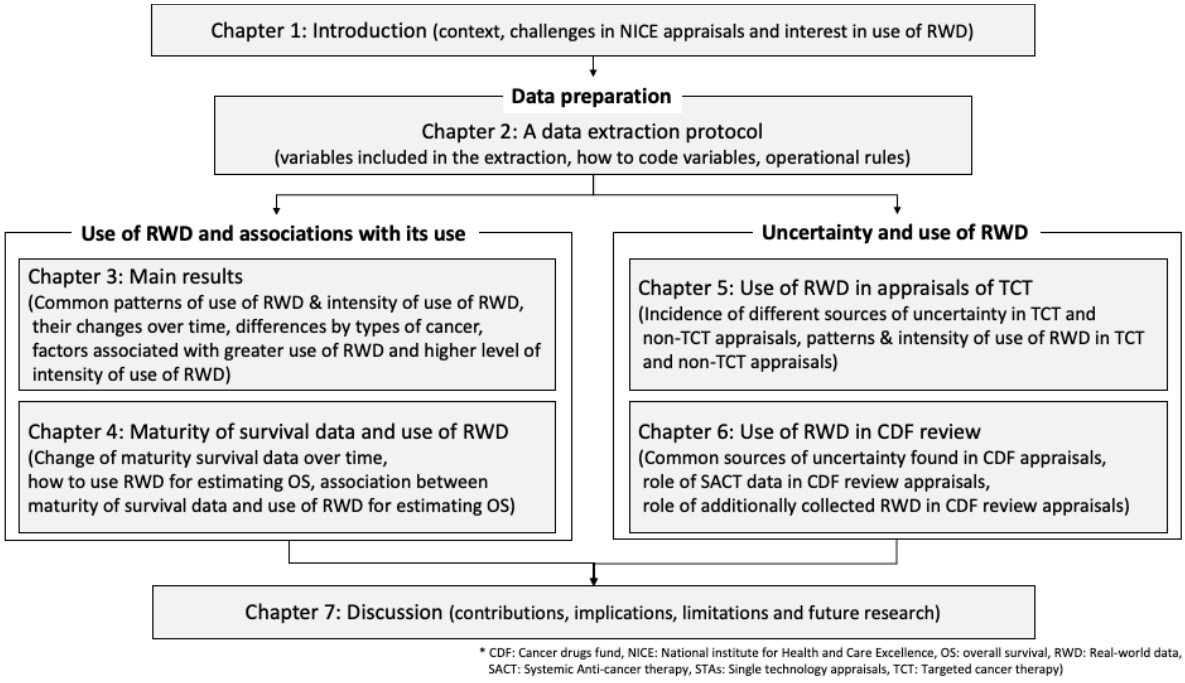


Figure 1.3 Contribution of each chapter to the thesis

The following is a summary of the remaining chapters:

- Chapter 2 describes a protocol for data extraction, which is an essential step in the preparation of data in this thesis. The variables included in the extraction, how to code the variables and operational rules are summarised in this chapter. This chapter will give information how the data used in thesis were prepared.

- Chapter 3 reports the outcome of data extraction. Two approaches - patterns of use of RWD and intensity of use of RWD are used to review the use of RWD in economic modelling. The patterns are reviewed by three types of use: non-parametric, parametric use and any use without distinguishing non-parametric and parametric use. The intensity analysis is conducted by categorising the different patterns of use of RWD by level of intensity. Three major uses are chosen for this categorisation: use of RWD for assisting choice of comparators, use of RWD for estimating OS, use of RWD for estimating volume of treatments. This chapter reviews the changes over time and differences by type of cancer in patterns and intensity of use. In addition, the regression model is estimated to test hypotheses regarding greater use of RWD or higher level of intensity of use of RWD. Here, the factors related to uncertainty or data availability are used in the regression analysis. This chapter will help understand the use of RWD in economic modelling.

- Chapter 4 focuses on the maturity of OS data used in economic models and the use of RWD for estimating OS. It begins with differences in data maturity over time and by type of cancer. Then an ordinal logistic regression model is fitted to test whether maturity of OS data is associated with time/introduction of the 2016 Cancer Drugs Fund (CDF). This chapter also covers how and what sources of RWD are used for estimating OS. Finally, the chapter explores whether maturity of OS data is associated with increased use of RWD in estimation of overall survival by fitting binary logistic regression model. This chapter provides more specific information on the

use of RWD for estimating OS, reflecting current interest in using RWD for comparative treatment effects.

- Chapter 5 reports the sources of uncertainty and how RWD have been used in TCT appraisals. These appraisals are compared with non-TCT appraisals with respect to external validity, availability of direct treatment comparison and maturity of OS data. The patterns and intensity of use of RWD in TCT and non-TCT appraisals are compared. The chapter investigates what and why the differences are found by highlighting the specific features of TCT. This chapter is designed to facilitate the research interest of the funding institute, Centre for Cancer Biomarker directly. However, it notes that the funder is not involved in any aspect of the study conduct.

- Chapter 6 more systematically reviews common sources of uncertainty and the extent to which RWD help reduce the uncertainty in CDF review appraisals. The first 24 drugs exiting the 2016 CDF are reviewed to identify common sources of uncertainty and explore the role of RWD. The original appraisals where recommendation "to use within the CDF" was made and the CDF review appraisals are respectively reviewed to identify the main sources of uncertainty. Then, the extent to which additionally collected RWD, especially the SACT dataset, are able to reduce the uncertainty in the original appraisals is assessed.

- Chapter 7 summarises the contributions, implications and limitations of the thesis as a whole.

1.8 Funding and ethics approval

This PhD was funded by the Centre for Cancer Biomarkers, a Norwegian Research Council Centre of Excellence at the University of Bergen and included tuition fees and an annual stipend over a period 3 years. Ethical approval was given by the London School of Hygiene and Tropical Medicines (LSHTM) ethics committees (ethics reference. 17315). This study uses publicly open data, not related to any human or animal.

Chapter 2. Data preparation: a data extraction protocol

This thesis aims to investigate the use of real-world data (RWD) in appraisals by the National Institute for Health and Care Excellence (NICE). How to extract the necessary data for this thesis needs to be clearly justified and thoroughly explained as this is an essential and critical step to meet the research aim. The rest of the thesis chapters use the data extracted following this protocol. Note that there is some repetition concerning the definition of RWD and the rationale for focusing on single technology appraisals (STAs).

This chapter introduces a data extraction protocol used for extracting information, describing what components are included in the extraction and how these data are extracted. The first section introduces why systematic data extraction is required, followed by a definition of RWD used throughout the thesis. The second section presents what the data extraction template includes and what decisions are made for clear and comprehensive data extraction. It describes the variables included in data extraction, followed by distinguishing between non-parametric and parametric use. After explaining the extraction template, how to code them and the operational rules for extraction are described. This chapter highlights three methodological issues (unclearly stated information, level of aggregation, and no consensus on the definition of RWD) and how this protocol is designed to mitigate the issues.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	157809	**Title**	Ms
First Name(s)	Jiyeon		
Surname/Family Name	Kang		
Thesis Title	Improving economic evaluation and decision-making for oncology drugs using real-world data		
Primary Supervisor	John Cairns		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Open		
When was the work published?	January 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I had primary responsibility for conceptualising and designing the extraction tool and drafting the protocol manuscript.

SECTION E

Student Signature	Jiyeon Kang
Date	7 November 2022

Supervisor Signature	
Date	

Research paper 1

A protocol for data extraction: how real-world data have been used in the National Institute for Health and Care Excellence appraisals of cancer therapy

Jiyeon Kang[1], John Cairns[1]

[1]Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

[2]Centre for Cancer Biomarkers (CCBIO), University of Bergen, Bergen, Norway

Jiyeon Kang: ORCID: http://orcid.org/0000-0002-1587-9674

John Cairns: ORCID: http://orcid.org/0000-0001-6442-0440

Corresponding author

Jiyeon Kang, PharmD, MSc.

Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Address: 15-17 Tavistock place, London, WC1H 9SH, UK

Email: Jiyeon.Kang@lshtm.ac.uk

Keywords: Real-world data (RWD), Health Technology assessment (HTA), economic evaluation, oncology medicine, National Institute for Health and Care Excellence (NICE)

Abstract

Introduction Due to the limitations of relying on randomised controlled trials, the potential benefits of real-world data (RWD) in enriching evidence for health technology assessment (HTA) are highlighted. Despite increased interest in RWD, there is limited systematic research investigating how RWD has been used in HTA. The main purpose of this protocol is to extract relevant data from National Institute for Health and Care Excellence (NICE) appraisals in a transparent and reproducible manner in order to determine how NICE has incorporated a broader range of evidence in the appraisal of oncology medicines.

Methods and analysis The appraisals issued between January 2011 to May 2021 are included following inclusion criteria. The data extraction tool newly developed for this research includes the critical components of economic evaluation. The information is extracted from identified appraisals in accordance with extraction rules. The data extraction tool will be validated by a second researcher independently. The extracted data will be analysed quantitatively to investigate to what extent RWD has been used in appraisals. This is the first protocol to enable data to be extracted comprehensively and systematically in order to review the use of RWD.

Ethics and dissemination This study is approved by the Ethics Committee of the London School of Hygiene and Tropical Medicine on 14 November 2019 (17315). Results will be published in peer-reviewed journals.

2.1 Introduction

In the last few years, interest in real-world data (RWD) has grown in health care decision-making (33). Health Technology Assessment (HTA) refers to the systematic evaluation of clinical- and cost-effectiveness of health technology (3,81). Health technologies include drugs, medical devices, diagnostics, surgical procedures to mitigate health issues and improve the quality of life (82). HTA requires valid and reliable information to evaluate such technologies. Randomised controlled trials (RCTs) have mainly provided the information (83). However, it is challenging to meet all information needs from RCTs since the new generation of therapies pose several assessment challenges. For example, when treatment options are expanding rapidly, it is increasingly unlikely that there are RCTs featuring of all the relevant comparators. Furthermore, the traditional design of RCTs is possibly less appropriate for new technologies such as those targeting rare genetic mutations where it is harder to recruit patients from the clinically relevant populations (84). Moreover, RCTs often have strict inclusion criteria reducing generalisability (85). Another barrier to obtaining the information required for HTA from RCTs relates to the extrapolation of survival. Extrapolation is required in order to incorporate the survival data from RCTs in the health economic model (20). It is more challenging to identify the most appropriate extrapolation the shorter the duration of the trial. If survival data from RCTs are based on a very limited observation period, the extrapolation of the survival curve is likely to fail to predict the long-term effect (86).

The potential benefits of RWD in enriching evidence for HTA are highlighted by the limitations of relying on RCTs (87). This research focuses on the use of RWD in HTA by the National Institute for Health and Care Excellence (NICE). NICE has achieved an international reputation for rigorous development and application of scientific methods to appraise new health technologies to provide its decisions with robust and fair justification (88). More importantly, NICE is noted for the transparency of its processes, responsiveness to change, and commitment to using the best available evidence (89). The structure of the relevant documents facilitates identification of the key information, and the documents are available on the NICE website. Therefore, review of these appraisals can provide

comprehensive information on the evidence used for decision-making. In April 2020, NICE signalled its intention to integrate broader types of data in developing NICE guidance (90). Although it is primarily a statement of intent, it is not a new development in NICE practice since NICE already incorporates a diverse range of published scientific evidence when developing its guidance on health technologies. For example, UK audit data (TA255, 2012), Hospital Episode Statistics (TA559, 2018) and registry data such as the Edinburgh Ovarian Cancer Database (TA598, 2019), Surveillance, Epidemiology, and End Result program (TA562, 2019) have been used in the development of NICE technology appraisal (TA) guidance. While a wide range of data are already used in NICE guidance, there is limited understanding regarding how and where RWD has been used, and in which circumstances RWD is accepted as relevant. Research is required to investigate systematically patterns in the use of RWD and to understand the driving forces behind its use in NICE appraisals.

Several researchers have reviewed practice across HTA bodies (91,92) or reported the use of RWD in HTA (93). However, little systematic research has been conducted. Important information is missing such as how they included literatures without selection bias, which parts of the evidence were reviewed, whether they have clearly defined RWD and justified or explained why this definition is relevant and how different HTA systems were compared given their different practices. Roberts et al. addressed the potential role of RWD in bridging the evidence gaps (94). However, they illustrate the use of RWD with a few examples, rather than providing a fuller picture of current practice when using RWD. Bullement et al. recently reviewed how RWD informed single technology appraisals of cancer drugs in NICE (80). Although this study follows a more systematic approach to the review of the use of RWD, a data extraction table was not provided and the authors focused only on how RWE influenced the cost-effectiveness analysis, and not on how RWE was used to support or establish the appraisal. Due to limited information presented concerning the review process in this study, it is unclear whether the information presented provides a full picture of the use of RWD. Bullement et al. included 113 STAs issued between April 2011 and October 2018. As interest in RWD is increasing over time, it may miss relevant information from recent years. This extraction protocol is required to

help extract the data systematically from appraisals, to increase the reliability of the results of the analysis and to permit a more detailed description of the use of RWD and analysis of factors influencing its use.

A protocol is required to ensure the consistency of data extraction so that the risk of unsystematic data collection is reduced. The main purpose of this protocol is to extract data from NICE appraisals in a transparent and reproducible manner to answer, “how has NICE incorporated a broad range of evidence in the appraisal of oncology medicines.” Without proper justification and operational rules, the data may not be extracted consistently, with a risk of biasing the analysis. The extracted data are expected to be objective and less biased. By consolidating these data, subsequent analysis can provide more robust answers to questions regarding how RWD has been used in NICE technology appraisals. Furthermore, this protocol facilitates the development of a rich dataset which can highlight not just where RWD has been used but also what types of evidence have been used in the HTA process in line with NICE’s interest in incorporating a broad range of evidence. The data can be analysed to answer several research questions including “how has RWD been used in NICE appraisals” and “which factors are associated with increased likelihood of the use of RWD” in depth.

2.2 Methods and analysis

NICE appraisal documents are identified following inclusion criteria (Figure 2.1). The information is extracted from identified appraisals in accordance with extraction rules. The detailed extraction rules can be found in supplement 1. The extraction tool includes evidence-related information such as characteristics of the main clinical evidence and the economic evaluation model and other information. Using this tool, information will be collected about which parts of the cost-effectiveness analyses used RWD. Analyses of the intensity of use of RWD and regression analyses are planned. The data analysis is expected to start from January 2022 and be completed by December 2022.

Inclusion criteria
- STA of oncology medicine
- Appraisals issued from January 2011 to May 2021
Exclusion criteria
- Appraisal of technology for preventing the complications of cancer
- Appraisal of surgical practice and other therapeutic therapies
- Appraisals for which evidence is not available (withdrawn appraisals) or was never supplied (terminated appraisals)

Figure 2.1 Inclusion/exclusion criteria

2.2.1 Definition of real-world data

A definition of RWD is clearly required before extracting information about the use of RWD in NICE. RWD is an umbrella term which covers broad categories of data. Although RWD is increasingly addressed in the literature, there is no consensus over the definition. One of the commonly used definitions of RWD is that of the US Food and Drug Administration (FDA)(95). Another widely cited study regarding the definition of RWD is Makady et al. (33). Each definition has relatively large operational flexibility to be used for data extraction. For example, companies sometimes present phase 1 clinical trial as RWD. However, these data hardly provide insights in the discussion of the use of RWD in HTA. Requiring data to meet both definitions can help to reduce the discretionary interpretation of RWD. Hence, this study uses a definition combining a category of the study designs of collecting RWD explored by Makady and his colleagues' study and the FDA's definition of RWD focusing on routinely collected data. In this research, RWD is defined as the data relating to patient health status and/or the delivery of health care routinely collected from non-experimental settings.

2.2.2 Step 1 Appraisal selection

The first step of the research identifies the NICE TA guidance which meets the eligibility criteria. TA guidance are publicly available on the NICE website (www.nice.org.uk). This study focuses on four types of appraisal documents, the final scope, the manufacturer's submission, the evidence review

group (ERG) report, and the final appraisal determination. These documents are reviewed to establish whether RWD is used to determine any components of the economic evaluation.

Data sources

This research exclusively includes single-technology appraisals (STA) of oncology medicines. Figure 2.1 shows the inclusion and exclusion criteria. One aim is to understand how and where RWD has been used in the appraisal process. Therefore, it is necessary that the appraisal process should be identical. However, the STA and multiple technology appraisal (MTA) processes differ substantially. The MTA has different format of appraisal documents to assess several drugs or treatments used for one or more condition. It is challenging to gather the same information in the MTA process as different actors are responsible for producing and reviewing the main pieces of evidence (12). Besides, STAs are the predominant form in practice, 93% of appraisals of oncology. The small number of the MTAs, only eighteen oncology appraisals, limits the scope for a comparison of MTAs and STAs in terms of the use of RWD. Therefore, this study focuses on STAs, which assess a single treatment. It also limits analysis to appraisals published between January 2011 and May 2021 in order to have a long enough time period to capture potential changes over time in how RWD has been used but also recognising that STAs from earlier years might be of less interest because enthusiasm for RWD was largely absent. Here, the date when guidance was published refers to the date of issuing the final appraisal determination document (FAD) which can be regarded as an end point of the evidence synthesis process (in the absence of a successful appeal). Figure 2.2 summarises a process to identify relevant appraisals for this study.

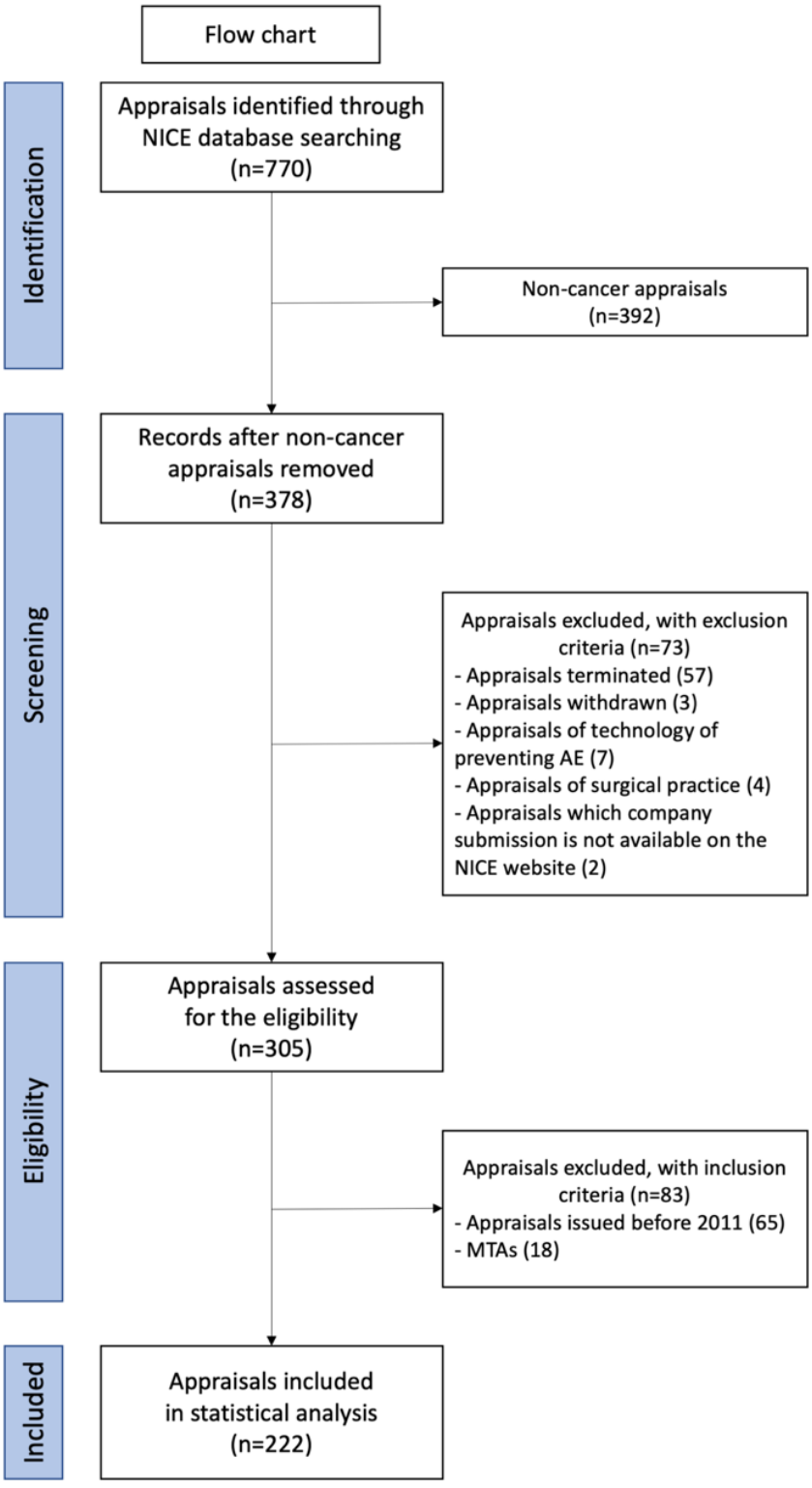


Figure 2.2 Flowchart

Operational separation

Following the inclusion and exclusion criteria, appraisals are identified. Among these appraisals, some TAs have more than one clinical indication or involve combination therapy. It is possible that different evidence was used for the different patient populations in the appraisal. Hence, these appraisals are separated by clinical conditions or treatment lines and reviewed in order to avoid losing information. For example, olaparib for maintenance treatment of relapsed platinum-sensitive ovarian, fallopian tube or peritoneal cancer (NICE TA620) has two separate recommendations for different indications (96). While a patient who has a BRCA1 or BRCA2 mutation and has had three or more courses of platinum-based chemotherapy is eligible for the treatment, a patient who has a BRCA1 or BRCA2 mutation and has had two courses of platinum-based chemotherapy is able to use the treatment within Cancer Drug Fund. Consequently, these indications are included separately in the analysis.

2.2.3 Step 2: Data extraction

A detailed protocol is developed to guide the extraction of essential data for each appraisal in order to investigate the use of RWD in NICE technology appraisals in a systematic and reproducible manner. The protocol is designed to extract information from both the manufacturer's submission (manufacturer's cost-effectiveness analysis) and the final appraisal document (the model preferred by the committee) regarding where RWD was used, and to determine the extent to which the committee supported the use of RWD in these appraisals and understand what factors are associated with supporting or not supporting their use. Figure 2.3 shows the structure of the data extraction template. In summary, the extraction tool consists of three parts – general information, explanatory variables, and outcome variables. The outcome of interest being the use of RWD. The outcome variables record use or non-use of RWD for different elements of the economic evaluation. The information in the base-case analysis and sensitivity analyses will separately extracted. The tool includes all important elements of an economic evaluation. The study will analyse the data to investigate patterns in the use of RWD in NICE appraisals, and the association between several factors and the use of RWD.

Explanatory variables are suggested based on the hypotheses presented under Step 4: data analysis. All items in the extraction template and how to code them are described in the glossary (supplement 1). To convey the type of information to be extracted, some examples from a preparatory review are presented in the glossary.

Parametric and non-parametric use

This protocol distinguishes two categories of outcome variable, parametric and non-parametric use of RWD. Parametric use of RWD is the use of such data to define the numerical value of a specific variable in the economic evaluation, whereas non-parametric use is where data are utilised to develop the model structure or to determine the scope of the evaluation. For example, when RWD are used to estimate survival, this will be counted as parametric use with respect to clinical outcomes (OS/PFS). Parametric use is reviewed and recorded for the intervention and comparators separately as different data could be used in the cost-effectiveness analysis. An example of non-parametric use of RWD can be found in the appraisal of palbociclib for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer (NICE TA495) (97). In this appraisal, the company used information from a study of medical records to determine the subsequent treatments to be assumed in the economic model. This case is regarded as non-parametric use since RWD was used to specify the treatment sequence but not the quantity and cost of subsequent treatment.

Parametric and non-parametric use of RWD and the different categories shown in Figure 2.3, facilitate more consistent data extraction by highlighting the different ways RWD might be used, and provide greater flexibility when testing hypotheses regarding the use of RWD, and the exploration of ways to measure the intensity of use of RWD.

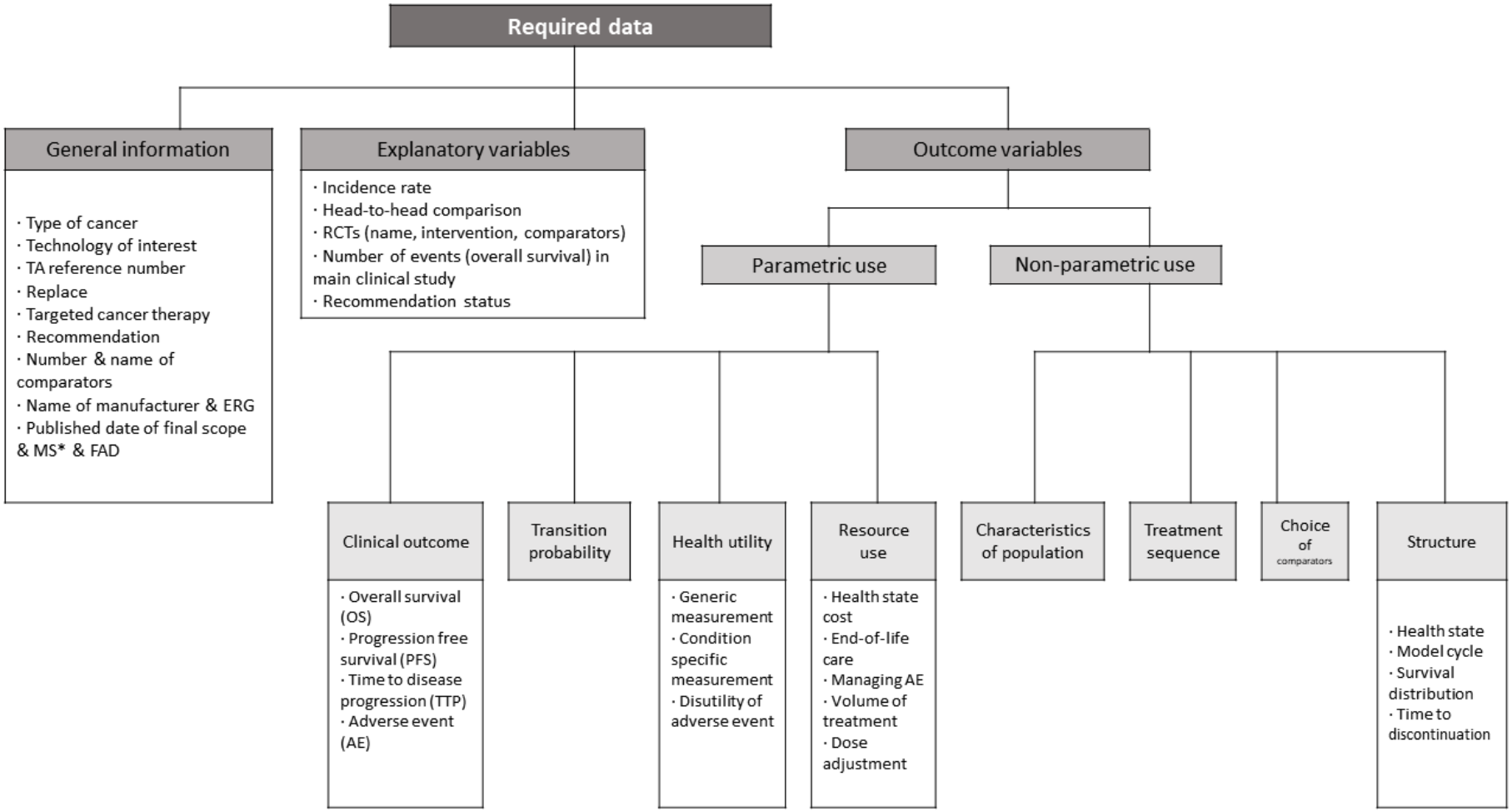


Figure 2.3 The framework for data extraction

Coding

A key issue with respect to improving the reliability of data extraction is how many distinct variables to identify and how finely to divide are the potential responses to these variables. One option, in order not to lose information, is to have many distinct variables with binary responses. Another option is to merge many variables but have multi-level responses. This coding system has advantages which include avoiding information loss, and also grouping together 'similar' information used during appraisals to establish patterns of the use of RWD. This is closely linked to the reason for not using multiple responses in the coding. The template takes an "including all and combining trivia" approach. It helps to include all relevant variables where RWD data can potentially be used, but also to list variables more concisely by merging unnecessarily trivial variables so that the outcome of the extraction can be concretely analysed. Based on two categories, the parametric and non-parametric use of RWD, the areas where data are likely to be used are carefully searched. As a backbone of the extraction structure, distinguishing two categories helped to search each component systematically. Under parametric use, clinical effectiveness, health utility and cost and healthcare resource use were thoroughly reviewed. After sorting variables, they were aggregated if the information is minor and can be categorised into one variable. The area where aggregation is mostly required is resource use. In order to reflect routine clinical practice, especially the cost part has naturally incorporated RWD into the analysis. Estimates of unit costs are usually informed by NHS reference costs (a form of RWD) and thus in order to provide a more sensitive measure of the use of RWD the extraction template focuses on resource use (with respect to cost). However, the measures of resource use are not fully differentiated. Different health technologies include different elements of resource use reflecting their characteristics. Distinguishing all resource use is not an accurate way to understand why and how RWD was used. Although all individual resource uses are not identified, some resource uses, which can be critical in appraisals are differentiated. Variables such as volume of treatment or dose adjustment have potentially critical impacts on the result of economic evaluation. Therefore, these variables are separated from overall resource use.

2.2.4 Step 3: Validation of data extraction tool

The data extraction tool will be validated by a second researcher independently repeating the data extraction for a random sample of appraisals (20% of all appraisals). This validation is required to check the replicability of the data extraction and the clarity of the extraction tool. Any disagreements between the researchers will be resolved by discussion. Peer discussion following the validation process is important not only to check the clarity of this protocol but also to investigate any deviations caused by unclear information. It will help pinpoint where a higher degree of subjectivity may arise in the data extraction.

2.2.5 Step 4: Data analysis

The extracted data will be analysed quantitatively in two different ways. First, counts and proportions will summarise where and how RWD has been used in appraisals. This will be supplemented by an analysis of the intensity of use of RWD in order to explore changes in the pattern of use of RWD over time and differences with respect to cancer type. In addition to descriptive statistics, the association between years and the intensity of use of RWD will be examined. Secondly, a regression analysis will be performed to investigate which factors are associated with the greater use of RWD in a company's submission. As part of the protocol development, some appraisal documents were reviewed to identify factors potentially associated with the use of RWD. Five factors were identified and formulated into hypotheses about increased use of RWD (Figure 2.4).

· Poor internal/external validity of the clinical trial is associated with greater use of RWD.
· Absence of direct (head-to-head) comparison is associated with greater use of RWD.
· Low incidence rate of the disease is associated with greater use of RWD.
· Immature survival data in the clinical trial are associated with greater use of RWD.
· The technology having been recommended in previous NICE TA guidance is associated with greater use of RWD.

Figure 2.4 Hypotheses about increased use of real-world data

2.3 Methodological issues

The design of this data extraction protocol, in which information is reliably and repeatedly extracted across appraisals, will allow us to review evidence for the use of RWD more systematically than could be obtained from conducting several case studies. However, several methodological challenges can be anticipated. This section addresses these challenges and how they might be mitigated.

2.3.1 Unclearly stated information

Overall, NICE appraisals clearly describe the data used in the evidence synthesis. However, sometimes the search process may not be well-documented and the precise source of information may not be clear. Systematic literature reviews are carried out to identify all relevant evidence in appraisals. Clinical effectiveness evidence is carefully examined and described in detail, with clear reasons for the inclusion and exclusion of studies. On the other hand, the systematic search for resource use and cost information usually enumerates miscellaneous studies with bibliographic information and a summary, but the critical review of minor components of health cost is sometimes missing. While manufacturers provide the result of their assessments, some manufacturers' submissions do not clearly state whether a particular study was used to determine an element of resource use making up the health state costs. However, it appears to be rare for there not be an explicit statement regarding the evidence used, mostly with respect to resource use.

2.3.2 Level of aggregation

An important question is the most appropriate level of aggregation. This is best illustrated with respect to healthcare costs. It would be possible to have a variable indicating use or non-use of RWD for every single element of cost (distinguishing GP visits, frequency of hospitalisation, and so on). At the opposite extreme there could be a single cost variable which indicated whether RWD was used for any element of cost. The more aggregated the measure the greater the loss of information, but some elements of cost are much more important than others and the potential analyses of the use of RWD will multiply greatly if there is no attempt at aggregation. The current protocol tries to balance the advantages and

disadvantages of different levels of aggregation by combining several elements into a health state cost variable but distinguishing other important components of cost, such as volume of treatment, dose adjustment and resource use for adverse events.

2.3.3 No consensus on the definition of real-world data

This research uses a definition of RWD merging definitions from the FDA and Makady et al. The distinctive part of the definition used in this research is 'routinely collected' data from a 'non-experimental study'. Although this definition provides a specific and clear definition for this research, there is no consensus on the best definition of RWD. Even the same definition can be interpreted in different ways. For example, some researchers interpret that 'routinely collected' in the FDA definition is 'collected in routine care' whereas other interpret it as 'how frequently data are collected.' It is likely that other definitions of RWD are preferred by other researchers and the data extracted will be influenced by the definition of RWD chosen. While the use of multiple definitions of RWD was considered, it would create practical problems such as multiplying the number of potential analyses and making data extraction take longer. Although the chosen definition can be questioned by other researchers who have different views, the various definitions overlap considerably. It is thus unlikely there will be a marked divergence in the data extracted when using the different definitions.

2.4 Design to mitigate methodological issues

Several operational rules have been designed to minimise bias likely to come from the methodological issues encountered in the data extraction. First, 'not clear' is recorded separately in order to provide a more accurate description of the use of RWD. However, for purposes of data analysis, we anticipate treating these instance as "no RWD" since the code 'not clear' cannot be independently analysed. In addition, having a 'not clear' category in analysis is unlikely to improve data quality since we anticipate that this problem will arise in very few appraisals. Also, information which is not clearly recorded in the appraisal documents is usually not important information with respect to the evidence synthesis.

The approach (extracting all relevant information which can provide meaningful data for analyses) is also closely linked to the reason for using binary code for analysis in this research. Decomposing levels of codes into several small parts can facilitate data extraction. However, it is more likely to increase the complexity since trivial information is individually recorded. The extracted trivial data should be interpreted based on another operational rule. It is subject to increased error, particularly when testing hypotheses. For these reasons, the benefit of using multi-level codes does not outweigh the benefit of binary codes while separation is much more time consuming. Instead of adapting multilevel codes, this study will adopt an alternative approach, an intensity analysis which helps to identify important differences within the diverse patterns of use of RWD. When looking at the pattern of use of RWD, the intensity of use will be analysed. Simply counting the number of times RWD are used is not an accurate way to understand why and how RWD were used. Alternatively, this study focuses on variables which are potentially important determinants of cost-effectiveness in appraisal. Variables such as survival outcome, volume of treatment and choice of comparators are more likely to influence estimated cost-effectiveness. Especially, the survival outcome is the most important information in both clinical and cost-effectiveness as well as one of the controversial areas where to use RWD. The intensity analysis is a framework to show whether RWD is used in these components alongside the quantity of the use of RWD. It can offer more benefits in deeper understanding of the use of RWD than counting all miscellaneous uses of RWD.

2.5 Strengths and limitations

Strengths
· This protocol enables data to be extracted in a transparent and systematic manner for the study of how RWD has been used in NICE appraisals including all the different ways an economic evaluation might use RWD. · This study facilitates systematic understanding of the use of RWD in NICE appraisals over the last 10 years.
Limitations
· Since it is focussed on cancer, the methods and eventually the findings are to some extent cancer-specific. · The protocol could be modified to reflect the HTA context in different countries although the extraction protocol is not fully applicable to the practice of other HTA bodies as much of the protocol reflects the NICE appraisal process. · Since data extraction is based on the four main types of appraisal document it is possible, but not likely that some relevant information concerning RWD is missed.

Figure 2.5 Strengths and limitations of this study

Figure 2.5 summarised the strengths and limitations of this study. To the best of the authors' knowledge, this is the first study protocol to investigate to what extent RWD has been used in NICE appraisals. It allows the practice of extracting information to be reproducible, systematic and transparent. Strengthening the reproducibility and transparency of data extraction can maximise understanding of the use of RWD by allowing more accurate interpretation and use of findings. This protocol could be relevant to researchers or HTA agencies who aim to understand how various data resources are used in HTA in England. Analysis of data generated using this protocol can provide a detailed picture of the use of RWD in NICE appraisals over ten years. Moreover, the study findings could add value to NICE's ongoing work to broaden the evidence used in appraisals.

The protocol has the limitation that it has been developed to study the use of RWD in NICE appraisals of oncology drugs. Consequently, the data extraction protocol may not be fully applicable to appraisals in other disease areas or to the different practice of other HTA bodies. Since the documentation is

significantly different depending on each country's context, it may not be feasible to extract the same information as in the English context. However, many of the distinctions are of wider application, e.g. parametric vs non-parametric use of RWD, and the taxonomy of where in an economic evaluation it might be relevant to look for use of RWD. Also, the hypotheses are potentially of wider application. The results are going to be specific to NICE but otherwise the structure of this research has wider application. Although not fully transferrable, this protocol can be modified for use in other HTA contexts. Lastly, this protocol focuses on four main documents. Relevant RWD may arise at the clarification or technical engagement stage. It is possible there is some information regarding use of RWD that is not reported in any of the four main documents. However, only a small number of such cases are anticipated. If RWD is critically used in a revised model and the committee thinks it is an important change, this evidence is likely to be addressed in FAD.

Chapter 3. Use of real-world data and factors associated with its use

The main objective of this thesis is to explore the use of real-world data (RWD) and factors associated with the use of RWD in the National Institute for Health and Care Excellence (NICE) appraisals of oncology. The previous chapter has described how the data are extracted. The data extraction tool covers extensive information required to understand the previous use of RWD. This thesis describes the use of RWD beyond simple counting. Two ways are used to review the use of RWD in this thesis: patterns of use of RWD and intensity of use of RWD. These methods are newly designed for this thesis. This chapter carefully presents how patterns are identified and the intensity of use of RWD is rank-ordered. The detailed description of these methods facilitates the understanding of patterns and intensity of the use of RWD. As these methods recur in later chapters, some of the information will be repeated because of the research paper style of the thesis.

The first section starts with an introduction to the patterns of use of RWD followed by categorising the patterns into three groups: non-parametric, parametric and any use regardless of type. After a description of the patterns of use of RWD, this chapter describes how the intensity of use of RWD is ordered based on three major uses (for choice of comparators, for estimating overall survival, for estimating volume of treatment). The intensity of use of RWD is illustrated in several different ways: Venn diagram, changes over time and by type of cancer. The second section starts with an introduction to logistic models. Binary logistic regression (logit model) is estimated to investigate the factors associated with greater use of RWD in economic modelling. After the binary logistic regression analysis, ordinal logistic regression (generalised ordered logit model) is fitted to explore the association between the level of intensity of use of RWD and a series of factors. This chapter focuses on investigating the patterns and intensity of use of RWD and the factors associated with using RWD using the logistic models. This chapter achieves the main objective of the thesis.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	157809	**Title**	Ms
First Name(s)	Jiyeon		
Surname/Family Name	Kang		
Thesis Title	Improving economic evaluation and decision-making for oncology drugs using real-world data		
Primary Supervisor	John Cairns		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	Jiyeon Kang, John Cairns
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I had primary responsibility for obtaining the data, performing the analysis and drafting the paper.

SECTION E

Student Signature	Jiyeon Kang
Date	7 November 2022

Supervisor Signature	
Date	

Research paper 2

What is associated with use of real-world data in single technology appraisals of oncologic medicine by the National Institute for Health and Care Excellence?

Jiyeon Kang[1], John Cairns[1]

[1]Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

[2]Centre for Cancer Biomarkers (CCBIO), University of Bergen, Bergen, Norway

Jiyeon Kang: ORCID: http://orcid.org/0000-0002-1587-9674

John Cairns: ORCID: http://orcid.org/0000-0001-6442-0440

Corresponding author

Jiyeon Kang, PharmD, MSc.

Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Address: 15-17 Tavistock place, London, WC1H 9SH, UK

Email: Jiyeon.Kang@lshtm.ac.uk

3.1 Introduction

Real-world data (RWD) are increasingly being used in the life cycle of drug development (98). Clinical trials mainly generate the evidence to support decision making in regulatory and health technology assessment (HTA) whereas RWD have been used complementarily for post-marketing surveillance. Interest in the adoption and implementation of RWD in the context of HTA has exploded in recent years (99) as evidence gaps are more recognised when heavily relying on clinical trials (100,101). Many HTA bodies share positive views toward the use of RWD as acceptable sources of data in the context of HTA (102,103). In most European HTA organisations, RWD are accepted to inform epidemiological data, cost and resources uses (48) and used to fill evidence gaps by supporting the assumptions in economic models (104). Health economics and outcome research organisations such as European Network for HTA (EUnetHTA) and International Society for the Professional Society for Health Economics and Outcomes Research (ISPOR) have engaged in evaluation of technology and recently more involved in how RWD can be collected and analysed in HTA context (37,105–107).

The National Institute for Health and Care Excellence (NICE) has also demonstrated their interest in RWD (108). NICE has highlighted the value of RWD and shown their ambitions to improve the understanding of health care, resolve gaps in knowledge and drive early access to innovations using RWD (109). In June 2022, NICE implemented a real-world evidence framework for the more comprehensive use of RWD in NICE guidance (32). In this framework, the use of RWD is expected to reduce uncertainties that potentially come from the challenges of RCTs such as generalisability of clinical trials, absence of direct comparison and limited long-term follow-up. There is a common ground that RWD can provide useful information in challenging circumstances such as with small patient populations, rare diseases and cases where robust evidence is lacking including single-arm trials (102,110).

Although use of RWD in HTA appears to be relatively new, the use of RWD is not entirely new in NICE. RWD have already been used in NICE appraisals to some extent to develop their guidance in

diverse ways. Registry or hospital data commonly referred to as RWD have been used for economic modelling. For example, NICE issued technology appraisal (TA) guidance for lenalidomide with rituximab for previously treated follicular lymphoma (NICE TA627) using UK registry data for the clinical outcomes of the comparators due to the absence of data from direct treatment comparison (111). Another example is European Chart Review data used to make indirect treatment comparisons when comparative data were absent in an appraisal of ibrutinib for treating Waldenstrom's macroglobulinaemia (NICE TA795) (112).

The use of RWD in NICE appraisals has been reviewed in several studies. When reviewing NICE guidance issued in 2015 and 2016, increasing prominence of RWD was found despite limited use (113). Also, RWD have been used to predict long-term effectiveness in the submissions to different HTA agencies (114). Several studies have discussed the role of RWD to support health care decision-making and have broadly outlined the benefits and challenges of the use of RWD (94,115,116). These studies show that awareness of the value, and use of RWD, in HTA decision making has increased over time. The value of RWD described in these studies is to provide information in the situations where RCTs provided limited information. However, current studies have been limited to understanding how RWD were used in HTA decision making since the discussions about RWD are generally based on case studies. A more comprehensive review of STAs of cancer drugs found that RWD were extensively used in the analysis of cost-effectiveness (80). This review was the first study to report in which part of economic models RWD were used. However, the ability of this study to answer the question "how have RWD been used?" is limited since it only reports the number of times RWD were used as model inputs in appraisals.

Despite tremendous interest in RWD, how RWD are used in NICE STAs and what factors are associated with greater use of RWD are not fully understood. Currently, no study shows how use of RWD has changed over time or varied by type of cancer or explores whether there are distinct patterns of use of RWD within appraisals. More specifically, no studies consider the intensity of use

of RWD in economic models. Current literature and examples in the NICE framework show that RWD have been used in decision-making where RCTs are of poor quality, there is a lack of long-term follow-up and questions concerning generalisability of the results. Although literature illustrate scope for use of RWD, the circumstances associated with increased or decreased use of RWD are unknown. This study investigates the patterns and intensity of use of RWD as well as tests hypotheses about the use of RWD in economic models in STAs of oncology medicines. This study hypothesises that a range of factors are associated with greater use of RWD/higher level of intensity of use of RWD. Exploring the patterns and intensity of using RWD and testing the hypotheses clarify to what extent RWD have been used in appraisals of oncology medicine and how RWD were involved in supplementing the evidence in the model. The specific purposes of this study are (1) to identify and characterise patterns and intensity of use of RWD in economic models and (2) to measure and investigate associations between factors and increased intensity of use of RWD.

3.2 Methods

This study used the data extracted from NICE STAs of oncology medicines. In total, 229 STAs of oncologic medicines for which NICE issued guidance between January 2011 and December 2021 were included. The data were extracted following a protocol developed to document information about the use of RWD in economic models in NICE STAs of oncologic medicines (117). Extracted data include general information about STAs, evidence-specific information such as characteristics of primary clinical evidence, and the use of RWD in economic models. The use of RWD was separately extracted for the base case analysis and sensitivity analyses. The data were processed in several ways for a review of use of RWD. Figure 3.1 shows how the data used in this study were prepared. Alongside the data extraction from appraisals, a semi-structured interview was conducted with HTA stakeholders for more comprehensive understanding of use of RWD. The summary of the interview questions can be found in Appendix 2.

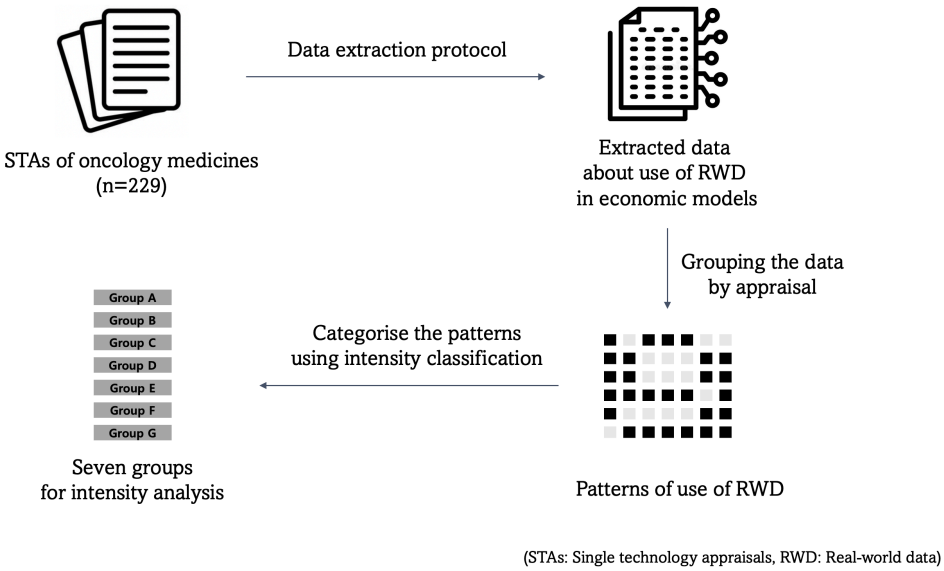


Figure 3.1 Diagram of data preparation

3.2.1 Pattern review

The patterns of use of RWD were reviewed. Simply reporting the number of uses of RWD has limitations to see to what extent RWD were used in an appraisal. Reviewing the use as a pattern can provide more comprehensive information about use of RWD in each STA. Also, it facilitates comparisons of the use, over time or by type of cancer. The extraction protocol distinguished 31 areas in economic evaluation where RWD might be used (Table 3.1). It generated numerous patterns. This study reviewed and identified patterns in three ways, any use of RWD, non-parametric and parametric use of RWD. Any use of RWD refers to use of RWD in any part of the economic model regardless of how RWD were used. Parametric use means that RWD provides numerical values for specific variables in the economic model. For example, the use of data to give estimates for overall survival (OS) or resource use in the economic model are categorised as parametric use. Non-parametric use is to the use of RWD to develop the structure of economic model and to support assumptions in the model. Using RWD to select comparators or to validate the choice of survival distribution are examples of non-parametric use. This separation provides a more informative review of how RWD have been used in justifying a model of economic evaluation as well as estimating parameters in models.

Table 3.1 Components in use of real-world data in data extraction

Type of use	Elements
Non-parametric use	Characteristics of population
	Treatment sequence
	Choice of comparators
	Health state
	Model cycle
	Survival distribution (intervention)
	Survival distribution (comparators)
	Time-to-discontinuation (intervention)
	Time-to-discontinuation (comparators)
Parametric use	Overall survival (OS) of intervention
	Progression-free survival (PFS) of intervention
	Response rate (intervention)
	Time-to-progress (intervention)
	Adverse event (intervention)
	Overall survival (OS) of comparators
	Progression-free survival (PFS) of comparators
	Response rate (Comparators)
	Time-to-progress (Comparators)
	Adverse event (Comparators)
	Transition probability
	Health utility (generic measure)
	Health utility (cancer specific measure)
	Disutility
	Resource use of health state cost
	End-of-life resource use
	Resource use of adverse event cost (intervention)
	Volume of treatment (intervention)
	Dose adjustment (intervention)
	Resource use of adverse event cost (comparators)
	Volume of treatment (comparators)
	Dose adjustment (Comparators)

While both parametric and non-parametric use of RWD were extracted from base case analysis, only parametric use of RWD was extracted from sensitivity analysis since the model structure is often shared between the two analyses. Distinguishing between use of RWD in the base case and in sensitivity analyses enabled to look at not only change of the use in sensitivity analyses but also the gradual move over time from use in sensitivity analyses to increased use in the base case. However, due to the limited observation from sensitivity analyses, the comparison wasn't conducted in this study. The reason of less use of RWD in sensitivity analyses was reviewed in Discussion.

3.2.2 Intensity analysis

The patterns were reviewed by characterising the intensity. Given that the interest in use of RWD is more shifting to leverage RWD to varying extent mainly with respect to treatment effectiveness (118), a review of the patterns by differentiating where RWD were used can facilitate more diverse comparison beyond the general description of use of RWD. The level of intensity was used to differentiate the use of RWD in the patterns. Intensity was determined by how many times RWD were used for major and minor components of the economic evaluation in one appraisal. Three components (OS of intervention/comparator, volume of treatment of intervention/comparators, choice of comparators) were identified, as being highly likely to influence the estimated incremental cost-effectiveness ratio (ICER). They were labelled as major uses of RWD. The experts' opinions from the interview supported this assumption (Appendix 2). The remaining components were considered as minor uses of RWD. The groupings by intensity of use of RWD are shown in Figure 3.2, which uses a Venn diagram to describe how the different groups relate to each other. This grouping is a straightforward way to follow the logic illustrating the relationships between each intensity group. It is possible to see the commonalities and differences as well as to compare the size of groups.

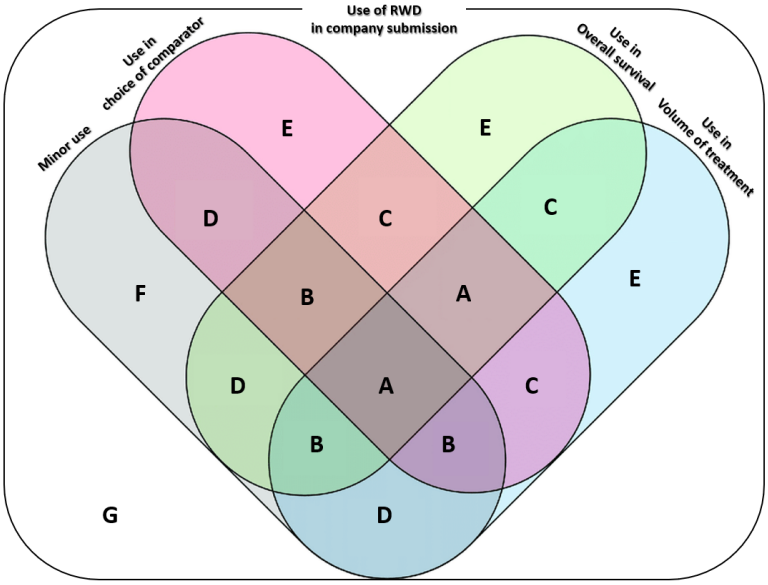


Figure 3.2 Grouping by intensity

The identified patterns were categorised into seven levels of group by characterising major and minor uses of RWD (Figure 3.3). Categorising the patterns by level of intensity is helpful not only to document changes in the patterns of use of RWD but also to identify factors associated with higher or lower intensity of use. Two classifications were suggested. One counts the number of each major and minor component; another is a trimmed classification that only counts the number of major uses. The group with all three major components is the highest intensity group of use of RWD. It was impractical to rate a score for each intensity group given the limited methods to measure it in the appraisal. Instead of rating the score, the intensity was rank-ordered according to the number of major use of RWD in this study.

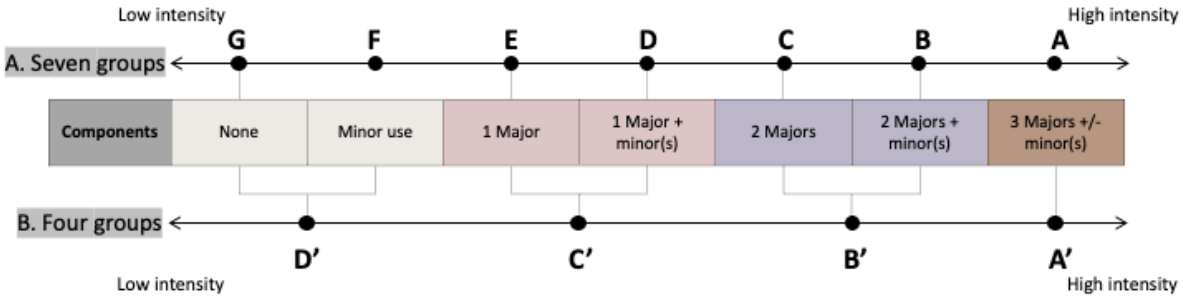


Figure 3.3 Scale of intensity

The sources of RWD used in three major components were also reviewed. In the NICE real-world evidence framework, common sources of RWD found in NICE guidance were summarised such as registry, medical history and other resource use data (32). The sources of RWD were classified based on the common data sources presented in NICE framework. If the relevant category was not available, extra category was added for further classification.

3.2.3 Regression analysis

Regression analysis was carried out to understand the association of greater use of RWD and higher level of intensity of use with a set of factors. There were two types of outcome variables. One is *use of RWD* and another is *level of intensity of use of RWD* in economic models. Two different regression analyses, binary logistic regression and ordinal logistic regression were suggested depending on the

outcome variables. The primary explanatory variable $time$ was measured by the date when final appraisal determination (FAD) issued. The month was used as a unit of $time$. Six additional variables were identified for both regression analyses either because they were all potentially related to data gaps or data availability: availability of direct treatment comparison (AD), incidence rate (IR), maturity of survival data (MS), external validity (EV), internal validity (InV) and previous technology recommendation by NICE (PR). Table 3.2 summarises how the explanatory variables used in the analysis were coded.

Table 3.2 How to code the explanatory variables

Variables	Code
Availability of direct treatment comparison (AD)	· Not available · Some available · All available
Incidence rate (IR)	· Number of expected patients per 10,000 (annual)
Maturity of survival data (MS)	· Extremely immature: Proportion of death events ≤ 20% · Immature: 20% < Proportion of death events < 50% · Mature: 50% ≤ Proportion of death events
External validity of RCT (EV)	· Acceptable external validity · Moderate external validity · Questionable external validity
Internal validity of RCT (InV)	· High quality of internal validity · Low risk of internal validity · Moderate risk of internal validity · High risk of internal validity
Previous technology recommendation by NICE (PR)	· Yes, recommended in other indications · No, it is not recommended yet

Binary logistic regression

A binary multivariate logistic regression was estimated to test the research hypotheses as the outcome variables, different types of *use of RWD,* were binary. The primary outcome variable of this study was *any use of RWD*. This outcome variable was extended to non-parametric, parametric use of RWD and use of RWD for individual components in the economic models. Seven explanatory variables were used for binary logistic regression analysis. Multicollinearity was tested before

conducting the analysis. After checking the correlation between predictors, the variable InV had several correlations with other variables, hence it was excluded in this study. The hypotheses tested in the binary logistic regression were summarised in Figure 3.4.

· Time is associated with greater use of RWD.
· Absence of direct (head-to-head) comparison is associated with greater use of RWD.
· Low incidence rate of the disease is associated with greater use of RWD.
· Immature survival data in the clinical trial are associated with greater use of RWD.
· Poor external validity of the clinical trial is associated with greater use of RWD.
· The technology being recommended in other NICE TA guidance is associated with greater use of RWD.

Figure 3.4 Hypotheses about greater use of real-world data

This regression analysis was carried out using the Logistic procedure in STATA Version 17. Except for $time$ and IR, four predictors were categorical variables, which dummy coding was required for the following models. Odd ratios (ORs) compared the relative odds of the greater use of RWD occurring in the reference given the exposure to the explanatory variables ($time$, AD, IR, MS, EV, PR). The regression coefficients were also presented in the results.

Ordinal logistic regression

A multivariate ordinal regression model was considered as the outcome variable, *level of intensity of use of RWD* is categorical ordinal data. A generalised ordered logit model was fitted to the data in this study since it allows the effect of explanatory variables to vary when the proportional odds (PO) assumption is violated (119). The Brant test indicated the PO assumption was untenable for the variable PR (Appendix 5.5). Therefore, generalised ordered logit model is considered a better option.

With respect to the outcome variable, the simplified classification was used for this regression to provide more observations in each group. Due to the small number of observations in the higher intensity groups, intensity groups A' and B' were merged. As a result, the outcome variable, *level of intensity* had three categories: no major use, 1 major use and more than 2 major uses. As the odds of being beyond a certain category are estimated against being at or below that category in the

generalised ordered logit model (120), each binary model was used for two categories (Table 3.3). Level 1 and 2 were presented in binary logistic model 1 and 2 respectively in the results.

Table 3.3 Category comparisons for generalised ordered logit model (j=1,2)

Category	$logit\ P(Y > j)$	Odds	Probability comparisons
Level 1	$logit\ P(Y > 1)$	$\frac{P(Y > 1)}{P(Y \leq 1)}$	No major use *vs*. 1 major use & more than 2 major uses
Level 2	$logit\ P(Y > 2)$	$\frac{P(Y > 2)}{P(Y \leq 2)}$	No major use & 1 major uses *vs*. more than 2 major uses

The hypotheses were setup for generalised ordinal logistic regression model. A main hypothesis, *the time is associated with higher level of intensity of use of RWD*. The univariate model was extended to multivariate ordinal regression including additional predictors (AD, IR, MS, EV and PR). Same as in the binary logistic regression model, InV was excluded for the analysis due to the multicollinearity. Six hypotheses tested using ordinal regression were presented in Figure 3.5.

· Time is associated with higher level of intensity of use of RWD.
· Absence of direct (head-to-head) comparison is associated with higher level of intensity of use of RWD.
· Low incidence rate of the disease is associated with higher level of intensity of use of RWD.
· Immature survival data in the clinical trial are associated with higher level of intensity of use of RWD.
· Poor external validity of the clinical trial is associated with higher level of intensity of use of RWD.
· The technology being recommended in other NICE TA guidance is associated with higher level of intensity of use of RWD.

Figure 3.5 Hypotheses about higher level of intensity of use of real-world data

The ordinal logistic regression analysis was conducted in STATA version 17. The generalised ordered logit model is specified as follows (120):

$$\ln\left\{\frac{\pi(Y > j|x_1, x_2, \cdots, x_p)}{\pi(Y \leq j|x_1, x_2, \cdots, x_p)}\right\} = a_j + (\beta_t X_t + \beta_p X_p)$$

Where a_j is the intercept and X_t and β_t are the variable and coefficient of interest, *time*. X_p and β_p represent the covariates and coefficients of the four additional explanatory variables ($p = AD, IR, MS, EV, PR$). Odds ratios (ORs) were used to measure the associations between level of

intensity use of RWD and a set of predictors. As with the binary logistic regression, ORs were used to investigate the change of use of RWD by exposure to the factors.

3.3 Results

In this section, the results of the patten review and intensity analysis are presented, followed by the results of the regression analyses. The patterns of use of RWD were reviewed in three ways – any use, parametric use and non-parametric use of RWD. The intensity of use of RWD was reviewed by highlighting the use in three major components. The sources of RWD used for the major components were also summarised. Then changes in intensity of use in whole sample over time and by type of cancer were explored. From two different regression analyses, the associations between use of RWD/level of intensity of use and the predictors were demonstrated.

3.3.1 Pattern review of the use of RWD

No dominant pattern of use of RWD in economic models was identified in these appraisals. Among identified patterns of the use of RWD (n=111), only fifteen patterns appeared in more than two appraisals, cumulatively 52% of all appraisals (Table 3.4). Identified patterns are presented in the table allocated to the different intensity groups. Patterns of use of RWD observed on a single occasion are not separately identified in the table, but are grouped under *Others*. 16% of included STAs did not use RWD in any part of economic models. The most commonly observed pattern was the pattern, estimating overall survival of intervention and comparators (6% of all patterns). It is followed by a pattern, estimating end-of-life resource use (5% of all patterns). Most patterns identified in more than two appraisals, belonged to the low and medium levels of intensity groups, 33% and 18% respectively, in which RWD are used to inform fewer than two major components.

Table 3.4 Description of patterns of use of real-world data without considering non-parametric/parametric use

Patterns	Number (%)	Intensity group	
		A	B
No use of RWD	37 (16.16%)	G	D’
Estimating overall survival of intervention and comparators	13 (5.68%)	E	C’
Estimating resource use of end-of-life	12 (5.24%)	F	D’
Estimating resource use of end-of-life & health state cost	8 (3.49%)	F	D’
Estimating resource use of health state cost	7 (3.06%)	F	D’
Estimating overall survival of intervention and comparators and resource use of end-of-life & health state cost	6 (2.62%)	D	C’
Estimating overall survival and progress free survival of intervention and comparators and resource use of health state cost	5 (2.18%)	D	C’
Validating survival distribution of intervention and comparators and estimating resource use of end-of-life	5 (2.40%)	F	D’
Estimating overall survival and progress free survival of intervention and comparators	5 (2.40%)	D	C’
Estimating resource use of end-of-life and dose adjustment of intervention and comparators	4 (1.75%)	F	D’
Estimating volume of treatment of intervention and comparators	3 (1.31%)	E	C’
Estimating overall survival of intervention and comparators and resource use of health state cost	3 (1.31%)	D	C’
Validating survival distribution of intervention and comparators	3 (1.31%)	F	D’
Choosing comparators	3 (1.31%)	E	C’
Choosing comparators and estimating resource use of health state cost	3 (1.31%)	D	C’
Others	112 (48.1%)		
Total	229 (100%)		

The non-parametric and parametric use of RWD were separately reviewed. Sixty per cent of all included appraisals made no non-parametric use of RWD (Table 3.5). The commonest pattern of non-parametric use of RWD was to validate the choice of survival distribution for the intervention and comparators (9% of all patterns of non-parametric use), followed by use of RWD for choice of

comparators (6% of all patterns of non-parametric use). Since only one major component, use of RWD for choice of comparators was classified as non-parametric use, the patterns of non-parametric use observed are classified as low intensity use.

Table 3.5 Description of patterns of non-parametric use of real-world data

Pattern	Number (%)	Intensity	
		A	B
No use of RWD	136 (59.39%)	G	D’
Validating survival distribution of intervention and comparators	20 (8.73%)	F	D’
Choice of comparators	14 (6.11%)	E	C’
Validating survival distribution of comparators	13 (5.68%)	F	D’
Treatment sequence	7 (3.06%)	F	D’
Characteristics of population	7 (3.06%)	F	D’
Validating survival distribution of intervention	4 (1.75%)	F	D’
Treatment sequence & validating survival distribution of intervention and comparators	4 (1.75%)	F	D’
Choice of comparator & validating survival distribution of intervention and comparators	3 (1.31%)	D	C’
Choice of comparator & validating survival distribution of comparators & time-to-discontinuation of comparators	2 (0.87%)	D	C’
Treatment sequence & time-to-discontinuation of intervention and comparators	2 (0.87%)	F	D’
Treatment sequence & validating survival distribution of comparators	2 (0.87%)	F	D’
Other*	15 (6.55%)		
Total	229 (100%)		

With respect to parametric use of RWD, parametric use showed more diverse patterns than for non-parametric use. 24% of included appraisals made no use of RWD to inform any parameter in the economic model (Table 3.6). Using RWD for estimating end-of-life resource use was the commonest pattern (10% of all patterns of parametric use of RWD), followed by use of RWD to estimate OS for the intervention and comparators (7% of all patterns of parametric use of RWD). 23% of patterns

identified in more than two appraisals showed a medium level of intensity which includes use of RWD for one major component.

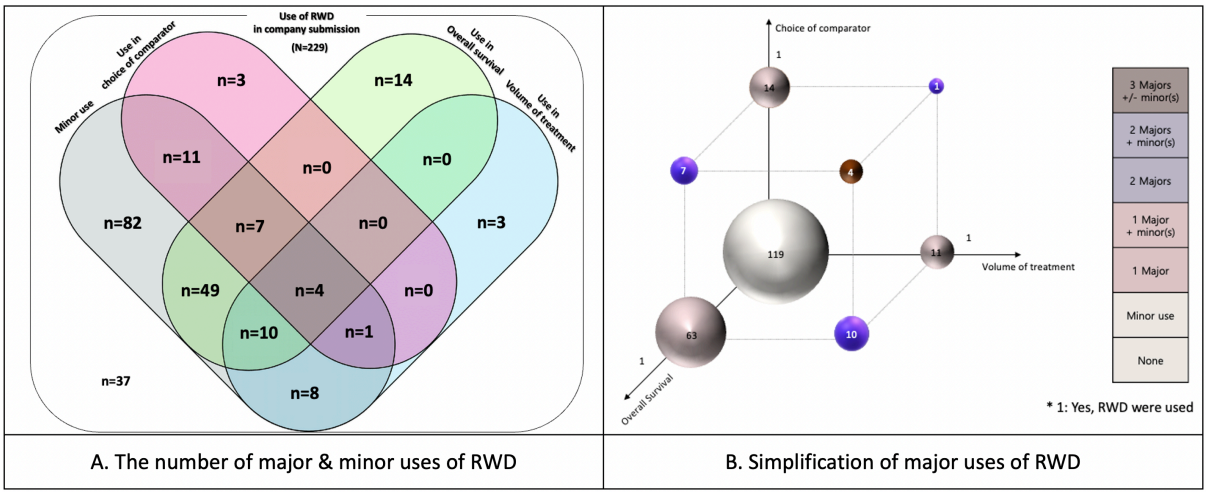
Table 3.6 Description of patterns of parametric use of real-world data

Patterns	Number (%)	Intensity	
		A	B
No use of RWD	55 (24.02%)	G	D’
estimating end-of-life cost	23 (10.04%)	F	D’
estimating OS of intervention and comparators	17 (7.42%)	E	C’
Using RWD for estimating end-of-life & health state cost	14 (6.11%)	F	D’
Using RWD for estimating health state cost	13 (5.68%)	F	D’
estimating OS & PFS of intervention and comparators	10 (4.37%)	D	C’
estimating end-of-life cost & dose adjustment of intervention and comparators	9 (3.93%)	F	D’
Using RWD for estimating OS of intervention and comparators & end-of-life & health state cost	6 (2.62%)	D	C’
Estimating OS & PFS of intervention and comparators & health state cost	6 (2.62%)	D	C’
Estimating volume of treatment of intervention and comparators	4 (1.75%)	E	C’
Estimating OS & PFS of comparators	3 (1.31%)	D	C’
Estimating OS of intervention and comparators & end-of-life cost	3 (1.31%)	D	C’
Estimating OS of intervention and comparators & health-state cost	3 (1.31%)	D	C’
Estimating OS & PFS of intervention and comparators & health state cost & end-of-life cost	3 (1.31%)	D	C’
Other*	60 (26.2%)		
Total	229 (100%)		

3.3.2 Intensity analysis

Use of RWD for three major components

Intensity of use of RWD was analysed in all appraisals included in this study. These appraisals were classified into intensity groups following the two classifications in Figure 3.3. The groups of intensity classification were presented in Venn diagram (Figure 3.6A). Of 229 STAs, 84% used RWD at least once in the economic model. Among the uses of RWD for major components, use for the estimation of OS is the most common, followed by determining the volume of treatment. Four appraisals were identified, which included RWD in all major components. Figure 3.6B shows the use of RWD in simplified classification which focuses the major components. Nearly half of appraisals didn't include any form of RWD with respect to any of the three major components of economic evaluation.



A. The number of major & minor uses of RWD | B. Simplification of major uses of RWD

Figure 3.6 Major uses of RWD

Sources of the real-world data used for the three major components

Figure 3.7 summarises the sources of RWD used for the major components. As data can be differently used for intervention and comparators in estimating OS and volume of treatment, the use of RWD was reviewed separately for the intervention and the comparators. Depending on the components, frequently used sources of RWD were different. With respect to the sources of RWD used for estimating overall survival, data from Office for National Statistics (ONS) were commonly

used while market share data were much more common in other components. Market share data appeared to be frequent sources of RWD for determining relevant comparators in economic models. These data were often collected by the company or international health information private entities such as IQVIA. Three sources of RWD (registry, hospital data, medical chart review) were found across the three major components. On average, 22% of sources of RWD used for major components were registry data. The Surveillance, Epidemiology, and End Results Program (SEER) and Flatiron were frequently used registry data. Hospital data were the sources on 16% of the occasions that RWD were used for major components.

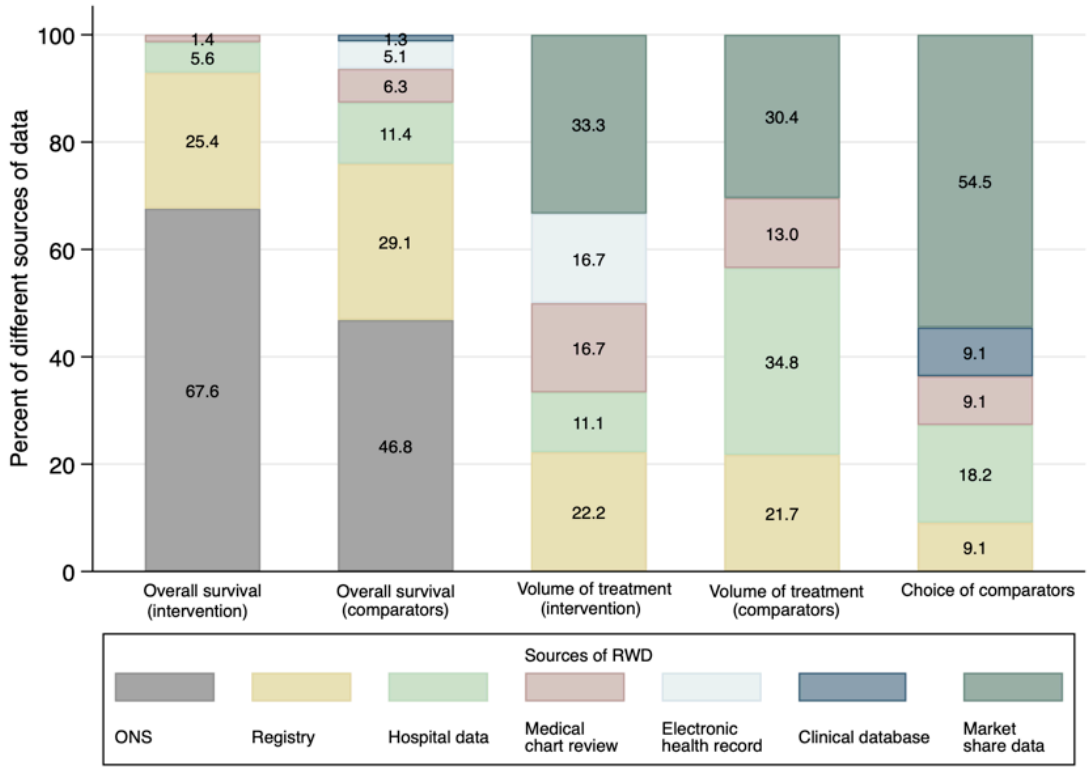
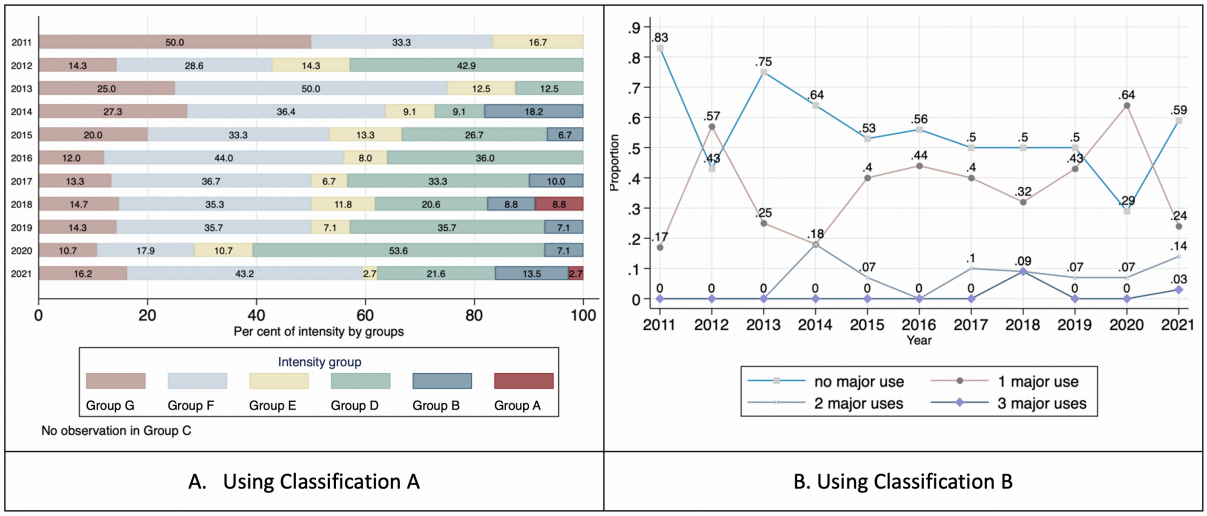


Figure 3.7 Sources of RWD used in three major components

Clinician surveys or expert opinions are often used in economic models when determining the resource use or distribution of subsequent treatments. For example, the appraisal of nivolumab for previously treated unresectable advanced or recurrent oesophageal cancer (NICE TA707) used a clinician survey to determine the frequency of resource use in order to estimate disease management costs (121). However, data from clinician surveys or expert opinion are not classified as RWD following the definition used in this study.

Changes of intensity over time and by type of cancer

Intensity of use of RWD in each year was reviewed using classification A. Over time, the major use of RWD has increased in appraisals (Figure 3.8A). In 2020, about 60% of appraisals made at least two major uses of RWD in the economic models. Cases involving three major uses of RWD were observed in 2018 and 2021. There does not appear to have been a clear change in the intensity of use of RWD using the simpler classification B (Figure 3.8B).



A. Using Classification A | B. Using Classification B

Figure 3.8 Intensity over time

Figure 3.9 shows the intensity of use of RWD by cancer type. In five types of cancer (skin cancer, oesophageal cancer, bone & marrow cancer, lung cancer and prostate cancer), RWD were used to inform more than two major components in small proportion of their appraisals. In STAs of treating skin cancer, RWD were more often used for major components than other types of cancers. Ovarian cancer and head & neck cancer show less use of RWD in economic models; only minor use or no use were observed.

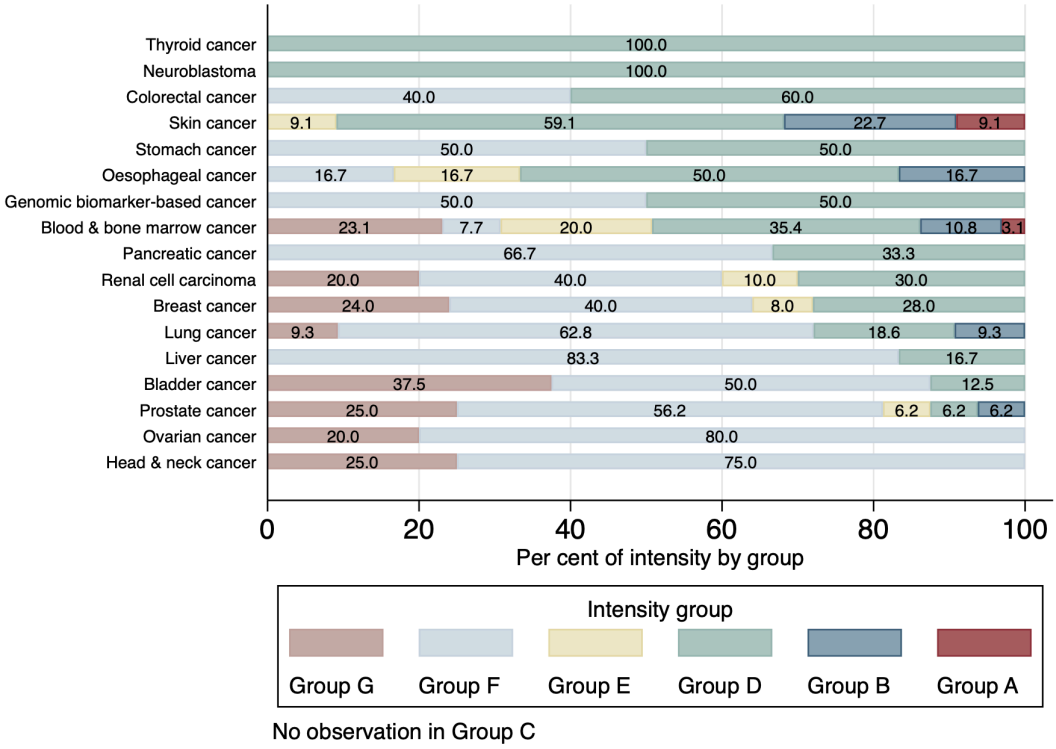


Figure 3.9 Intensity of use of RWD by the cancer type

3.3.3 Regression analysis

Test for multicollinearity of variables

A correlational analysis was conducted to investigate the relationships among the explanatory variables. Since the variables are categorical variables (except for $time$ and IR), Pearson Chi-squares were calculated to detect the correlation between the categorical variables (Table 3.7).

The Chi square value for AD and InV, $\chi^2(6) = 106.270,\ p = 0.000$, indicates there was a strong, positive relationship between the two variables. Internal validity also has positive relationships with external validity ($\chi^2(6) = 23.102,\ p = 0.001$) and maturity of survival data ($\chi^2(6) = 15.244,\ p = 0.018$).

Table 3.7 Correlation matrix of categorical variables (Chi-square values, n=229)

Variables	M	SD	1	2	3	4	5
1. Availability of direct treatment comparison (AD)	1.061	0.825	-				
2. Internal validity (InV)	1.201	1.069	106.270***	-			
3. External validity (EV)	0.790	0.635	9.166	23.102**	-		
4. Previous recommended technology (PR)	0.437	0.497	0.482	2.943	1.456	-	
5. Maturity of survival data (MS)	1.939	0.825	17.441**	15.244*	6.129	0.270	-

* p < 0.05, ** p < 0.01, *** p < 0.001

The correlation between IR and other categorical variables was tested by Kendall's rank correlation (Table 3.8). The strong negative correlations were found between IR and InV ($\tau_b = -0.328, p = 0.000$). Positive correlations were found with AD ($\tau_b = 0.131, p = 0.011$) and MS ($\tau_b = 0.176, p = 0.0006$).

Table 3.8 Correlation matrix (τ_b, n=229)

	Availability of direct treatment comparison	Internal validity	External validity	Previous NICE recommendation	Maturity of survival data	Time
Time	-0.073	0.062	-0.025	0.321***	-0.135**	-
Incidence rate (IR)	0.131*	-0.328***	-0.091	0.094	0.176***	0.024

* p < 0.05, ** p < 0.01, *** p < 0.001

In the tests of multicollinearity, InV had multiple associations with other variables. In order to reduce multicollinearity in the regression, InV was omitted in the analysis. The results of univariate analysis were compared with the results of multivariate analysis to see the differences (Appendix 5.1). While other variables showing correlations with some variables were still included in regression analyses, this model was compared with the regression model without these variables to observe their effects in prediction (Appendix 5.3).

Binary logistic regression

a. Associations with any, non-parametric and parametric use of RWD

A multiple logistic regression analysis was conducted to investigate the association between the use of RWD in economic models and six variables ($time, IR, AD, EV, PR$ and MS). In the models, dummy variables were automatically generated with level 1 in categorical variables as the reference level. The log likelihood ratio chi-square test statistic for the full model A (any use of RWD) $LR\ {\chi^2}_{(9)} = 32.80, p < 0.0001$, indicated that the overall model with all explanatory variables was significant. The model with seven variables including InV is presented in Appendix 5.2.

Table 3.9 presents the logit coefficients, standard errors, and odds ratios for the full model for three different outcome variables (any use, non-parametric use, parametric use). There was no statistical significance between explanatory variables and the non-parametric use of RWD whereas a few statistically significant associations were found in the case of any use and parametric use of RWD. The variable AD has a negative association with any use of RWD for both *some available* ($OR = 0.136, p = 0.002$) and *all available* ($OR = 0.188, p = 0.007$). The association with *parametric* use of RWD aligned with the results of *any use* of RWD. When direct treatment comparison was either *some available* or *all available*, the *parametric use* of RWD was negatively associated ($OR = 0.270;\ p = 0.008, OR = 0.268; p = 0.005$, respectively). For the dummy variable *moderate EV*, $OR = 2.668$ ($p = 0.021$) for any use of RWD, 2.174 ($p = 0.027$) for parametric use of RWD, which indicated that *any* and *parametric* use of RWD were likely to increase than when the level of EV was low. PR had positive associations with *any* use ($OR = 2.673, p = 0.032$) and *parametric* use of RWD ($OR = 2.222, p = 0.035$). The variable MS had a positive association with *any* use of RWD. When survival data were mature, *any* use of RWD is likely to increase ($OR = 5.348, p = 0.002$).

Table 3.9 Results of multivariate binary logistic models

Covariate	Model A Any use of RWD		Model B Non-parametric use		Model C Parametric use	
	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Time						
	0.008 (0.007)	1.008 (0.995,1.021)	0.006 (0.005)	1.006 (0.996,1.016)	0.002 (0.006)	1.002 (0.991,1.013)
Direct treatment comparison *(AD)*						
Not available[a]						
Some available	-1.995** (0.655)	0.136** (0.038, 0.491)	-0.671 (0.369)	0.511 (0.248, 1.053)	-1.309** (0.492)	0.270** (0.103, 0.708)
All available	-1.668** (0.618)	0.189** (0.056, 0.633)	-0.401 (0.340)	0.669 (0.344, 1.303)	-1.318** (0.471)	0.268** (0.106, 0.674)
Incidence rate *(IR)*						
	-0.000 (0.000)	1.000 (1.000, 1.000)	0.000 (0.000)	1.000 (1.000, 1.000)	-0.000 (0.000)	1.000 (1.000, 1.000)
Maturity of survival data *(MS)*						
Extremely immature[a]						
Immature	1.103* (0.490)	3.014* (1.155, 7.867)	-0.079 (0.353)	0.924 (0.462, 1.845)	0.287 (0.413)	1.332 (0.594, 2.990)
Mature	1.677** (0.547)	5.348** (1.830, 15.632)	0.529 (0.347)	1.697 (0.859, 3.353)	0.605 (0.426)	1.832 (0.794, 4.225)
External Validity *(EV)*						
Low risk[a]						
Moderate	0.981* (0.424)	2.668* (1.163, 6.123)	0.081 (0.310)	1.085 (0.590, 1.994)	0.777* (0.351)	2.174* (1.092, 4.329)
Questionable	0.606 (0.682)	1.834 (0.482, 6.975)	-0.215 (0.488)	0.806 (0.310, 2.096)	1.025 (0.632)	2.786 (0.807, 9.621)
Previously recommended *(PR)*						
No[a]						
Yes	0.98* (0.457)	2.673* (1.091, 6.548)	0.260 (0.303)	1.297 (0.717, 2.348)	0.799* (0.378)	2.222* (1.059, 4.664)
Constant	0.761	2.141	-0.865	0.421	1.015	2.760
Observations	229	229	229	229	229	229

*p<0.05, **p<0.01

[a] Reference group

b. Association with use of RWD in single components

The multiple logistic regression analysis was carried out to test the hypotheses regarding use of RWD for individual components of the economic model (Appendix 5.4). The variable, IR did not have any associations with the use of RWD in this research. Results regarding the association between use of RWD in estimation of OS and the explanatory variables is reported in Chapter 4 with more information.

Time

$Time$ had statistical associations with the use of RWD in validating survival distribution for intervention ($OR = 1.015, p = 0.049$), estimating progression-free survival (PFS) for intervention $OR = 1.022, p = 0.017$) and transition probability ($OR = 1.033, p = 0.048$).

Availability of direct treatment comparison

Since several predictors were categorical variables, dummy variables were generated in the logistic regression. $All\ AD$ had a negative association with the use of RWD in estimation of PFS for comparators ($OR = 0.170, p = 0.000$). There was also statistically a significant association between $some\ AD$ and use of RWD in PFS for comparators ($OR = 0.129, p = 0.000$). Within the same variable, statistical significance was partially found in other outcome variables. $All\ AD$ had a negative association with volume of treatment for comparators ($OR = 0.313, p = 0.034$) while no association was found between another dummy predictor variable, $some\ AD$ and this variable.

Maturity of survival data in clinical trials

MS showed both positive and negative associations depending on the outcome variables. In the use of RWD in estimating PFS for comparators, MS was negatively associated (*immature survival data*: $OR = 0.347, p = 0.026$; *mature survival data*: $OR = 0.295, p = 0.016$). However, with a dummy variable, *mature survival data*, positive associations were found in several outcome variables. *Mature survival data* had positive associations with use of RWD in treatment sequence ($OR = 5.291, p = 0.014$), resource use for health state ($OR = 2.583, p = 0.019$), end-of-life ($OR = 2.475, p = 0.014$) and dose adjustment for intervention ($OR = 6.070, p = 0.009$) and comparators ($OR = 4.707, p = 0.014$).

External validity

A statistically significant association with EV was only found in the use of RWD for dose adjustment. $Moderate\ EV$ had positive associations with use of RWD for dose adjustment for both intervention ($OR = 11.046, p = 0.024$) and comparators ($OR = 4.048, p = 0.038$). However, there was no statistical significance between $questionable\ EV$ and the use of RWD for dose adjustment.

Previous technology appraisals recommendation in NICE

The logistic regression results showed that PR had positive associations with the use of RWD in estimating transition probabilities ($OR = 8.544, p = 0.011$) and resource use of end-of-life ($OR = 2.276, p = 0.010$).

Ordinal regression analysis

An ordinal regression analysis (generalised ordered logit model) was conducted to investigate the association between the level of intensity of use of RWD and a set of factors: *time*, AD, EV, MS, IR and PR. To test the main hypothesis, *intensity of use of RWD has increased over time*, a univariate model with *time* as the explanatory variable is fitted first. And then the full model with all variables is fitted. The log likelihood ratio test is used to compare the two models ($\chi^2_{(9)}$ = 53.24, p=0.0000). The result indicated that the full model fitted data better than the univariate model. For the full model, *LR* $\chi^2_{(10)}$ = 56.71, p=0.0000, which indicates that the full model with eight variables from five explanatory variables to test hypotheses provides a better fit than the null model with no independent variables in predicting the ordinal response variable. In the full model, the likelihood ratio R^2_L=0.133, which suggests that the relationship between the outcome variable, intensity of use of RWD, and explanatory variables, is small. The results for both the single-predictor model and the full model are presented in Table 3.10.

Table 3.10 Results of the Generalised Ordered logit Models: Univariate Model and Full model

Covariate	Univariate model		Full model	
	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Model 1 (Y > 1 vs Y≤ 1)				
Time				
	0.008 (0.004)	1.008 (1.000, 1.016)	0.004 (0.005)	1.004 (0.995, 1.015)
Availability of head-to-head comparison				
Not available[a]				
Some available			-1.707*** (0.378)	0.181*** (0.087, 0.380)
All available			-1.467*** (0.340)	0.231*** (0.118, 0.451)
Incidence rate				
			-0.000 (0.000)	1.000 (1.000, 1.000)
Previous recommendation states				
			-0.309 (0.320)	0.734 (0.392,1.377)
Maturity of survival data				
Extremely immature[a]				
Immature			-0.618 (0.341)	0.539 (0.276, 1.051)
Mature			-0.948** (0.348)	0.388** (0.196, 0.767)
External validity				
Low risk[a]				
Moderate risk			0.381 (0.308)	1.464 (0.800, 2.678)
High risk			0.045 (0.477)	1.046 (0.410, 2.667)
Model 2 (Y > 2 vs Y≤ 2)				
Time				
	0.008 (0.004)	1.008 (1.000, 1.016)	0.005 (0.005)	1.004 (0.995, 1.015)
Availability of head-to-head comparison				
Not available[a]				
Some available			-1.707*** (0.378)	0.181*** (0.087, 0.380)
All available			-1.467*** (0.342)	0.231*** (0.118, 0.451)
Incidence rate				
			-0.000 (0.000)	1.000 (1.000, 1.000)
Previous recommendation states				
			0.774 (0.488)	2.619 (0.833, 5.645)
Maturity of survival data				
Extremely immature[a]				
Immature			-0.618 (0.341)	0.539 (0.276, 1.052)
Mature			-0.948** (0.348)	0.388** (0.196, 0.767)

External validity				
Low risk[a]				
Moderate risk			0.381 (0.308)	1.464 (0.800, 2.677)
High risk			0.045 (0.477)	1.046 (0.410, 2.667)
α_1	-0.736 (0.383)	0.479 (0.226,1.015)	1.179 (0.611)	3.251 (0.981,10.774)
α_2	-2.919*** (0.436)	0.054*** (0.023, 0.127)	-1.942** (0.666)	0.143** (0.039, 0.530)
Observations	229		229	
LR R^2	0.008		0.133	
Log likelihood	-211.865		-185.243	
LR χ^2	3.47		56.71	

**p<0.01, ***p<0.001

[a] Reference group

Note: Standard errors are shown in parentheses.

Five explanatory variables, $Time$, $AD\ IR$, MS, EV met the PO assumption in the model. As the PO assumption test was tenable for these variables, the same logit regression coefficients for each variable were found across the binary logit models. In the univariate model, there is no statistical significance ($OR = 1.01, p = 0.07$). In the full model, AD shows the statistical association with lower odds of increased intensity of use of RWD ($some\ AD$: $OR = 0.181, p = 0.000$; $all\ AD$: $OR = 0.231, p = 0.000$). This indicates that the odds of being above a particular level of intensity of use decreased by a factor of 0.20 for a one-unit increase in the variable, AD, when all the other predictors remained constant. Another predictor, *mature survival data* had statistical significance for the odds of decreased intensity of use of RWD ($OR = 0.39, p = 0.007$). The variable, PR that violated the PO assumption showed different odds ratios across two binary models. However, none of them show any statistical associations. The model with seven variables including InV is presented in Appendix 5.6.

3.4 Discussion

This study identified the patterns and the intensity of use of RWD in economic models in NICE appraisals of oncology medicines, rather than a simple counting of the number of times RWD were used. Also, the factors associated with use of RWD/intensity of its use were explored. When looking

at use of RWD over time, it is notable that RWD have been used since the early appraisals. NICE has accepted relevant and available data to assess new technologies systematically and unbiasedly. Regardless of data types, it is standard practice to use the best available evidence in NICE technology appraisals (122). There are a series of NICE decision support unit (DSU) technical support documents (TSDs) to "help improve the quality of analysis, reporting, critical appraisal and interpretation of estimates of treatment effect from non-RCT studies (110,123)." These documents guide the analysis of non-RCT data and observational data more generally. This may show the NICE's longstanding interest in using diverse sources of evidence for drug appraisals.

The most common pattern was no use of RWD. Among patterns showing at least one use of RWD, there was no dominant pattern in this review. This suggests that there is no agreed approach to the use of RWD in STAs. Each appraisal has different issues in the decision making. Differences in the decision context are likely to lead to diverse approaches to the use of data. Also, the availability of data or capacity for data analysis would vary in each appraisal context. Hence, how to use and where to use data can be different, although RWD are available for similar situations. Subsequently, approaches and opportunities to incorporate supplementary data such as RWD might vary across appraisals. Although there was no dominant pattern of non-parametric use, about half of the identified patterns included uses of RWD to validate or corroborate the survival distributions for either the intervention or the comparators. This could be seen as a considerable effort to validate the clinical feasibility of survival models following NICE methods guidance, which the NICE DSU document issued in 2013 proposes that the justification of the extrapolated survival model could be achieved through the use of external data sources (124).

Having documented how RWD have been used, the intensity of use has been gauged by classifying patterns of use in terms of the intensity of use of RWD. This study found that RWD were not often used in estimation of OS, volume of treatment and choice of comparators although RWD were widely used across the economic models. Clinical experts' opinions were often used to justify assumptions

regarding the volume of treatment and the choice of comparators. Expert opinions are helpful when relevant data are unlikely to exist such as for rare cancers. However, inconsistent application and insufficient reports on the use of expert opinions are a concern (125,126). Diverse RWD resources are available in the NHS, to describe the volume of subsequent treatments. For instance, SACT data are used for estimating volume of subsequent treatments in a few Cancer Drugs Fund (CDF) review appraisals. Along with expert opinions, these RWD could provide more systematic information to better describe routine practice. However, the reliability of the RWD can be questioned in some cases. With respect to the use of RWD in the estimation of OS, the nature of RWD such as non-randomised and subject to confounding creates evident problems for the use of RWD. The recently launched NICE real-world evidence framework highlights the use of RWD for estimating comparative treatment effectiveness including building an external control arm. Based on this framework, accumulated experience in use of RWD could incentivise more intensive use of RWD in estimation of OS in future appraisals.

Regression analyses were carried out to investigate the statistical associations. The correlation analysis between variables identified that the *Internal validity (InV)* was collinear with other variables. The information about InV was extracted from 'Quality assessment of the relevant clinical effectiveness evidence' in the Evidence Review Group (ERG) reports. ERGs usually use a tool for risk of bias assessment recommended by NICE when assessing the internal validity of clinical trials (127). The criteria include randomisation, concealment of treatment allocation, selective reporting, and completeness of reporting outcomes. The criteria are, to some extent, related to the explanatory variables used in this study. For example, an appraisal using single-arm clinical trials as the main clinical evidence was reported as having a high risk of bias due to the absence of double blinding and randomisation. While different criteria were usually used to evaluate the internal validity for non-randomised clinical evidence, this study regarded single-arm clinical trial as high risk of bias due to the absence of randomisation. As the InV had the highest correlation, this variable was excluded for the analysis to reduce multicollinearity.

While the highly correlated variable InV was excluded, two variables IR and MS showing some correlations with other variables were included in regression model to avoid missing too many independent variables. This study has highlighted the likelihood rather than the estimates of regression coefficient. The literature indicates that multicollinearity is less likely to influence the predictions direction (128) although it can undermine the regression coefficient. Instead of excluding all correlated variables, excluding only the highly correlated variable can have a benefit of not losing many independent variables. Nonetheless, how the correlated variables affected the prediction needed to be clearly stated. The model including two variables was compared with one without these variables to inspect the impact of the multicollinearity. The results of the univariate regression analyses were also compared. The results showed that they had less impact on the prediction of the regression models.

This study analysed statistical associations between the increased use of RWD and six variables using binary logistic regression. Several results supported the initial hypotheses. The primary variable, $Time$ had an association with the use of RWD for validating survival distribution of intervention. In many appraisals, RWD have been used to support the choice of survival distribution for the clinical outcome. This finding could be a result of the NICE DSU recommendation that external data should be used to assess the clinical plausibility of survival curves (124). While the recommendation of NICE DSU possibly has influenced on the use of RWD for supporting the choice of survival distribution, it is difficult to distinguish its impact from the overall time trend. The variable, PR had positive associations with any use of RWD, parametric use and several uses in individual components (resource use of end-of-life treatment, estimating transition probability). This hypothesis assumed that the data from routine clinical practice would be more likely to be available if the technology was already recommended for routine commissioning or use within CDF for other indications. When reviewing RWD used in estimating transition probability and resource for end-of-life treatment, previous guidance or literature were frequently used to in the appraisals.

In some explanatory variables, the partial relationships were found. Since most variables are categorical variables, dummy variables were generated in the logistic regression. Within the same variables, statistical significance was partially found in some dummy variables. With respect to the variable AD, both dummy variables have negative associations with *any use* of RWD and *parametric use* of RWD. However, two individual uses of RWD for estimating volume of treatment for comparators have negative associations only with $all\ AD.$ A possible explanation might be that the difference between $some\ AD$ and the reference category would not be substantial due to frequent uses of experts' opinions. Clinical experts' opinions are often used for making assumptions about volume of treatment. Compared to the reference group ($no\ AD$), $some\ AD$ could involve less use of external information, including experts' opinions. However, in this study, such opinions are not regarded as RWD. Therefore, differences are less readily observed.

Between the dummy variables in the variable EV, partial associations were found. The statistical associations with the $moderate\ EV$ in several variables (*any use*, *parametric use* and use of RWD in dose adjustment of intervention and comparators) were found whereas the associations between increased use of RWD and $questionable\ EV$ were not found in these variables. The reasons of this finding might lay on the limited number of observations in $questionable\ EV$. It was found in only 12 % of included appraisals. $Moderate\ EV$ was recorded in more than half of STAs. This study tried to break down the levels of EV. However, it was challenging due to the language differences across the appraisals. The information about the external validity of clinical trials was extracted from ERG reports. Several different ERGs participate in the NICE appraisal process. Although these external academic organisations strictly follow the guidance on the review and quality of evidence and challenges are covered in ERG critiques, variation in the focus areas can be found (129). The assessment of generalisability, which is a different word of external validity was one of the areas where the differences are often found. The critiques on uncertainty around generalisability made by the ERGs might not be comparable due to the various language use and perspective from each ERG report across appraisals. This is one of the limitations of this study.

Some associations were completely opposite to those hypothesised. When survival data were mature, less use of RWD in economic modelling was expected because RCTs could provide more information for a long period. However, the opposite associations were found. The logistic regression model found that MS was positively associated with several variables such as any use of RWD, use of RWD in treatment sequence and resource use of health state cost, end-of-life and dose adjustment for intervention and comparators. One possible reason for these associations might be different approaches toward survival outcome and resource use. For treatment effects, RWD for clinical effectiveness was cautiously used due to potential biases. Also, since trial data are preferred, other sources of data are less used to mitigate the problem in predicting the treatment effect if survival data from clinical data were mature. On the other hand, more diverse ranges of evidence were included for resource use. The evidence quantifying the effect of the technology on resource use such as days in hospital or visit to a GP in practice was required. Compared with survival outcome, RWD for such information were often more available. Previous literature collecting data from routine practice alongside clinical trials was commonly used for resource use. When the survival data are mature, routine data about resource use could have a greater chance of being incorporated in economic models to predict cost-effectiveness in routine clinical practice.

In the binary logistic regression, the expected association with incidence rate was not found. It was anticipated that economic models of rare cancers used more RWD due to the limitation of conducting RCTs (130). However, the results of this study did not statistically support the hypothesis. The stakeholder interviews potentially can provide some insight into this result. During the interviews, the rareness of a disease was identified as a negative factor in the collection of meaningful RWD for appraisals. The number of patients in registries of rare cancers was not usually enough for drug appraisals. Also, a large proportion of the rare cancer patients might already be included in the clinical trials of treatments. Thus, rarity may be associated with greater reliance on clinical trial data rather than registry data.

The ordinal regression was conducted to explore the associations between level of intensity of using RWD and other explanatory variables. In both univariate and full models, the variable $time$ did not have a statistical association with level of intensity. This study tried to review the impacts of these published DSU TSDs (110,123) on the level of intensity of use of RWD. However, it was challenging to evaluate the association due to time lags to implement the methodology in appraisal practice. The results for the full model indicated that higher level of intensity use was negatively associated with AD. The outcome variables, OS and volume of treatment are essential information in the economic model for drug appraisal. This information can be sensitive to the availability of direct treatment comparison compared to other variables. Alternative data sources are required when clinical trials cannot provide suitable estimates. This result indicates where potentially RWD could be useful to answer the questions and how to prepare study design of RWD to answer the question. Given increased drug regulatory approval based on single arm trials (131), there is scope for increased use of RWD in OS and volume of treatment. Eventually the level of intensity of using RWD is expected to be increased.

In the oncology appraisals reviewed, several limitations to the use of RWD in HTA decision making were addressed. First, RWD cannot provide full information regarding the new intervention. Data for the new technology are not available at the point of appraisal as RWD can only be collected after product launch. Another limitation is that the RWD population is unlikely to be perfectly matched to the target population of appraisal. RWD are expected to provide information reflecting routine practice. However, depending on the timing and context of data collection, RWD may not properly capture routine practice. Due to changes in clinical practice or the small number of patients recruited, the sample population in RWD is potentially different from the whole population. In the STA of darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer (NICETA660), RWD were used for the estimates of health care resource use for the target population. The ERG was concerned that the study population was recruited over a wide time interval, which may have seen substantial changes in clinical practice, and that the target

sample was too small to understand the clinical benefits since the primary outcome was obtained from 44 patients diagnosed with the specific indication. This example shows that RWD do not always reflect current practice. How and what information was collected to answer the question is critical to the use of RWD in STAs.

The study also reviewed the use of RWD in sensitivity analyses, as distinct from the base case analysis. Remaining uncertainty around parameters is usually explored in sensitivity analyses using the alternative evidence which is reviewed but not used in the base case analysis (133). The use of RWD as supplementary data in sensitivity analyses was expected to more common than in the base case. However, it was found that parametric use of RWD in sensitivity analysis was made in only a few appraisals. The interview with key players in NICE appraisals helps draw an implication (Appendix 2). The manufacturer is less likely to present the results of analysis of RWD in the sensitivity analysis if the data hardly provide the additional benefit in appraisals. Processing RWD can require significant resources in terms of collection and analysis. If there is no absolute motivation to use RWD in sensitivity analysis, manufacturers prefer to use other published RCTs to explore the uncertainty in their model inputs and in the survival distribution.

Despite my best efforts, this study has several limitations. The first limitation is that the information about use of RWD was extracted from the company submission. The extracted data did not reflect the committee's preferences. The data in company submissions are primary evidence in a STA process. However, the Appraisal Committee do not necessarily accept the data presented by company. Their preferences regarding RWD could be different from those of the company. A study reviewing the Appraisal Committee's preferences regarding RWD is in progress, may be able to provide a more comprehensive understanding of previous use of RWD in NICE STAs. Another limitation of this study is that intensity of use can be defined differently depending one's perspective. In this study, intensity has been defined in terms of whether use of RWD was made for particular purposes, and thus is judged more by the overall pattern of use in an appraisal. There could be

different ways to define and measure the intensity of use of RWD in appraisals. Also, regarding the predictors of regressions, there can be other factors which are potentially associated with use of RWD. Although the predictors in the logistic regression covered most situations where additional data are required or where additional data are available in economic models, other factors might be influential in the use of RWD. For instance, companies can be incentivised to use more RWD due to their market access strategy. Also, some manufacturers could be more confident than others to use more RWD in appraisals. The regression only assumed that the identified variables were associated with use of RWD or the level of intensity of use.

This study is the first study to extract explicitly and systematically information on the use of RWD and to assess associations between use of RWD and the level of intensity of use and a range of factors in economic models in NICE STAs of oncology medicines. This study is one of the first attempts to look at the factors associated with increased use of RWD in economic model of NICE STAs. Most previous studies investigated the use of RWD in limited appraisals without clear operation rules or tried to identify its use from stakeholder interviews. Beyond simple description, the statistical analysis to identify associations with increased use of RWD can provide a clear picture on where and why RWD have been used. Understanding previous use of RWD in NICE STAs can provide a crucial perspective on how to organise the use of RWD in the future. It helps identify circumstances where RWD could be more actively used based on past observation. It can usefully inform future evidence generation strategies for use of RWD.

3.5 Conclusion

NICE has had a long-standing interest in the use of RWD in STAs. RWD have been widely used in economic models. Nonetheless, uses of RWD have mostly been minor. When randomised controlled trials failed to provide a comprehensive picture of the drug, due to absence of direct treatment comparison, issues concerning the external validity of trials and the maturity of survival data in

clinical trials, RWD were more likely to be used in economic modelling of oncology medicines. These results based on the systematic review of previous appraisals suggest that RWD were used in diverse situations and its uses were associated with data gaps in the economic modelling.

Chapter 4. Maturity of overall survival data and use of real-world data

The previous chapter has described the use of real-world data (RWD) in three ways: non-parametric, parametric and any use of RWD regardless of the type of use. Given the interest in using RWD for comparative treatment effects, this chapter focuses on the maturity of the data on overall survival (OS) in economic modelling and the use of RWD for estimating OS for intervention and comparators. This chapter highlights how the maturity of OS data has changed and how its maturity is associated with the use of RWD for estimating OS. Although this paper is not published yet, it has a paper-style structure, which can be read alone. Hence, repetition is found in this chapter with respect to the factors used in testing the hypotheses.

This chapter discusses the maturity of OS data in economic modelling and the use of RWD for estimating OS. The first part of the chapter starts with an introduction to issues of using immature survival data in decision-making. The maturity of OS data is reviewed: the change of maturity over time and by type of cancer, followed by regression analysis. The ordinal proportional model is estimated to determine the association between maturity and time/introduction of 2016 CDF. The second part presents the use of RWD for estimating OS, reviewed in several ways: change over time, by type of cancer, how to use RWD for estimating OS and which sources of RWD are used for estimating OS. The last part of this chapter describes the association between the use of RWD for estimating OS and the maturity of data on OS using a binary logistic model. This chapter contributes to understanding the use of RWD for estimating OS, where considerable research interest has been imposed.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	157809	**Title**	Ms
First Name(s)	Jiyeon		
Surname/Family Name	Kang		
Thesis Title	Improving economic evaluation and decision-making for oncology drugs using real-world data		
Primary Supervisor	John Cairns		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	Jiyeon Kang, John Cairns
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I had primary responsibility for preparing the data, developing the model, producing the results and drafting the paper.

SECTION E

Student Signature	Jiyeon Kang
Date	7 November 2022

Supervisor Signature	
Date	

Research paper 3

Using regression to explore associations between the use of real-world data and the maturity of survival data in single technology appraisals of National Institute for Health and Care Excellence

Jiyeon Kang[1], John Cairns[1]

[1]Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

[2]Centre for Cancer Biomarkers (CCBIO), University of Bergen, Bergen, Norway

Jiyeon Kang: ORCID: http://orcid.org/0000-0002-1587-9674

John Cairns: ORCID: http://orcid.org/0000-0001-6442-0440

Corresponding author

Jiyeon Kang, PharmD, MSc.

Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Address: 15-17 Tavistock place, London, WC1H 9SH, UK

Email: Jiyeon.Kang@lshtm.ac.uk

Abstract

Introduction Overall survival (OS) is a key driver of cost-effectiveness in appraising cancer drugs. As a common source of uncertainty, immature survival data are a challenge to the assessment of long-term clinical benefit. Additional data such as real-world data (RWD) are expected to help supplement projection of OS in the economic model. This study aims to investigate the immaturity of survival data and the use of RWD in estimating OS in appraisals of the National Institute for Health and Care Excellence (NICE). A particular focus is the association between the level of maturity of the data on OS and whether the appraisal uses RWD.

Methods Data were extracted for oncology medicines, for which single technology appraisal (STA) guidance was issued between January 2011 and December 2021, using a specially developed protocol. The maturity of the data on OS was classified into three groups, *extremely immature, immature* and *mature* using the percentage of death events in the primary clinical trial. The use of RWD to estimate OS in economic modelling was categorised as *used* or *not used*. Changes in maturity and uses of RWD over time were reviewed. Also, the use of RWD was explored according to different purposes of use. Lastly, binary logistic regression identified the association between the use of RWD to estimate OS and the maturity of the data on OS.

Results About 70% of the included oncology appraisals (n=229) used extremely immature or immature survival data. Although not statistically significant, appraisals with extremely immature survival data were more likely to be recommended within the CDF than appraisals with mature survival data. Immature data were more likely to be used in economic models after the 2016 CDF was introduced (OR=0.49, p=0.013). The main reasons for using RWD on OS were to adjust the background mortality and to extrapolate survival after the final data-cut in the intervention arm. Two negative associations with the increased use of RWD in estimating OS were found (mature survival data: OR=0.370, p=0.000 for intervention; OR=0.222, p=0.000 for comparators, availability of direct

treatment comparison between treatment of interest and comparators: OR=0.269, p=0.000 for comparators).

Conclusion The immaturity of survival data in these appraisals appears to be increasing over time. After the 2016 CDF, immature survival data are more likely to be used in appraisals. A negative association was found between the use of RWD for estimating OS and the maturity of survival data. Absence of direct treatment comparisons between the intervention and comparators was associated with greater use of RWD for estimating OS. While RWD was used to adjust the background hazard and to extrapolate survival curves beyond the trial period, only limited use was made of RWD to estimate comparative treatment effects. RWD can be useful to reduce the uncertainty arising from immature survival data in appraisals of oncology medicines.

4.1 Introduction

The essence of health technology assessment (HTA) is to inform decision-makers about health technology using scientific and systematic evidence (3). The National Institute for Health and Care Excellence (NICE) assesses the value of new health technologies and advises on improving health outcomes within National Health Service (NHS) in England and Wales (7). In NICE appraisals, cost-effectiveness is assessed to maximise heath gain from available resources by estimating costs of the interventions with regard to expected health benefits (134). As an important determinant of cost-effectiveness and clinical effectiveness, health benefits of treatment need to be assessed appropriately. Among several types of health benefits, overall survival (OS) is the main clinical outcome in oncology medicine appraisals.

The key issue for HTA is whether the survival time is sufficient to assess the differential survival, quality adjusted life years (QALYs), cost and hence cost-effectiveness of the intervention. In order to evaluate the differences, in economic models, costs and health benefits are compared over the long term, which is beyond the follow-up periods of trials. Since it is usually not feasible to observe all the

death events in a clinical trial, extrapolation based on observed data is necessary to estimate the long-term effects. Extrapolation is critical in drug appraisals as the magnitude of the change in survival can substantially change the results of economic evaluation. However, finding appropriate survival extrapolation is often challenging due to wide variations between the survival distribution used (135). When survival data are immature – too short-term to support successful observation and analysis of events related to survival – it increases uncertainty in the extrapolation (136,137). NICE is aware of the pitfalls associated with having immature data; that is, treatment effects could be over- or underestimated depending on which survival distributions are selected for the intervention and comparator(s) (138). Tai et al. assessed the risk of using immature data in NICE appraisals, concluding that if long-term trial data were used, a different decision might be made (21).

There is growing interest in how real-world data (RWD) might be used to evaluate healthcare interventions. The potential for RWD to support decisions made by drug regulators and HTA organisations has been highlighted (139). The drug regulators have made the case for using RWD to support their decision-making (95,140) and have approved drugs on the basis of RWD (141). NICE has also highlighted RWD as useful data for health technology appraisals. For example, methods for obtaining comparative treatment effects or extrapolating the survival curve from RWD have been studied (142–144). The NICE *Real-World Evidence Framework* documents opportunities for RWD to reduce uncertainty and for best practice for RWD studies (32). One of the areas where RWD can be useful is in response to the challenges of limited follow-up in clinical trials. Limited follow-up makes it harder to assess the long-term benefits of treatments. This is one of the frequently observed uncertainties in cancer appraisals. RWD can reduce uncertainty by providing information on baseline event rates and changes in disease hazards.

The attempts to use RWD to examine survival outcomes seem new in NICE appraisals. However, RWD such as registry data have already been used to inform the extrapolation of survival estimates when the clinical trial survival data were immature (124). Despite increased awareness of the

immature survival data problem, few studies have systematically reviewed the maturity of survival data in NICE appraisals, mostly using case studies (145). Also, it is unknown how the use of RWD in technology appraisals varies relative to the maturity of survival data in clinical trials. Therefore, this study examines the maturity of the survival data used in economic models in NICE appraisals of oncology medicine and whether the maturity of survival data is associated with increased or decreased use of RWD when estimating OS.

4.2 Methods

This study reviews the maturity of the survival data and the use of RWD to estimate OS in economic models in STAs. All the data required for this study were extracted from NICE Single Technology Appraisals (STAs) of oncology medicines (n=229) reported from January 2011 to December 2021. A data extraction protocol (117) was followed so that comprehensive information about the characteristics of the clinical trials and the use of RWD for economic evaluations in the STAs could be extracted. The extracted data included information about the number (and %) of death events reported in the main clinical trials used in the HTA, and the use of RWD for estimating OS in economic models. Although information about RWD can be extracted for both base-case and sensitivity analyses, this study focused only on the base-case analysis.

The maturity of the survival data was categorised according to the percentage of patients recorded as dying in the primary clinical trial. As there is no consensus on how to define 'immature' survival data, a criterion used in other literature was adapted. Tai et al. reviewed statements about the maturity of survival data in evidence review group (ERG) reports (21) and documented that most immature STAs had under 50% of death events. As this figure groups together trials displaying quite different levels of maturity, an additional cut-point of 20 % of death events has been used here to categorise further the maturity of survival data (Table 4.1). In order to understand the impact of using different maturity criteria, a secondary criterion for maturity was used for the sensitivity

analysis in this study. In Tai et al., data with more than 70% deaths were regarded as mature data; therefore, here, the secondary criterion of maturity for the sensitivity analysis used three points – 20%, 50% and 70%, which yield four different categories of maturity.

Table 4.1 Survival maturity classifications

Primary criterion		**Secondary criterion**	
Classification	Proportion of death events (p, %)	Classification	Proportion of death events (p, %)
Extremely immature	p < 20%	Extremely immature	p < 20%
Immature	20% ≤ p ≤ 50%	Immature	20% ≤ p ≤ 50%
Mature	p > 50%	Relatively mature	50% < p ≤ 70%
		Mature	p > 70%

The information about maturity was obtained from the manufacturer's submission or the ERG report. If the number of death events was redacted, first, the published papers of original research were checked for this information; then, the manufacturer submissions and ERG reports were checked for statements that could potentially indicate immaturity, such as 'survival data are immature'. In the event none of this information was available or the classification was unclear, the survival data were assumed to be immature.

While immature survival data are known as a common source of uncertainty (146), the frequency with which immature survival data are used in NICE appraisals has not been documented. The maturity of survival data was reviewed in three ways: changes over time and incidence by type of cancer and in the light of the appraisal recommendations. Ordinal regression was conducted to investigate whether maturity of survival data was associated with the month in which the FAD was published. The different units – quarterly and annually were used for running additional regression models (Appendix 5.7). The impact of Cancer Drugs Fund (CDF, from here 2016 CDF) on maturity of survival data was also examined as a covariate in the regression model. The 2016 CDF offers scope for conditional approval of a drug which cannot be recommended for routine commissioning due to uncertainties (28). The number of appraisals bearing uncertainty such as immature survival data can

be increasing. The hypothesis for this regression was: *Maturity of survival data is likely to decrease over time and after introduction of the 2016 CDF*. An ordinal logistic regression was estimated as the outcome variable, *maturity of survival data* has three levels. As the variable *after introduction of the 2016 CDF* failed the Brant test, generalised ordered logit model was fitted in this analysis. Two control variables, stage of cancer (*stage*), and targeted cancer therapy (*target*) were included in this analysis to test whether there was an intercept shift and a change in the time trend associated with the introduction of the 2016 CDF. The variable *stage* is a categorical variable that classify the clinical cancer cancer using description in indication. *Target* is a binary variable, whether the technology of interest is targeted therapy or not. While *target* is available from the extracted data, *stage* was newly extracted for this study. How the stage of cancer was categorised in this study is presented in Appendix 1.2. Two different equations were used in multivariate ordinal logistic regression analysis:

$$Maturity = \alpha + \beta_{time}X_{time} \quad \text{(A)}$$

$$Maturity = \alpha + \beta_{tme}X_{time} + \beta_{CDF}X_{CDF} \quad \text{(B)}$$

$$Maturity = \alpha + \beta_{time}X_{time} + \beta_{CDF}X_{CDF} + \beta_{stage}X_{stage} + \beta_{target}X_{target} \quad \text{(C)}$$

(CDF = after introduction of the 2016 CDF, stage = stage of cancer, target =targeted cancer therapy)

The appraisal of azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts (TA399) was regarded as the first appraisal included in post 2016 CDF. While the formal start date for the revised CDF was 1 April 2016, none of three cancer appraisals issued around April 2016 (TA391, TA395, TA396) mentioned the new CDF. TA399 was the first appraisal that discussed and concluded it could not be recommended for CDF. Hence, any STAs issued from TA399 onward were regarded as 'post 2016 CDF'.

Information about the use of RWD to estimate OS was extracted for the intervention and for the comparators as they could come from different data sources. The use of RWD for estimating OS was distinguished between the use to adjust OS for age-related or sex-related mortality, background disease hazard, treatment effect estimates and survival curve extrapolation. The sources of RWD used for estimating OS were also reviewed. The sources of RWD identified in this study were

classified based on the common data sources presented in the NICE real-world evidence framework. These include electronic health records, registries, retrospective chart reviews and other administrative data on health care services (32).

A binary multivariate logistic regression was used to test the hypothesis: *Immature survival data are associated with greater use of RWD for estimating OS*, because the outcome variable was 'use of RWD or not' and the primary predictor was the maturity of survival data (MS). This was extended to a multiple logistic regression model that included four additional predictors: external validity (EV), availability of direct treatment comparison (AD), incidence rate (IR), and previously recommended technology by NICE (PR), which were used in previous Chapter 3. These predictors are related to the sources of uncertainty or availability of data. The use of RWD was separately reviewed for intervention and comparator. All but IR were categorical variables and dummy coding was required. The logistic regression analysis was performed using the logistic procedure in STATA Version 17. Odds ratios (ORs) were used to measure the association between the outcome variable and explanatory variables. ORs compared the relative odds of the use of RWD occurring in the reference level given the exposure to the explanatory variables (MS, EV, AD, IR, PR). The regression coefficients, which is the log form of odds were also presented in the results.

4.3 Results

In the Results section, the incidence of maturity of survival data was reviewed over time and by type of cancer. Appraisal recommendations by maturity of survival and how maturity of survival data has changed in post-2016 CDF were also reviewed. The ordinal logistic regression was run to investigate that the maturity of survival data is associated with time. After the reviewing the maturity of survival data in appraisals, the use of RWD for estimating OS in economic models was examined in several ways including the change over time and by types of cancer, where to use RWD for estimation and sources of RWD. Finally, the results of regression analysis testing the association between use of RWD and the predictors including maturity were presented.

4.1.1 Maturity of survival data in NICE Single Technology Appraisals

STAs were assigned a maturity classification according to the author-specified classification outlined in Table 4.1 above. In the economic evaluations, 36.5% of reviewed STAs (n=76) used extremely immature survival data, while 33.2% (n=69) were based on immature data. Only 30.3% (n=63) used mature survival data. The proportion and average of death events in the STA clinical studies were calculated for each year and reviewed to understand how the maturity of survival data has changed over time (Figure 4.1). The left-hand axis referred to the % of patient with mortality observed within the follow-up period and the right-hand side was the average of these across the published appraisals within a given year. As seen in Figure 4.1, the average number of observed death events generally decreased over time, with two sharp drops between 2012 and 2013 and between 2016 and 2017 (14.9%, 12.8% respectively). Note that the exact proportion of death events was not reported for 23.1% of included STAs (n=53).

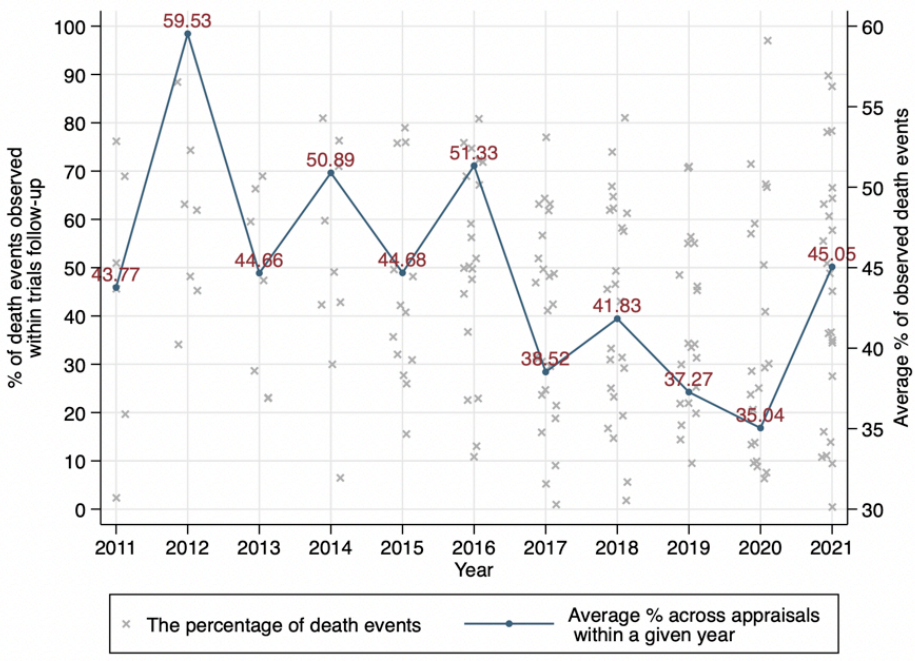


Figure 4.1 Summary of the percentage of death events in appraisals: 2011 - 2021

The maturity of the survival data used in the economic models in these STAs was reviewed by type of cancer (Figure 4.2). More than half of blood and bone marrow cancer appraisals (58.1%) and skin cancer appraisals (52.4%) used extremely immature survival data in the economic evaluations. For

lung cancer, 13.2% of appraisals used extremely immature survival data. Mature survival data were more likely to be used in bladder cancer appraisals (85.7%). All the oesophageal cancer appraisals (n=4) used mature survival data. In contrast, both genomic biomarker-based cancer appraisals (entrectinib, larotrectinib) used extremely immature survival data.

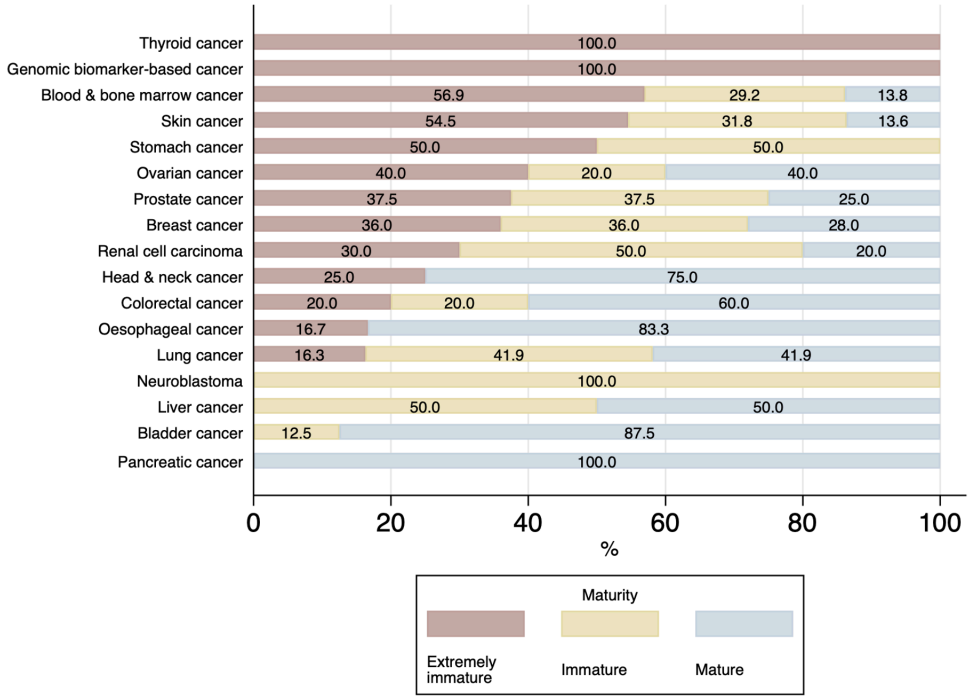


Figure 4.2 Proportion of maturity groups for each type of cancer

Table 4.2 presents the number and proportion of STA recommendations by survival data maturity. The proportion of appraisals based on mature survival data that were not recommended by NICE committees was similar for both time periods (61.1% and 62.5%, respectively). In appraisals recommended through routine commissioning pre- and post-2016 CDF, around 75% used extremely immature or immature survival data for their clinical evidence. Appraisals with extremely immature survival data were more likely to be recommended within the CDF than were appraisals with mature survival data. The statistical difference between recommendations and maturity groups was tested. The difference for post-2016 CDF was statistically significant (χ^2=18.3, p=0.02), whereas there was no statistical difference for pre-2016 CDF appraisals.

Table 4.2 Technology appraisal recommendations by survival data maturity

	Extremely immature (n=85)	Immature (n=73)	Mature (n=71)	Total (n=229)
Pre-CDF 2016				
Not recommended	1 (5.6%)	6 (33.3%)	11 (61.1%)	18 (100%)
Recommended	6 (20.7%)	16 (55.2%)	7 (24.1%)	29 (100%)
Optimised	2 (40.0%)	2 (40.0%)	1 (20.0%)	5 (100%)
Post-CDF 2016				
Not recommended	5 (27.8%)	2 (11.1%)	11 (61.1%)	18 (100%)
Recommended	33 (45.8%)	23 (31.9%)	16 (22.1%)	77 (100%)
Optimised	13 (31.7%)	11 (26.8%)	17 (41.5%)	41 (100%)
CDF	20 (57.1%)	9 (25.7%)	6 (17.1%)	35 (100%)
CDF Optimised	5 (45.5%)	4 (36.4%)	2 (18.2%)	11 (100%)

Figure 4.3 shows the technology appraisal recommendations pre- and post-2016 CDF. The proportion of the appraisals using mature survival data was 6% lower post the introduction of the 2016 CDF, while the proportion using extremely immature survival data was 2.5 times higher in the post-2016 CDF appraisals than in pre-2016 CDF appraisals.

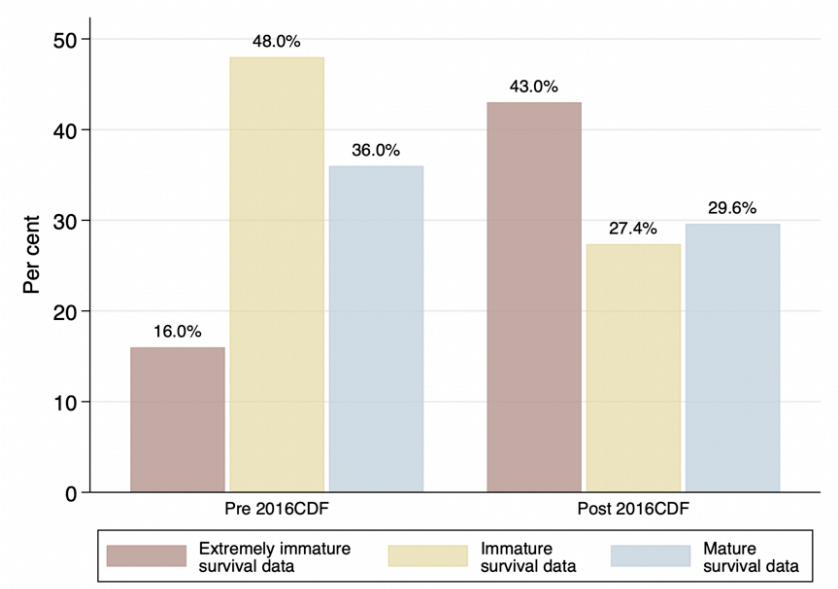


Figure 4.3 Maturity of survival data: Pre- and post-Cancer Drugs Fund 2016

Ordinal regression analysis

In order to investigate the change of maturity of survival data over time and the association between maturity of survival data and after introduction of the 2016 CDF, an ordinal logistic regression (generalised ordered logit model) was carried out. Table 4.3 shows the results of ordinal logistic regression analysis. The regression results with the variable, *time* is presented in Regression A. It showed that maturity of OS data had a negative association with *time* (OR=0.989, p=0.006). However, this statistical association disappeared in multivariate regression models. Two regression models are presented: Regression B excludes the control variables; *stage* and *target*, Regression C includes them. As the Regression B nested within the Regression C ($\chi^2 = 22.79, p = 0.000$), the later regression provided a better fit. The regression models showed that there was an intercept shift and a change in the time trend after introduction of the 2016 CDF. However, the statistical associations with time were found in none of the regression models. While *stage* was the control variable to investigate the intercept shift, *stage* had an association with increased maturity of survival data (advanced stage: $OR = 2.027;\ p = 0.024$, metastatic stage: $OR = 3.555;\ p = 0.000$). That is, more advanced stage of cancer is associated with use of mature survival data in economic modelling.

Table 4.3 Results of multivariate regression (generalised ordered logit model)

	Regression A (*Time*)		Regression B (*Time* and *CDF*)		Regression C (*Time, CDF, stage, target*)	
Covariate	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Model 1 (Y > 1 vs Y≤ 1)						
Time						
	-0.011** (0.006)	0.989** (0.982, 0.997)	-0.008 (0.006)	0.992 (0.980, 1.004)	-0.008 (0.006)	0.992 (0.980, 1.005)
After introduction of the CDF						
Before[a]						
After	-	-	-0.915 (0.542)	0.401 (0.138, 1.160)	-0.854 (0.557)	0.426 (0.143, 1.267)
Stage of cancer						
Early stage[a]						
Advanced	-	-	-	-	0.710* (0.312)	2.027* (1.095, 3.752)
Metastatic	-	-	-	-	1.271*** (0.328)	3.555*** (1.860, 6.794)
Targeted cancer therapy						
	-	-	-	-	0.295 (0.302)	1.342 (1.230, 9.048)

Model 2 (Y > 2 vs Y≤ 2)						
Time						
	-0.011** (0.006)	0.989** (0.982, 0.997)	-0.008 (0.006)	0.992 (0.980, 1.004)	-0.008 (0.006)	0.992 (0.980, 1.004)
After introduction of the CDF						
Before[a]						
After	-	-	0.178 (0.491)	1.195 (0.456, 3.130)	0.453 (0.510)	1.573 (0.579, 4.272)
Stage of cancer						
Early stage[a]						
Advanced	-	-	-	-	0.710* (0.312)	2.027* (1.095, 3.752)
Metastatic	-	-	-	-	1.271*** (0.328)	3.555*** (1.860, 6.794)
Targeted cancer therapy						
	-	-	-	-	-0.525 (0.323)	0.591 (0.314, 1.114)
α_1	1.460*** (0.368)	4.304*** (0.982, 0.997)	2.006*** (0.470)	7.437*** (2.959,18.689)	1.205* (0.509)	3.336* (1.230, 9.048)
α_2	0.095 (0.354)	1.100 (0.550, 2.202)	-0.241 (0.389)	0.786 (0.366, 1.686)	-1.016* (0.468)	0.362* (0.145, 0.907)
Observations	229		229		229	
LR R^2	0.015		0.032		0.078	
Log likelihood	-246.991		-242.745		-231.351	
LR χ^2	7.70**		16.20**		38.98***	

**p<0.05, ***p<0.01, [a] Reference group
CI: confidence interval, CDF: Cancer Drugs Fund

4.1.2 Use of real-world data for estimating overall survival

Figure 4.4 shows the proportion of STAs that used RWD to estimate OS for the intervention and comparator groups in the base case analysis. Overall, the use of RWD for estimating OS has increased. This is generally higher for the comparators than for the interventions. A marked drop in the use of RWD for both groups was observed in 2019. One year later, the use of RWD in the estimation of OS rose again, to the highest proportion observed from 2011 to 2021 (intervention, 48%; comparators, 50%).

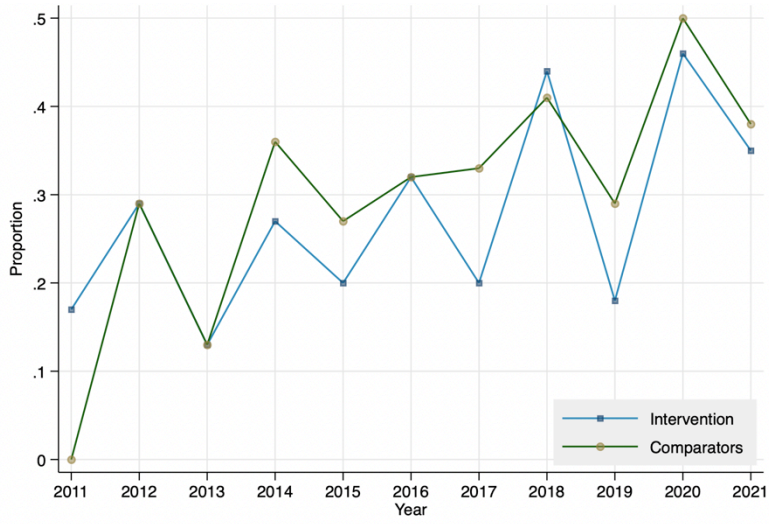


Figure 4.4 Annual proportion of appraisals using real-world data to estimate overall survival: 2011-2021

The STAs using RWD for estimating OS were reviewed by type of cancer (Figure 4.5). Except for neuroblastoma and thyroid cancer, where only one STA was issued, skin cancer showed the highest proportion of use of RWD to estimate OS. RWD were used for intervention and comparator groups in 90% of appraisals. Half of blood and bone marrow cancer appraisals also used RWD (intervention 48.4%, comparators 54.8%). Breast cancer and lung cancer appraisals, which are relatively numerous, do not often use RWD to estimate OS.

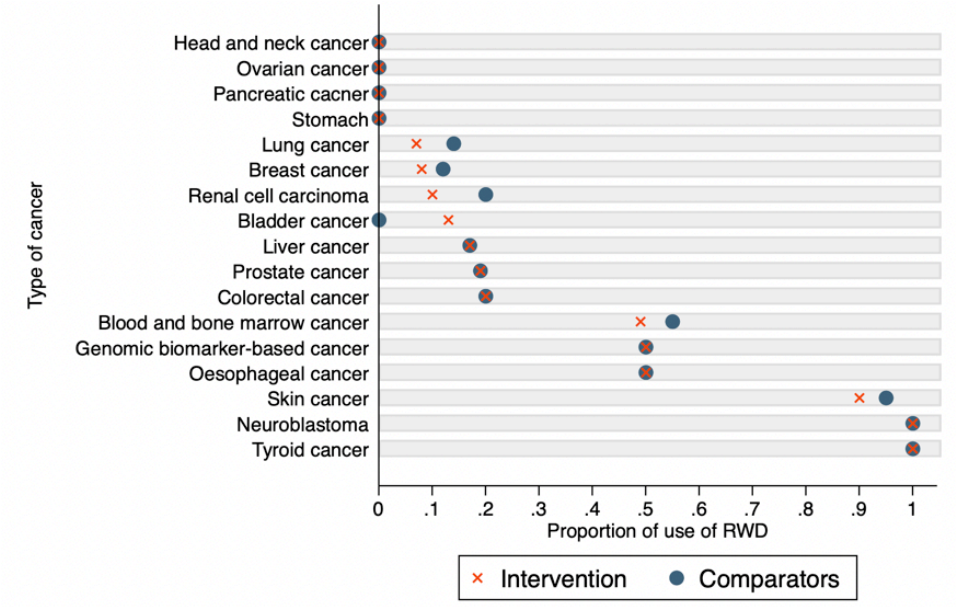


Figure 4.5 Proportion of appraisals using real-world data to estimate overall survival: 2011 - 2021

How real-world data are used to estimate overall survival

How RWD was used to estimate OS was reviewed for a more comprehensive understanding of the process (Figure 4.6). Of the four ways considered, adjusting background mortality is dominant (intervention 62.0%, comparators 42.5%). For interventions, there was no evidence of use of RWD to estimate treatment efficacy. However, of STAs using RWD to estimate OS for comparators, 25% used RWD to estimate the comparative treatment effect of comparators. RWD were used to extrapolate the survival curve in STAs of oncology medicine (intervention 32.4%, comparators 25%). A few appraisals used RWD to adjust for the baseline hazard of disease (intervention 5.6%, comparators 7.5%).

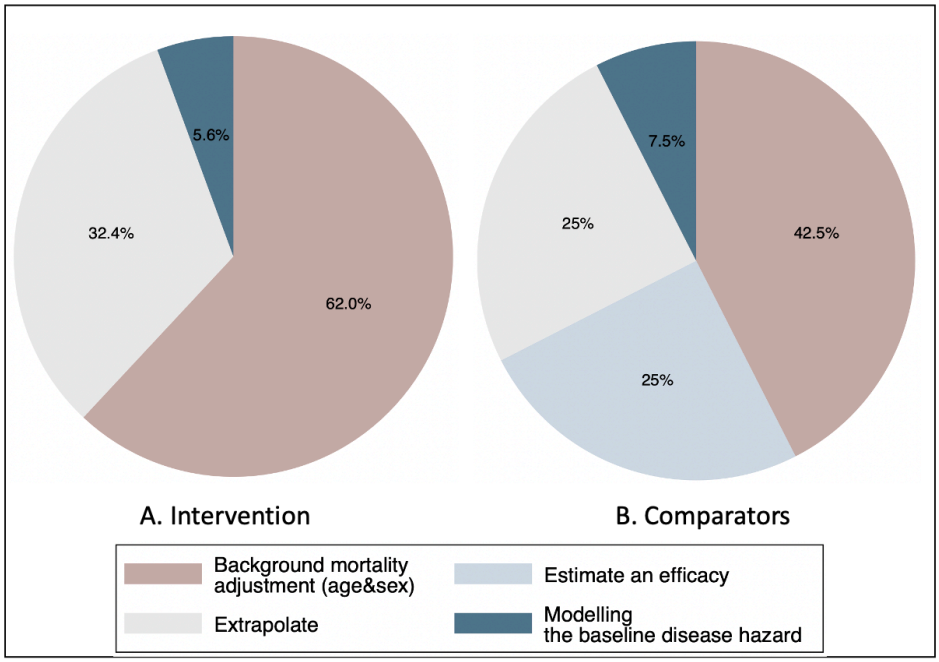


Figure 4.6 How real-world data have been used to estimate overall survival

Sources of real-world data used to estimate overall survival

The sources of RWD used to estimate OS in economic evaluations were reviewed (Figure 4.7). More diverse sources of RWD were found in the comparator group (seven sources of data) than in the intervention group (four sources). In both intervention and comparator groups, the commonest sources of RWD used were national statistics from the Office for National Statistics (ONS). Registry

data were the second most commonly used source of RWD for both interventions (n=17, 27.42%) and comparators (n=20, 28.57%). Four registry databases were frequently used: the American Joint Committee on Cancer (AJCC), the Surveillance, Epidemiology, and End Result Program (SEER), the Haematological Malignancy Research Network (HMRN), and Flatiron Health. In using RWD to estimate OS in comparators, electronic health records and clinical databases were identified as additional data resources.

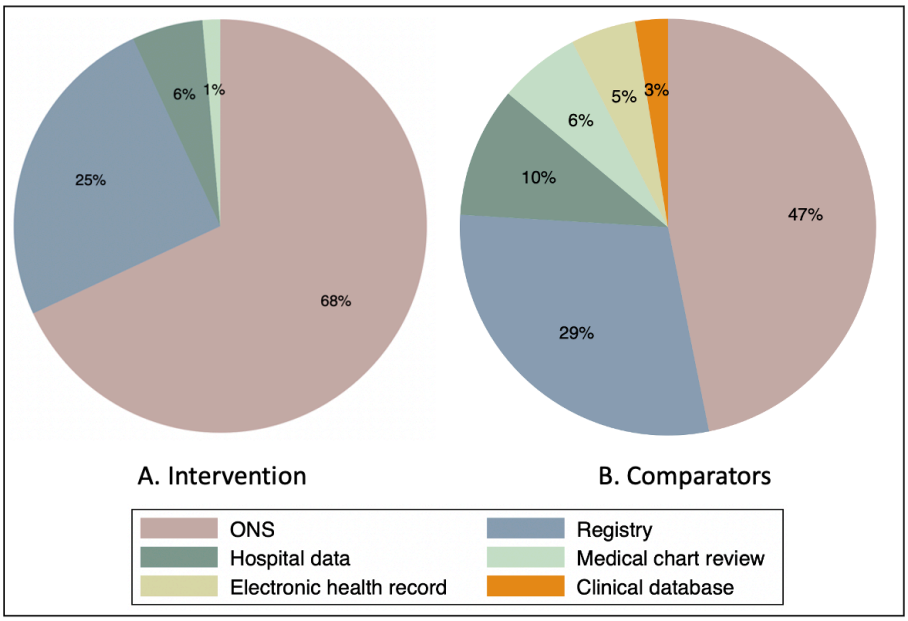


Figure 4.7 Sources of real-world data used to estimate overall survival

The association between the use of real-world data for estimating overall survival and the maturity of survival data

The association between the use of RWD in estimating OS and the maturity of survival data in intervention/comparators, accounting for covariates, was reported according to odds ratios (OR). Regression D analyses the use of RWD to estimate OS for the intervention treatment, and Regression E analyses the use of RWD to estimate OS for the comparator. The log-likelihood ratio chi-square test statistic for Regression D is $LR\ {\chi^2}_{(11)} = 33.02, p < 0.0005$ and for Regression E is $LR\ {\chi^2}_{(11)} = 61.05, p < 0.0000$; therefore, both the models with predictors were significant. Table 4.4 presents the logit coefficients, standard errors and odds ratios.

Table 4.4 Multivariate logistic regression results

	Regression D		Regression E	
	Use of RWD in estimating OS for intervention		Use of RWD in estimating OS for comparators	
Covariate	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Maturity of survival data				
Extremely immature[a]	-	-	-	-
Immature	-0.993**	0.370**	-1.154***	0.315***
	(0.384)	(0.174, 0.786)	(0.394)	(0.146, 0.683)
Mature	-1.061***	0.346***	-1.503***	0.222***
	(0.390)	(0.161, 0.744)	(0.410)	(0.100, 0.496]
Incidence rate				
	-0.000	1.000	-0.000	1.000
	(0.000)	(1.000, 1.000)	(0.000)	(1.000, 1.000)
Direct treatment comparison				
Not available[a]	-	-	-	-
Some available	-1.210**	0.298**	-1.692***	0.184***
	(0.510)	(0.110, 0.811)	(0.522)	(0.066, 0.512)
All available	-0.716	0.489	-1.314***	0.269***
	(0.451)	(0.202, 1.184)	(0.469)	(0.107, 0.674)
External Validity				
Low risk[a]	-	-	-	-
Moderate risk	0.107	1.113	0.613	1.845
	(0.357)	(0.553, 2.240)	(0.378)	(0.884, 3.854)
High risk	0.185	1.203	0.175	1.191
	(0.530)	(0.426, 3.400)	(0.569)	(0.391, 3.630)
Previously recommended				
No[a]	-	-	-	-
Yes	0.128	1.137	-0.251	0.778
	(0.319)	(0.608, 2.124)	(0.334)	(0.404, 1.500)
Constant	0.498	1.646	1.005	2.731
Observations	229		229	
LR R^2	0.1171		0.2069	

**p<0.05, ***p<0.01.

[a] Reference group, CI: confidence interval.

In both regressions, an association between the use of RWD for estimating OS and the maturity of survival data was found. In Regression D, the immature survival data (OR = 0.370, CI: 0.174, 0.786) and mature survival data (OR = 0.346, CI: 0.161, 0.744) odds ratios were less than 1, indicating that for each one-unit increase in the maturity of survival data, the odds of using RWD for estimating OS in interventions decreased by 0.370 and 0.346, respectively. Similarly, in Regression E, for immature survival data (OR = 0.315, CI: 0.146, 0.683) and mature survival data (OR = 0.222, CI: 0.100, 0.496)

there were significant associations and the ORs indicated decreased use of RWD for estimating OS in comparators; the odds decreased by 0.315 and 0.222 for immature and mature survival data, respectively.

The availability of direct treatment comparison was another indicator showing a statistical association. In Regression E, direct comparisons available for all treatments (OR = 0.269, CI: 0.107, 0.674) and some treatments (OR = 0.184, CI: 0.066, 0.512) were associated with decreased use of RWD compared to the reference group (*not available direct treatment comparison*). In Regression D, the OR for some available direct treatment comparisons was 0.298, which indicates that the odds of using RWD for estimating OS in interventions are 0.298 times the odds for the reference case. The availability of direct comparisons for all treatments did not significantly influence whether RWD was used or not to estimate OS for the intervention (p=0.113).

An alternative categorisation of maturity was used to determine the sensitivity of the results to the classification of maturity. While a similar likelihood and significance was observed, the levels of statistical significance were not identical. The results are presented in Appendix 5.8.

4.4 Discussion

This chapter reviewed the maturity of survival data in primary clinical evidence and the use of RWD for estimating OS in economic models in appraisals of oncology medicine. The use of immature survival data appears to increase in economic modelling in appraisals of oncology medicines. 70% of included STAs reported less than 50% of death events, which introduces a high level of uncertainty into the appraisal process. Especially, after introduction of the 2016 CDF, more extremely immature survival data were used in appraisals. RWD have been used in several appraisals to estimate OS when adjusting for background mortality, treatment effects for comparators, and extrapolating the survival curve. This study also found that, in economic models of oncology medicines, mature survival data were negatively associated with the use of RWD for estimating OS.

Over the 11 years examined here (2011 to 2021), the maturity of survival data has decreased. This could be because of an increasing emphasis on early approval for drug access. To enhance innovative drug access, regulatory authorities have accelerated approval for promising new medicines (147,148). Consequently, assessments and appraisals need to be completed earlier, which means that interim results based on limited observations are likely to be used for decision-making. Also, clinical trials for new cancer drugs are more likely to be based on indirect evaluation or measure the surrogate end-point of drug efficacy, such as progression-free survival (PFS) instead of OS (149,150). PFS is increasingly used in drug approval as an alternative intermediate measure to link the surrogate end-point in clinical trials with long-term patient survival (151). Although interim analysis may expedite drug development process (152), such estimates are limited as they often fail to show true treatment effects (153–155). NICE methods guide indicates that when sufficient evidence on surrogate endpoint is not provided, a conditional recommendation can be made and when more mature data become available, the technology should be re-appraised (156). The tendency toward early approval using immature data has increased uncertainty about the clinical benefit and cost-effectiveness for payer's decision-making (16). Consequently, it might lead to a provision through CDF.

In the NHS, provision through CDF is an option for drugs with high uncertainty about the clinical and cost-effectiveness. The immature survival data are one of the most common sources of uncertainty in CDF review appraisals (146,157). Compared to pre-2016 CDF, immature survival data were more likely to be used in the economic models. The results of the regression models showed that the introduction of the 2016 CDF has impacted on the shift of intercept when the OS data got immature over time. It indicated that the less mature survival data were used over time and this trend was accelerated by introducing 2016 CDF. While the regression models showed the intercept shifts after the introduction of the 2016, it is unclear that this also changed the speed of more use of immature survival data. Additional regression model would help understand this change of the use of immature survival data.

One of the important findings of this study was that the broad use was made of RWD to estimate OS in appraisals. The most common way RWD has been used to estimate OS is through adjusting for background mortality. Accurate estimates of changes in survival are important since economic evaluation usually compares differences between treatments over the long-term. Population mortality adjustment is more crucial for extrapolations of immature data than for those of more mature data (158). Understanding changing hazards throughout the disease process can help us choose the right survival curve and adjust the extrapolation more rigorously. RWD could be used to provide the information about the background mortality and change of disease hazard data. The indication is, therefore, that RWD can be used to estimate OS for diverse purposes in economic models in STAs. Incorporating background mortality and registry data into the survival modelling process is discussed in NICE DSU (Decision Support Unit) 21, issued in 2020 (159). Yet generally, in STAs, even background mortality adjustments have been based on limited national statistics. Having a more relevant RWD dataset, such as a nationwide hospital database, can be of benefit for both mortality adjustment and extrapolation calculations.

Another way that RWD have been used to estimate OS in appraisals was to extrapolate the survival distribution. Notably, RWD were incorporated into the survival curve directly during extrapolation in several appraisals. The role of RWD has been highlighted for supplementing information about long-term clinical effects as extrapolating the survival curve is one of the critical elements of appraisal. RWD can help identify the potential long-term effects in early-stage cancers or cancers that progress slowly. Appraising the effectiveness of clinical melanoma studies is one instance where RWD can be useful. It is more likely that mature survival data can be obtained from trials of treatments for patients with short life expectancy. This was observed in the ordinal logistic regression model. More advanced stages of cancer were associated with more mature OS data. Where patients have early-stage cancer or slowly progressing cancer, economic modelling of long-term effects might be improved through greater use of RWD. For example, survival data in skin cancer clinical trials are

often not mature enough due to the slow disease progression. Some companies used data from the AJCC registry, a form of RWD for extrapolation after data-cut.

Interest in RWD has moved to how it can inform comparative treatment effects (160). In this study, there were a few cases (n=19) that used RWD to examine comparative treatment effects. In appraisal of a new health technology, it is nearly impossible to obtain robust RWD for the intervention for a certain indication unless drugs are re-appraised for routine commissioning after the provisional recommendation such as with the CDF. Data from routine practice are not available before the time of appraisal as NICE STAs occur in parallel with the drug approval process. When the drug of interest is not approved and is not used in clinical practice, there are no RWD available indicating how the drug would work. Hence, RWD are less likely to directly inform effectiveness for the treatment of interest. On the other hand, more RWD are available to examine the effectiveness of comparator treatments. There are several studies reviewing the use of RWD to generate an external comparator arm in both drug regulation and HTA submissions (58,161,162). This study found that RWD were already used to estimate the treatment effectiveness of comparators when treatment comparison data were not available from clinical trials. This was shown in the results of regression analysis. The absence of direct treatment comparisons was another indicator that RWD can be used for estimating OS, particularly for comparators. However, it notes that in most identified cases, the comparators were usually basic supportive care, not active treatments.

Overall, the use of RWD is associated with the maturity of survival data – more reliable survival estimates can be achieved with more complete RWD – but it leaves a question about how to deal with the limitations of RWD such as poor data quality and potential bias (163,164). The common sources of RWD are populated from registries. The Flatiron Health or SEER registries were often used for the extrapolation of the survival curve. However, these have problems because they are not representative of NHS practice and have different subsequent treatment lines. The Systemic Anti-Cancer Therapy (SACT) dataset collates the mandatory national systemic anti-cancer therapy activity

data from all NHS England oncology therapy providers (165). This dataset could be more used to inform survival outcomes when adjusting for background mortality or indicating a hazard change. However, currently, the SACT dataset is mainly used to review CDF appraisals and the limited follow-up and the small number of patients have been identified as important limitations of SACT data (146).

Also, the nature of RWD, non-randomisation, challenges examining treatment effectiveness and understanding the actual treatment effect. In appraisals of oncology medicine, population-adjusted treatment comparisons such as matching-adjusted indirect comparisons (MAIC) have been used to control for other confounders and identify treatment effects. Despite these efforts, the true size of the benefits is still uncertain since it is challenging to include the full range of effect modifiers from RWD. Despite increasing interest in RWD, there was little guidance on the use of RWD for estimating clinical effectiveness (166). In the recent published NICE real-world evidence framework, methods guide for real-world studies of comparative effects recommends identifying potential biases originated from different data sources and time-varying confounders and using a statistical method to address confounding. This can help reduce the confounding issues. Along with the NICE framework, accumulated experience of using RWD should increase understanding regarding controlling for confounders in appraisals.

This research examined the use of RWD for estimating OS in NICE STAs over 11 years. The study has a few limitations. First, maturity data were not fully available for all the STAs due to the absence of relevant information. In NICE guidance, information designated as commercial in confidence and academic in confidence can be redacted (167). Clinical data are redacted in many appraisals because they are academic in confidence (168). In this study, despite efforts to identify the relevant information from the appraisal documents and the published clinical study, 20% of survival data were assumed to be extremely immature because the information was not available. Although this could make the findings less accurate, this assumption was reasonable given ERG and committee

statements about immaturity of the survival data in the appraisal documents. Second, the criterion for maturity is not universally agreed upon. This study presented a sensitivity analysis to determine how the results changed according to what criteria were chosen: the findings suggest that results could differ given the criteria used, although the overall tendency is aligned. Third, having a large proportion of death events does not necessarily mean that the survival data are mature. There could be cases where the OS curve has plateaued despite the number of patients at risk. These cases might not be mature enough to observe differential survival and long-term effects. Fourth, the classification of maturity is arguably more from a 'clinical' perspective. It might not properly capture immaturity from an economic point of view. Even if survival data are reasonably mature, if there are big differences in the tails of the survival distributions, with respect to the economics, this can be hugely influential for the QALY gain and ultimately the economic evaluation. Fifth is the lack of agreement over the definition of RWD. This study used a specific definition of RWD. This definition excluded some sources of data such as compassionate use programmes and collecting data in one-off health surveys. Study findings could vary depending on the definition used. Last, the regression model might not include all the predictors for effectively demonstrating the use of RWD for estimating OS. Although the variables that could potentially affect estimates based on RWD were included here, each STA can have its own decision-making context. The regressions might not have captured these contexts. The analysis was only able to adjust for a few other factors that might have been associated with maturity, and also with use of RWD.

Although there are limitations in this research, this study comprehensively investigated maturity of survival data and use of RWD when estimating OS in STAs conducted over 11 years. Data were prepared in a systematic and unbiased manner following a published data extraction protocol. Also, this study is the first to explore the association between use of RWD and maturity of survival data using a regression model. This review provides a systematic and comprehensive examination of the maturity of survival data and the use of RWD in NICE appraisals of oncology medicines.

4.5 Conclusion

In the oncology STAs, survival data appears to be becoming more immature over time. After the 2016 CDF, immature survival data are more likely to be used in appraisals. This study found a negative association between the use of RWD for estimating OS and the maturity of survival data. Absence of direct treatment comparison between the treatment of interest and its comparators was also found to be a predictor of greater use of RWD for estimating OS. While RWD was used not only to adjust the background hazard but also to extrapolate survival curves beyond the trial period, its role in estimating comparative treatment effects has been limited. RWD can help to reduce the uncertainty arising from immature survival data in appraisals of oncology medicines.

Chapter 5. Use of real-world data in appraisals of targeted cancer therapies

Previous chapters have described the use of real-world data (RWD) and the factors associated with the use of RWD. Notably, RWD have been used in appraisals since 2011, and some factors related to the sources of uncertainty are associated with increased/decreased use of RWD. The question that the degree to which sources of uncertainty vary across the appraisals and how RWD are used in the context is naturally followed. It is worthwhile to investigate the common sources of uncertainty and the use of RWD.

This chapter focuses on the appraisals of targeted cancer therapy (TCT) to review the extent to which the uncertainties and use of RWD differ from non-targeted cancer therapy (non-TCT) appraisals. This chapter directly facilitates the research interest of the funding institute, the Centre for Cancer Biomarker (CCBIO). It is noted that CCBIO is not involved in any aspects of the study conduct.

This chapter applies the same methods, patterns and intensity of use described in the previous chapters; therefore, the repetitions are found in a few parts of the methods. This chapter was accepted by a peer-reviewed journal, *BMC Cancer,* in November 2022.

This chapter describes sources of uncertainty and the use of RWD in TCT appraisals compared with non-TCT appraisals. It starts with an introduction to the potential challenges of TCT appraisals. It is followed by a description of TCT appraisals: the number of appraisals over time and by type of cancer, appraisals recommendations and size of the trial population. After the introduction, this chapter describes different sources of uncertainty found in TCT and non-TCT appraisals. Three sources of uncertainty are compared: external validity of a clinical trial, availability of direct treatment comparisons, and maturity of the data on overall survival. After comparing sources of uncertainty, this chapter presents the patterns and intensity of use of RWD in both TCT and non-TCT appraisals. These patterns and intensity are compared to investigate to what extent the use of RWD differs in these appraisals.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	157809	**Title**	Ms
First Name(s)	Jiyeon		
Surname/Family Name	Kang		
Thesis Title	Improving economic evaluation and decision-making for oncology drugs using real-world data		
Primary Supervisor	John Cairns		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	BMC Cancer
Please list the paper's authors in the intended authorship order:	Jiyeon Kang, John Cairns
Stage of publication	**In press**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I had primary responsibility for processing the data, performing the analysis and drafting the paper.

SECTION E

Student Signature	Jiyeon Kang
Date	7 November 2022

Supervisor Signature	
Date	

Research paper 4

Exploring uncertainty and use of real-world data in the National Institute for Health and Care Excellence single technology appraisals of targeted cancer therapy

Jiyeon Kang[1,2], John Cairns[1,2]

[1]Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

[2]Centre for Cancer Biomarkers (CCBIO), University of Bergen, Bergen, Norway

Jiyeon Kang: ORCID: http://orcid.org/0000-0002-1587-9674

John Cairns: ORCID: http://orcid.org/0000-0001-6442-0440

Corresponding author

Jiyeon Kang, PharmD, MSc.

Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Address: 15-17 Tavistock place, London, WC1H 9SH, UK

Email: Jiyeon.Kang@lshtm.ac.uk

Abstract

Objectives Dealing with uncertainty is one of the critical topics in health technology assessment. The greater decision uncertainty in appraisals, the less clear the clinical- and cost-effectiveness of the health technology. Although the development of targeted cancer therapies (TCTs) has improved patient health care, additional complexity has been introduced in drug appraisals due to targeting more specific populations. Real-world data (RWD) are expected to provide helpful information to fill the evidence gaps in appraisals. This study compared appraisals of TCTs with those of non-targeted cancer therapies (non-TCTs) regarding sources of uncertainty and reviewed how RWD have been used to supplement the information in these appraisals.

Methods This study reviews single technology appraisals (STAs) of oncology medicines performed by the National Institute for Health and Care Excellence (NICE) over eleven years up to December 2021. Three key sources of uncertainty were identified for comparison (generalisability of clinical trials, availability of direct treatment comparison, maturity of survival data in clinical trials). To measure the intensity of use of RWD in appraisals, three components were identified (overall survival, volume of treatment, and choice of comparators).

Results TCTs received more recommendations for provision through the Cancer Drugs Fund (27.7%, 23.6% for non-TCT), whereas similar proportions were recommended for routine commissioning. With respect to sources of uncertainty, the external validity of clinical trials was greater in TCT appraisals (p=0.026), whereas mature survival data were available in fewer TCT appraisals (p=0.027). Both groups showed similar patterns of use of RWD. There was no clear evidence that RWD have been used more intensively in appraisals of TCT.

Conclusions Some differences in uncertainty were found between TCT and non-TCT appraisals. The appraisal of TCT is generally challenging, but these challenges are neither new nor distinctive. The same sources of uncertainty were often found in the non-TCT appraisals. The uncertainty when appraising TCT stems from insufficient data rather than the characteristics of the drugs. Although RWD might be expected to play a more active role in appraisals of TCT, the use of RWD has generally been limited.

5.1 Introduction

In England, the National Institute for Health and Care Excellence (NICE) has a role in assessing health technology, such as drugs and medical devices, in informing the best value of using the National Health Service (NHS) resources. Cost-utility analysis is the primary method to assess value for money in appraisals of cancer treatments. Uncertainty is unavoidable when appraising the clinical- and cost-effectiveness of new drugs. Uncertainty refers to the fact that we do not know the expected costs and effects of an intervention in a particular population of patients with absolute precision (133) — the more uncertainty there is in the clinical and cost-effectiveness evidence base for a health technology, the less clear is the appropriate decision. Limited clinical evidence, such as non-comparative studies, studies with small numbers of patients and studies with limited follow-up, could be sources of increased uncertainty in health technology assessment (HTA) decision-making (18). Although data are not sufficient, a decision must still be made. Charlton highlighted that NICE has made decisions based on weaker evidence than previously, which can diminish fairness (169). Hence, understanding and dealing with uncertainty has become more critical than ever in HTA, given the increasing use of uncertain evidence.

Targeted cancer therapy (TCT) refers to treatments that act on specific molecules associated with cancer growth, progression and spread guided by biomarker results (170). Lung cancer is one of the cancers for which TCTs are actively developed. Several altered driver oncogenes characterise non-small cell lung cancer, including KRAS, EGFR, ROS1, ALK, and MET exon 14 alterations (171). These biomarkers are actively used to develop the targeted therapy. Most of the latest lung cancer treatments are targeted therapies (172). Over the last decades, TCT has aroused interest because of the prospect of achieving better health outcomes (173). TCT selects a treatment population based on the expression of biomarkers. Such population targeting can introduce appraisal challenges, for instance recruiting an adequate sample size in clinical trials or choosing relevant comparators based on patient stratification (174,175). In some trials, subgroups are used to show the clinical effectiveness with a suitable biomarker expression. However, subgroups are likely to be too small to

demonstrate statistical significance. These challenges potentially make clinical trials less generalisable to NHS clinical practice. Ultimately, they are likely to be potential sources of uncertainty in appraisals of TCT (176).

Real-world data (RWD) are suggested as a means of overcoming evidence gaps and helping appraisal of innovative drugs in light of the challenges of obtaining the required information from randomised controlled trials (RCTs) (80,177). For example, electronic health records (EHR), a form of RWD, are a potential source of mature survival data which can reduce uncertainty regarding long-term outcomes (61). Also, the use of RWD has been highlighted as a means of constructing external control arms and supporting indirect treatment comparison in decision-making when the treatment effectiveness of comparators is not available from clinical trials (178,179). Furthermore, RWD could provide clinical and environmental information at the patient level, reflecting routine practice (180).

The uncertainty in appraisals is one of the significant concerns in HTA decision-making. RWD has received attention as a means of reducing uncertainty. However, there are caveats with using RWD due to confounders, biases and data quality (37). Also, it is unclear whether RWD can provide the appropriate information in an HTA decision-making context. The Cancer Drugs Fund (CDF) in England offers patients access to drugs while collecting additional information to reduce uncertainty using managed access agreements (181). A recent paper has highlighted RWD's limited role in reducing uncertainty in CDF review appraisals (146). Despite awareness of uncertainty in TCT appraisals and the potential for using RWD, it is unknown to what extent the uncertainties differ between appraisals of TCT and non-TCT and whether RWD are more widely used in economic evaluations of TCT. This study compares appraisals of TCT and non-TCT regarding sources of uncertainty and reviews the use of RWD in these appraisals.

5.2 Methods

This study compared single technology appraisals (STAs) of TCT and non-TCT in terms of appraisal recommendations, the size of clinical trials, types of uncertainties and use of RWD. Chi-square tests were used to show whether any differences between TCT and non-TCT were statistically significant. This analysis includes NICE STAs of oncology medicines for which guidance was issued between January 2011 and December 2021 (n=229). NICE technology appraisal guidance is publicly available (https://www.nice.org.uk/guidance). The appraisals were manually screened to identify the relevant appraisals. This study uses data extracted following a protocol developed to record information about the use of RWD in NICE appraisals of oncology medicines (117). This protocol was designed to extract data used in the economic evaluation, such as general information about technology appraisals, primary clinical evidence characteristics, and the use of RWD. All necessary data for the analysis are available from this dataset.

This research required a definition of TCT. One broadly accepted definition is a cancer treatment that targets specific genes and proteins involved in the growth and survival of cancer cells. However, the definition of TCT has changed over time (182), and TCT, precision medicine and personalised medicine are used interchangeably. Moreover, a biological definition of targeted therapy is less relevant to capture the issues when appraising TCT, as targeting biological molecules does not directly cause the problem. The issues often arise from specifying the population using biomarkers. Hence, in this paper, TCT is defined as an anti-cancer therapy where the indication approved by medical regulators distinguishes patients using biomarkers. In contrast, non-TCT is a cancer treatment not defined as TCT. This implies that some drugs can be categorised differently depending on the indication.

Any analysis of NICE recommendations needs to recognise that a new option became available in 2016 with the advent of a revised CDF. As the available options differ, this study reviewed the NICE appraisal recommendations separately before and after the 2016 CDF. The revised 2016 CDF was

introduced in April 2016. The first STA of a cancer medicine after the 2016 CDF was introduced was the appraisal of azacitidine for treating acute myeloid leukaemia with more than 30% bone marrow blasts (TA399). Any STAs issued after TA399 were regarded as 'after 2016 CDF'.

The size of clinical trials was also reviewed in this study. The number of patients included in the trials was summarised in a histogram to look at the distribution of the trial size. Kernel Density estimation was used to approximate the histogram with a continuous distribution. This estimation compared the similarities and differences between TCT and non-TCT appraisals, focusing on the average number of patients in the trials.

This study focuses on three potential sources of uncertainty in NICE appraisals: the external validity of clinical trials, the availability of direct treatment comparisons, and the maturity of survival data. The sources of uncertainty identified by Morrell et al. (157) were classified into three groups. Appraisal Committees often discuss these sources of uncertainty. The external validity of the clinical study to NHS practice is assessed primarily using the Evidence Review Group's (ERG) assessment of external validity, which the authors have used to classify studies into three groups (acceptable, moderate, and questionable external validity). Three issues potentially affecting external validity (appropriateness of comparators, subsequent treatments received by trial participants, and patient characteristics) are selected to discuss external validity (183,184). When one or more of these issues is identified, the study is coded as of *questionable external validity*. External validity is considered moderate if the ERG raises a few minor concerns. A comment such as "younger and fitter patients" without mentioning performance status is classified as a minor concern. External validity is classified as *acceptable* if there are no specific critiques.

The type of treatment comparison made by manufacturers in their evidence submissions is reviewed to identify the availability of direct treatment comparisons. A sixfold classification of treatment comparisons in NICE appraisals can be made using the information on the availability of head-to-head comparison for all comparators, indirect treatment comparison, anchored/unanchored

treatment comparison and population-adjusted treatment comparison. The possible combinations of treatment comparison are presented in appendix 6.1.

Lastly, the maturity of survival data is highlighted as a source of uncertainty. This study uses three categories (extremely immature, immature, mature) based on the percentage of death events in the primary clinical studies. 20% and 50% were used in this study to classify appraisals, adapting the findings from Tai et al. (21). If the proportion of death events is less than 20%, the maturity of survival data is recorded as extremely immature. When the proportion of death events is between 20% and 50%, the survival data are immature, and greater than 50%, the survival data are considered mature. The published clinical studies were consulted if this information was redacted in the appraisal document. If the proportion was not reported in the results of the original research, comments on maturity in the ERG report were checked. If none of this information was available, the survival data were considered extremely immature.

There are many potential uses of RWD in an appraisal and several ways of reporting the use of RWD. Simple counts of the number of occasions when RWD are used in an appraisal may not be a good guide to how differently one appraisal utilises RWD compared to another. This study used a few different methods, such as pattern review and intensity analysis, to review the use of RWD. The patterns of use of RWD were reviewed to provide a clearer picture of how RWD have been used. The extraction protocol distinguished 31 economic evaluation components where RWD might be used, giving rise to many different patterns. The patterns were reviewed by distinguishing between the parametric and non-parametric use of RWD. Parametric use involves basing the numerical value of specific variables in the economic model on RWD. For example, the use of data to provide values for overall survival (OS) or resource use in the economic model is categorised as parametric use. Non-parametric refers to using RWD to develop the model structure and support or validate assumptions in the model. Using RWD to select comparators or validate the survival distribution choice are examples of non-parametric use. This separation provides a more comprehensive review of how

RWD have been used in appraisals. All components where RWD could be used are presented in appendix 3.

The intensity of use of RWD in different appraisals was investigated by classifying different patterns in terms of the extent to which RWD are drawn upon in different economic evaluation components. Three components (OS of intervention/comparator, volume of treatment of intervention/comparators, choice of comparators) are identified as major uses of RWD, which are likely to have a high impact on the outcome of the economic evaluation, the incremental cost-effectiveness ratio (ICER). The remaining components are regarded as minor uses of RWD. The identified patterns were categorised into seven groups by distinguishing major and minor uses of RWD (Figure 5.1). Two classifications are suggested. One counts the number of major and minor components; another is a simplified classification that only counts the number of major components. The group with all three major components is the highest intensity use of RWD.

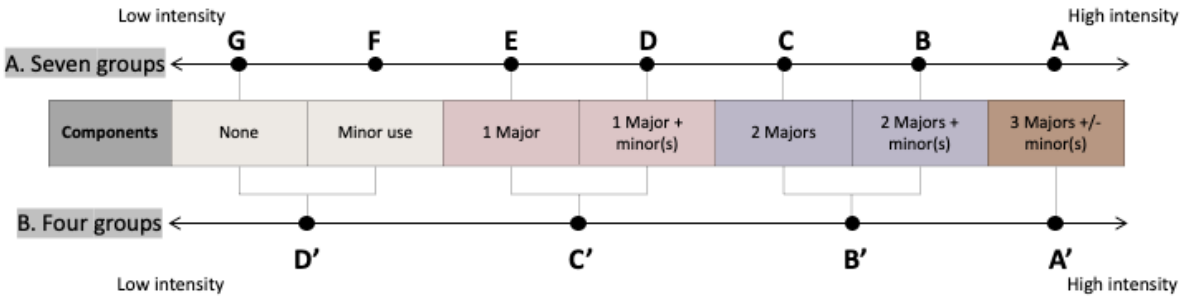


Figure 5.1 Classifications distinguishing major and minor use of real-world data

5.3 Results

5.1.1 Appraisals of targeted cancer therapy and non-targeted cancer therapy

Figure 5.2 shows published STAs of TCT and non-TCT over time. All identified STAs were included in this analysis (n=229). The number of STAs of oncologic medicines has generally increased over time except for 2019 and 2020. Of included STAs, 36% were TCT appraisals. Although there were

fluctuations, the TCT proportion has increased over time. The highest proportion of TCT appraisals was in 2019 - 57% of oncology appraisals. Note there were no TCT appraisals published in 2011.

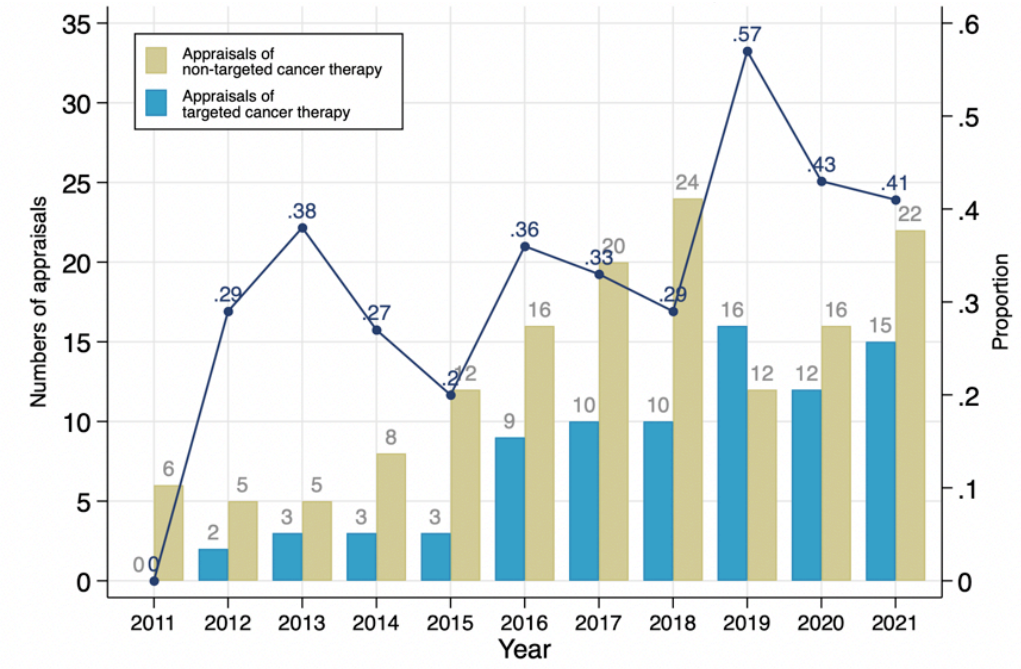


Figure 5.2 Appraisals of oncology drugs: 2011-2021

Figure 5.3 shows TCT and non-TCT appraisals by cancer type. Cancer areas where TCTs have been actively introduced are breast cancer (76% of breast cancer appraisals) and lung cancer (70% of lung cancer appraisals). In genomic biomarker-based cancer treatments known as histology-independent therapies, TCTs show the highest proportion because of the nature of the treatment. As a new generation of treatment, the genomic biomarker-based cancer treatment is histology-independent, which treats cancers based on a biomarker, not by the location of cancer. The two drugs, entrectinib and larotrectinib in this category, are currently recommended within the CDF.

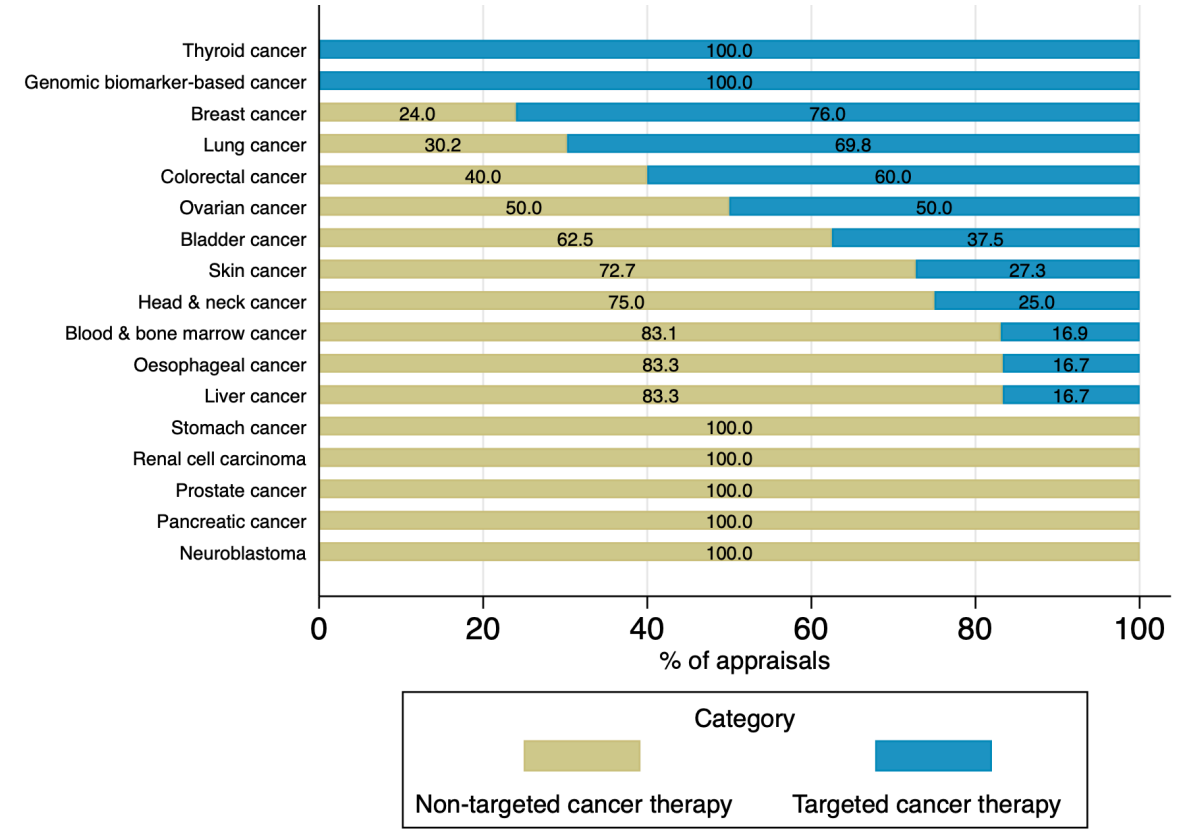


Figure 5.3 Targeted and non-targeted cancer appraisals: 2011-21 by cancer

The TCT and non-TCT appraisal recommendations are reported in Table 5.1. Overall, appraisals of TCT have a higher proportion of positive recommendations for routine commissioning, although the difference is not statistically significant. There has been no significant difference in recommendations to provide through the CDF between the two groups following the introduction of the 2016 CDF.

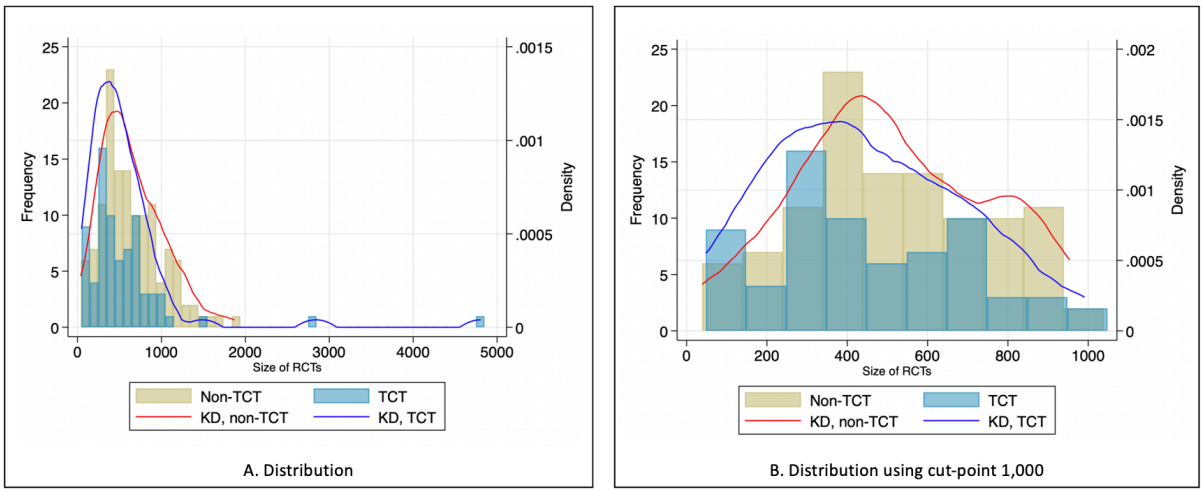
Table 5.1 Appraisal recommendations

	TCT	Non-TCT	χ^2 (p)
Overall			
Not recommended	7 (8.43%)	29 (19.86%)	6.7409 (0.150)
Recommended (routine commissioning)	40 (47.95%)	61 (41.78%)	
Optimised*	16 (19.28%)	30 (20.55%)	
CDF	14 (16.87%)	21 (14.38%)	
CDF, Optimised	6 (7.23%)	5 (3.42%)	
Total	83 (100%)	146 (100%)	

Before introducing the 2016 CDF			
Not recommended	3 (23.08%)	15 (38.46%)	1.2966 (0.523)
Recommended	9 (69.23%)	20 (51.28%)	
Optimised	1 (7.69%)	4 (10.26%)	
Total	13 (100%)	39 (100%)	
After introducing the 2016 CDF			
Not recommended	4 (5.71%)	14 (13.08%)	3.8190 (0.431)
Recommended (routine commissioning)	31(45.83%)	41 (38.32%)	
Optimised	15 (21.43%)	26 (24.30%)	
CDF	14 (20.00%)	21 (19.63 %)	
CDF, Optimised	6 (8.57%)	5 (4.67%)	
Total	70 (100%)	107 (100%)	

*"Optimised" is a recommendation for a smaller group of patients than originally stated by the marketing authorisation.

The number of patients in the clinical trials upon which treatment effectiveness in the economic models was based was reviewed to compare the sizes of the overall trials between TCT and non-TCT. Most clinical studies had fewer than 1,000 patients. Right skews were found (Figure 5.4A). These right-skewed distributions show that most values for both TCT and non-TCT are clustered around the left tail of the distribution. This distribution implies that most trials (of both TCT and non-TCT) are relatively small. To compare the distributions more clearly, the distributions have been trimmed at 1,000 in Figure 5.4B. Appraisals of TCT had their peak density around 300-400, whereas appraisals of non-TCT peaked at around 400-500.



TCT: Targeted cancer therapy, Non-TCT: non-targeted cancer therapy, KD: Kernel density

Figure 5.4 Distribution of trials by size

5.1.2 Sources of uncertainty in NICE appraisals

Potential sources of uncertainty are summarised in Table 5.2. While there is no statistical difference in the availability of direct treatment comparisons, the external validity of the clinical studies and the maturity of the survival data differ significantly.

Table 5.2 Sources of uncertainty in appraisals

	TCT	Non-TCT	χ^2 (p)
The external validity of clinical studies			
Acceptable external validity	36 (43.37%)	39 (26.71%)	7.2714 (0.026)
Moderate external validity	37 (44.58%)	90 (61.64%)	
Questionable external validity	10 (12.05%)	17 (11.64%)	
Total	83 (100%)	146 (100%)	
Availability of direct treatment comparison			
Not available	28 (33.73%)	43 (29.45%)	1.1922 (0.551)
Some available	28 (33.73%)	45 (30.82%)	
All available	27 (32.53%)	58 (39.73%)	
Total	83 (100%)	146 (100%)	

Maturity of survival data			
Extremely immature	29 (34.94%)	56 (38.36%)	7.2550 (0.027)
Immature	35 (42.17%)	38 (26.03%)	
Mature	19 (22.89%)	52 (35.62%)	
Total	83 (100%)	146 (100%)	

The external validity of the clinical study

The uncertainties concerning external validity raised in the appraisals were reviewed. These factors (appropriateness of comparators, subsequent treatment received by trial participants, and patient characteristics) are usually addressed in the ERG reports when assessing the generalisability of trial outcomes to NHS practice. Twenty-seven appraisals were identified, where the ERG highlighted the high level of uncertainty with respect to the external validity of the clinical evidence. Ten of these appraisals were TCT. Problems were identified with respect to the study population (70%), the comparators (20%) and subsequent treatment received by trial participants (10%). In appraisals of non-TCT, the external validity of evidence was heavily questioned in seventeen appraisals. The main reason was the study population (53%), followed by the issue of subsequent treatment received by trial participants (35%). The general problem of trial populations being younger and fitter than routine practice is widely noted by ERGs. However, this was not a major reason for the high level of uncertainty unless subgroups in the trial were very different from those in routine practice. More often, the issues with respect to the study population arose from differences in prior treatment, which might impact survival outcomes. For example, in an appraisal of nivolumab (NICE TA530), the ERG expressed serious concerns regarding the representativeness of the trial population to the UK population (185). One of the reasons was a mismatch of prior therapies. More than 75% of patients in UK clinical practice received a previous gemcitabine platinum-based therapy, while less than 40% of the trial population did. Another example is an appraisal of durvalumab (NICE TA578) (186). The ERG identified that the population in the clinical trial (PACIFIC) was narrower than in the scope

(patients expressing PD-L1 >1%). Also, they received different types of chemoradiation therapy cycles. UK patients received sequential rather than overlapping treatment, potentially affecting the treatment effect.

Types of treatment comparison in manufacturer submissions

The treatment comparisons made were not statistically different between TCT and non-TCT appraisals. The availability of head-to-head RCTs was reviewed to understand the patterns of indirect treatment comparison. The proportion of single-arm trials in TCT appraisals is higher than that of non-TCT. Nineteen TCT appraisals did not use RCTs as primary clinical evidence (23% of TCT appraisals). Several possible ways to compare treatments were found in these appraisals (Figure 5.5). In general, TCTs and non-TCTs show similar patterns of treatment comparisons. Thirty-one per cent of all appraisals made indirect treatment comparisons (ITC). Among the appraisals using ITC, 79% made unanchored ITC. TCT appraisals show a higher proportion of unanchored ITCs than non-TCT (23% of TCT appraisals, 14% of non-TCT appraisals).

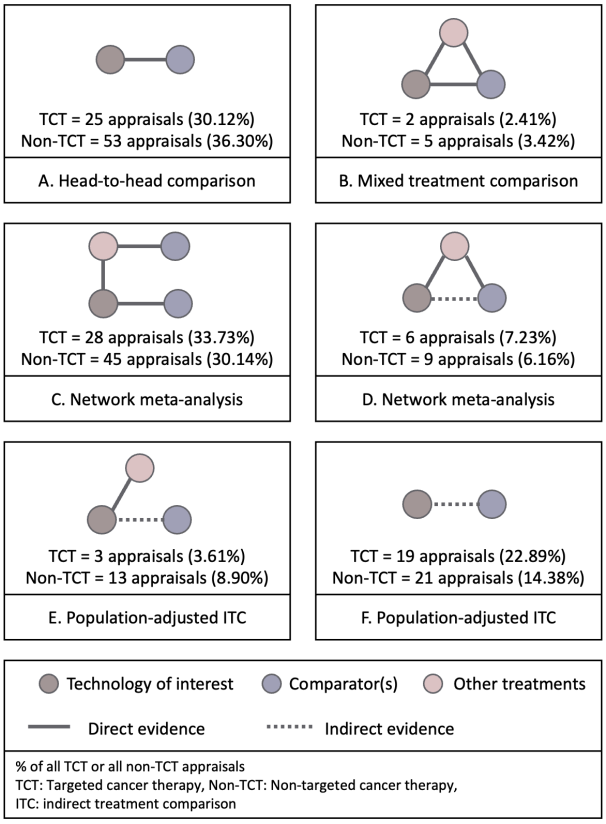
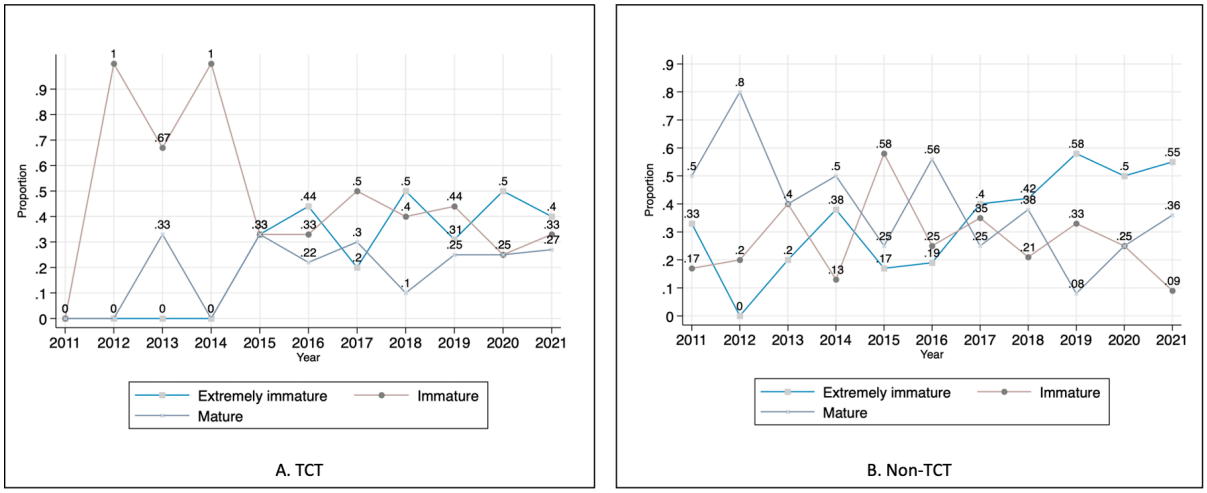


Figure 5.5 Illustration of treatment comparisons identified in company submissions

Maturity of survival data in clinical trials

The maturity of survival data showed a statistical difference between TCT and non-TCT appraisals. The proportion using extremely immature survival data was similar between the two groups, whereas immature survival data were used more in TCT appraisals. The changes in the use of extremely immature, immature and mature survival data over time are shown in Figure 5.6. Although it is difficult to see the clear patterns in the use of immature survival data, the proportion of the STAs using immature survival data tends to have increased over time in both groups.



TCT: Targeted cancer therapy, Non-TCT: non-targeted cancer therapy

Figure 5.6 Maturity of survival data in appraisals of targeted cancer therapy and non-targeted cancer therapy

5.1.3 The use of real-world data in the economic models of targeted cancer therapy and non-targeted cancer therapy

Pattern review

There is no dominant pattern of use of RWD in these appraisals. Fifteen different patterns of use of RWD can be identified, which appeared in three or more appraisals. These patterns cumulatively account for 51% of all appraisals (Appendix 4.1.a). The pattern, estimating overall survival of intervention and comparators, was the most commonly observed (13 appraisals, 6% of patterns), followed by the pattern estimating end-of-life resource use (12 appraisals, 5% of patterns). In

appraisals of TCT, using RWD for estimating end-of-life cost is the most common pattern (8 appraisals, 10% of patterns), whereas estimating OS of intervention and comparators was found in only one TCT appraisal (1%).

When looking at the non-parametric and parametric use of RWD separately, more diverse patterns were found for parametric use than for non-parametric use. Sixty-two per cent of all appraisals involved no non-parametric use of RWD (Appendix 4.1.b). The commonest pattern of non-parametric use of RWD was to validate the choice of survival distribution for the intervention and comparators (TCT: 11 appraisals, 13%; non-TCT: 9 appraisals, 6%). Some patterns found in non-TCTs were not identified in appraisals of TCT. Regarding the parametric use of RWD, 23% of appraisals did not use RWD to inform any parameter in the model (Appendix 4.1.c). In appraisals of TCT, using RWD for estimating end-of-life resource use (16 appraisals, 19%) and for estimating both end-of-life and health state resource use (7 appraisals, 8%) were common patterns. Fifteen non-TCT appraisals (10%) used RWD to estimate OS for the intervention and comparators.

Intensity analysis

For analysis of the intensity of use of RWD, all appraisals included in this study were classified into intensity groups using the two classifications in Figure 5.1. While classification A shows a statistically significant difference in intensity between appraisals of non-TCT and TCT (χ^2=14.66, p=0.012), classification B does not provide a significant difference (χ^2=6.8035, p=0.078). Over time, the major use of RWD has increased in both groups of appraisals. In 2020, about 60% of TCT and non-TCT appraisals made at least two major uses of RWD. The cases of three major uses of RWD were observed in the non-TCT group in 2018. Such a major use of RWD was not observed in the TCT group. Using classification A (Figure 5.7A & B), there does not appear to have been an evident change in the intensity of use of RWD. Whereas, using the simpler classification B, the intensity of use of RWD appears to have increased over time (Figure 5.7C & D).

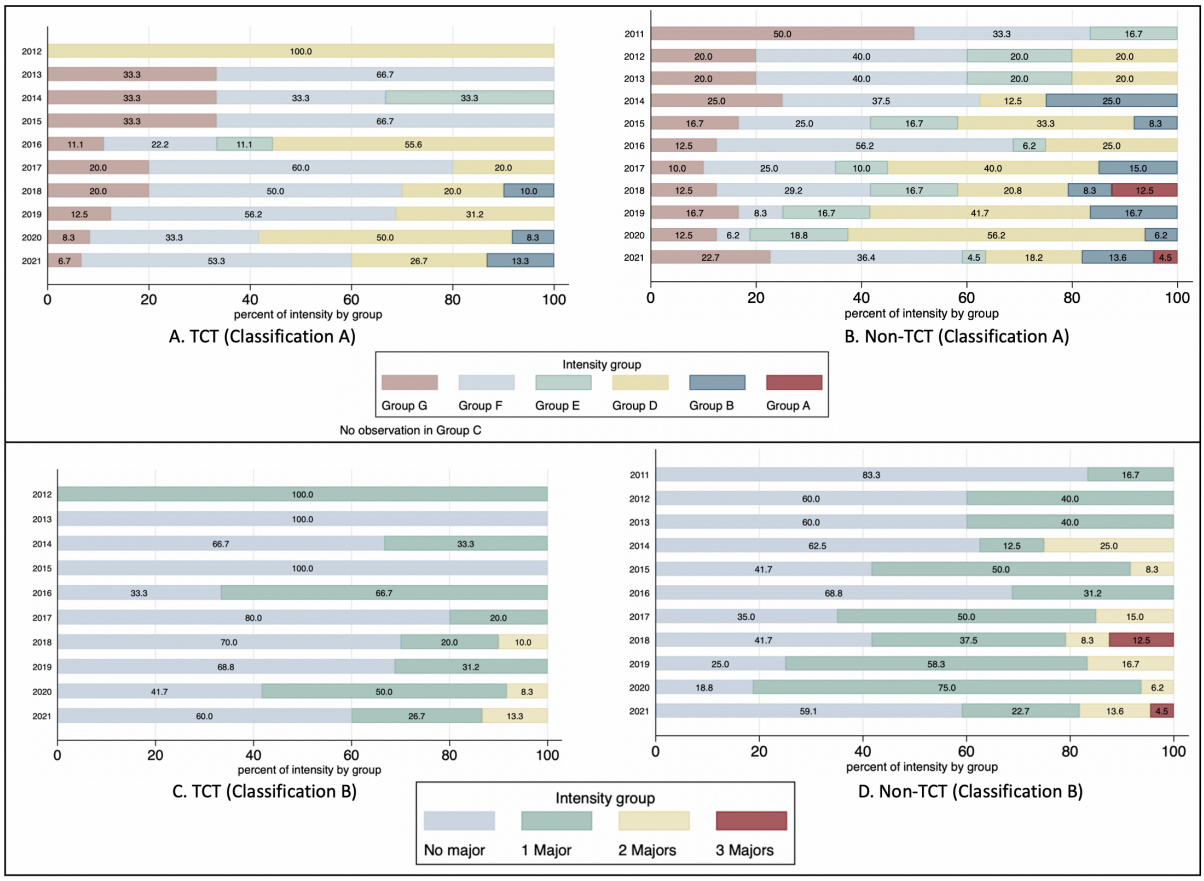


Figure 5.7 Intensity of use of real-world data over time

5.4 Discussion

This study compared appraisal recommendations, the size of clinical trials and sources of uncertainty and uses of RWD in STAs of TCT and non-TCT. TCT appraisals have higher rates of positive recommendation, although the difference was not statistically significant. The proportions of positive recommendations might vary in response to differences in the ICERs believed by the appraisal committee. However, the confidential nature of many drug prices limits reporting of precise ICERs and, thus, the exploration of differences in ICERs between TCT and non-TCT. Another possible explanation suggested by Cairns is that uniform pricing across indications combined with individual TCTs having fewer indications might explain the different recourse to the CDF (187). If a drug is already routinely commissioned for one indication, an extension of routine commissioning to other indications would be expected to be at the original price. In contrast, provision through the CDF could be at a different price.

The size of trials was compared between TCT and non-TCT appraisals. The cancers where TCTs have been actively developed were lung and breast cancer. Both cancers are common cancers (188). Also, some of the biomarkers found in these cancers are relatively common biomarkers. This implies that the "targeted population" is not necessarily small. Depending on the commonness of the disease and the proportion expressing the relevant biomarker, the target population size could be large enough to show statistical significance. An example is the human epidermal growth factor receptor-2 (HER2) as a prognostic and predictive marker for breast cancer. About 20 – 30% of breast cancer patients show overexpression of HER2. In the appraisal of trastuzumab emtansine for adjuvant treatment of HER2-positive early breast cancer (TA632), the primary clinical evidence, KATHERINE trial, recruited 1,486 patients randomised 1:1 to intervention and comparators. Given that the average trial size of non-TCT was 400-500, in TCT appraisals in these cancers, the extent to which the appraisal challenges are rooted in the characteristics of the TCT is diminished.

In contrast, rare cancers and rare biomarkers, which yield a significantly narrower population, could be a source of the risk when appraising drugs based on highly uncertain evidence in the future. The Neurotrophic tyrosine receptor kinase (NTRK) inhibitors (NICE TA630 larotrectinib (189), NICE TA644 entrectinib (190)) are good examples of likely future challenges . In these appraisals, the main clinical trials were basket trials, which is a novel trial design to evaluate the treatment effectiveness of TCT for one or more targets regardless of the pathology (191). Also, companion diagnostic tests for this biomarker were absent (192). In the entrectinib appraisal, the committee noted that "the population eligible for entrectinib is broader than the trial population, so entrectinib's clinical effectiveness in some groups is unknown" (p.13, Final Appraisal Determination of NICE TA644). Data from too few patients, immature survival data, and the absence of direct comparison were all addressed in the appraisal. Due to the uncertainty, these drugs are currently recommended within the CDF. Additional data, including RWD, are being collected to reduce uncertainty while these drugs are being provided through the CDF. However, to what extent these additionally collected data will help to reduce uncertainty is not clear (146).

To date, the targeting of treatment populations has not introduced significantly different appraisal challenges. However, the next generation of TCT, such as histology-independent therapy, might present more decision-making challenges, including identifying the eligible population and appropriate prices across the different populations in the future (193). Overall, TCT appraisals have fewer sources of uncertainty in the evidence despite the concerns about the poor quality of evidence. With respect to uncertainty around external validity, the characteristics of TCTs have some impact on these differences in uncertainty between TCT and non-TCT appraisals. The challenges inevitably increase when the population is restricted using specific biomarkers. Targeting specific populations leads to issues such as insufficient statistical power and eligibility depending on biomarker expression levels, increasing uncertainty regarding the external validity of trial outcomes to NHS practice. However, targeting the population is not the only source of uncertainty in TCT appraisals. Uncertainty is likely to increase with other factors, often found in non-TCT appraisals, such as finding the most suitable population for decision-making. In appraisals of TCT, differences in previous treatment options or subsequent treatment often raised questions concerning the representativeness of the trial data for NHS patients and the likely size of the treatment effect in practice. This adds to the uncertainty around the small size of the eligible population in appraisals of TCT but also of non-TCT.

Uncertain clinical outcomes due to immature survival data are commonly encountered in NICE appraisals (157). The immaturity of survival data introduces substantial uncertainty in the extrapolation of survival (135,136). The TCT appraisals used less mature survival data than appraisals of non-TCT. In appraisals of immunotherapy, a large portion of TCT in this research, appraisal committees often questioned the duration of the treatment effect when predicting the long-term effect. One of the novel response patterns reported in immunotherapy is a sustained response in a small number of patients after stopping immunotherapy (194). In NICE TA692, the duration of the continued treatment effect was described as an area of uncertainty for all immunotherapies (195). Immature survival data are more likely to increase the importance of this issue as no long-term data

are available. A longer follow-up would help reduce uncertainties concerning the duration of response to treatment and OS (151). However, this issue is not the only issue in TCT appraisals. A large proportion of non-TCT appraisals used immature survival data. It implies that the absence of long-term data introduces a great level of uncertainty in understanding long-term treatment effects and causes a problem in most cancer appraisals. This can be met by efforts to provide better quality evidence in appraisals and by managed access agreements such as the CDF, which can help to understand the long-term effect by following up the trial population.

The limited availability of direct treatment comparisons was identified as a source of uncertainty across appraisals. Regardless of the treatment type, obtaining head-to-head estimates of comparative effectiveness from a single trial becomes more challenging since the treatment options are rapidly expanding. When direct treatment comparison is not available in a trial, network meta-analysis has been used to identify the treatment effect indirectly. However, a network is not always available unless a common comparator links the available trials (196). The indirect treatment comparison is unanchored when the primary clinical evidence is a single-arm trial or the evidence cannot be linked to other clinical trials. Analytical techniques such as matching adjusted indirect treatment comparison (MAIC) or simulated treatment comparison have been used when making unanchored comparisons. However, these methods do not usually resolve the uncertainty around indirect comparison since it is not possible to adjust fully for all effect modifiers. An example is the appraisal of trastuzumab deruxtecan (NICE TA704) (197). In this appraisal, the main clinical evidence was a single-arm trial (DESTINY-Breast01). Due to the absence of direct comparative evidence, treatment effectiveness was assessed using an unanchored MAIC. The Appraisal Committee was concerned that important factors such as HER2 status and previous anti-HER2 therapy could not be adjusted for and concluded that the MAIC had limitations and the results were uncertain.

This study found that the evidence used in appraisals of new cancer drugs was uncertain across both TCT and non-TCT appraisals. The sources of uncertainty observed in TCT appraisals were not

essentially different from those in appraisals of non-TCT. The uncertainties decision-makers face are ones they have faced previously. Given the novelty of targeted therapy, a new approach was required, such as an innovative clinical trial design and strategy for early decision-making to improve operational efficiency (198). However, it is uncertain whether novel approaches such as enrichment trial design and trials with adaptive design can help the appraisal process more or introduce additional uncertainty (199,200). More importantly, current appraisal challenges arise from data insufficiency rather than the inherent characteristics of these drugs (201). The sources of uncertainty were more frequently found in the appraisals of non-TCT in this study. Regardless of the type of technology, NICE decision-making uses uncertain evidence.

RWD have been identified as supplementing RCT data. As the pattern review showed, RWD were used in diverse ways. However, while many are optimistic about the potential contribution of RWD (202), the use of RWD has contributed little to both TCT and non-TCT appraisals. RWD were generally only used for relatively unimportant aspects of the evaluation. This limited use of RWD could be explained by several concerns around RWD, including potential bias and study design limitations (163,164). Due to the limitations, using RWD might not particularly answer the questions about uncertainty. Also, given that fewer sources of uncertainty were found in TCT appraisals, there could be less incentive to use RWD. Further study of the factors associated with increased/decreased use of RWD would broaden understanding in the future.

Although limited use was made of RWD, it is notable that the intensity of use of RWD has increased over time. Among the patterns that appeared in three or more appraisals, five patterns included using RWD for estimating OS. It is a noteworthy result given the strong signal of NICE's interest in the use of RWD (109). Although this study cannot provide detailed information on how RWD were used for this purpose, RWD can be used in several ways to estimate OS, such as adjusting disease hazard and extrapolating the survival curve. Recently, NICE published a real-world evidence framework to

guide research on comparative treatment effects using RWD. Additional studies on how RWD have been used in estimating OS will help understand the opportunities and challenges of RWD.

This study explored several aspects of appraisals of TCT and non-TCT from an HTA perspective. Given the increased interest in using biomarkers to identify treatment groups, there will likely be growing challenges in appraising TCT. Although the findings of this study could change over time as more TCT are developed, this study is the first to document systematically the differences and similarities in sources of uncertainty and use of RWD between appraisals of TCT and non-TCT by reviewing over two hundred appraisals. However, this study has a few limitations. First, the information about external validity relies on the ERG reports. Although appraisal committees agree with ERG's assessments in general, committees do not necessarily always agree on all points with ERGs. What committees critically emphasise regarding external validity could be different.

Another limitation is the classifications of uncertainties and intensity of use of RWD. Although all the information used in this study was obtained from appraisal documents, how to categorise this information was based on the data extraction protocol. The maturity of survival data was classified using two values, 20% and 50%. However, these points are not agreed criteria to define data maturity. Committees can make different judgements with respect to maturity. With respect to the intensity classification, this study focuses on a specific assumption that the use of RWD in three major components would be intensive use of such data. However, the criteria to measure intensity are not universally agreed upon. Also, decision-makers might not be concerned about which RWD inform components of the economic model. More likely, they would concentrate on how RWD would help to address the decision problem. How to classify uncertainty and intensity of use could differ across researchers and decision-makers.

Finally, it is noted that there might be a difference in the number of appraisals depending on which criteria were used. In this study, the STAs for treating side effects of cancer drugs were excluded. When appraisals are collected, potentially also affects their number. Some appraisals available in this

study might not be available later due to the replacement of appraisals (CDF review, withdrawn etc.) Likewise, previously available appraisals might not be included in this study as the guidance was withdrawn. Despite this potential difference, this study included all STAs of cancer therapy which were available as of December 2021.

5.5 Conclusion

Some differences in uncertainty were found between TCT and non-TCT appraisals. The appraisal of TCT is generally challenging, but these challenges are neither new nor distinctive. The same sources of uncertainty were also often found in the non-TCT appraisals. The uncertainty in appraising TCTs is more likely to stem from insufficient data rather than the inherent characteristics of the drugs. Although RWD might be expected to take a more active role in appraisals of TCT, the use of RWD has generally been very limited.

Declaration

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the London School of Hygiene and Tropical Medicine on 14 November 2019 (17315). This study used the publicly available data. The consent to participate is not applicable in this study. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The data analysed during the current study are available in the National Institute for Health and Care Excellence website:

[https://www.nice.org.uk/guidance/published?ngt=Technology%20appraisal%20guidance&ndt=Guidance]

Competing interests

None declared.

Funding

JK is supported by the Centre for Cancer Biomarker, University of Bergen funded by the Research Council of Norway grant number (223250). The funder was not involved in any aspect of the study conduct or the decision to submit the manuscript for publication.

Authors' contributions

Both authors contributed to conceptualising and designing the study. JK analysed the data and drafted the protocol manuscript. JC revised the manuscript for important intellectual content and contributed to the methodology.

Acknowledgements

The authors thank Dr Alec Miners for helpful advice on the clarity of methodology and discussion.

Chapter 6. Use of real-world data in Cancer Drugs Fund reviews

While the previous chapter has identified the differences in the sources of uncertainty and the use of real-world data (RWD) in appraisals by comparing them between two different therapy groups, how RWD were used to reduce the uncertainty is not fully explained. This chapter focuses on the use of RWD to reduce uncertainty. This chapter contributes to this thesis by providing a detailed analysis of the use of RWD in the review appraisals of drugs exiting the 2016 CDF, paying close attention to the sources of additional data and the extent to which they reduced the uncertainties highlighted in the original appraisal. Some parts of the methods are repetitive, such as patterns of use of RWD and classification of the intensity of use. This paper was published in a peer-reviewed journal, *Pharmacoeconomics-Open,* in September 2022.

This chapter describes the common sources of uncertainty in appraisals where a recommended for provision through the CDF were made and the extent to which additionally collected RWD, mainly Systemic Anti-Cancer Therapy (SACT) data, help reduce the uncertainty in CDF review appraisals. It starts with an introduction to the features of CDF, especially highlighting the requirement for additional data collection. This chapter illustrates how to identify the sources of uncertainty in the CDF original appraisals, followed by how to identify key uncertainties. After a description of key uncertainties in the original appraisals, the patterns and intensity of use of RWD, identified using the methods in Chapter 3, are used to compare the use of RWD in original appraisals and review appraisals. This is followed by a review of the use of additional data to reduce uncertainty. Especially, the use of SACT data is highlighted to investigate to what extent such data help resolve the uncertainty described in original appraisals.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	157809	**Title**	Ms
First Name(s)	Jiyeon		
Surname/Family Name	Kang		
Thesis Title	Improving economic evaluation and decision-making for oncology drugs using real-world data		
Primary Supervisor	John Cairns		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Pharmacoeconomics-Open		
When was the work published?	20 September 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I had primary responsibility for conceptualising the study design, obtaining the data, performing the analysis and drafting the paper.

SECTION E

Student Signature	Jiyeon Kang
Date	7 November 2022

Supervisor Signature	
Date	

Research paper 5

"Don't Think Twice, It's All Right":

Using additional data to reduce uncertainty regarding oncologic drugs provided through managed access agreements in England

[Running title] Uncertainty and Additional Data in the CDF

Jiyeon Kang PharmD[1,2], John Cairns MPhil[1,2]

[1]Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

[2]Centre for Cancer Biomarkers (CCBIO), University of Bergen, Bergen, Norway

Jiyeon Kang: ORCID: http://orcid.org/0000-0002-1587-9674

John Cairns: ORCID: http://orcid.org/0000-0001-6442-0440

Corresponding author

Jiyeon Kang, PharmD, MSc.

Department of Health Service Research and Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Address: 15-17 Tavistock place, London, WC1H 9SH, UK

Email: Jiyeon.Kang@lshtm.ac.uk

Abstract

Objectives The Cancer Drugs Fund (CDF) in England uses managed access agreements to facilitate additional data collection to address uncertainties identified in the appraisals of new drugs. This study reviews the uncertainties highlighted in the original appraisals where recommendation 'to use within the CDF' were made and how additional data were used to address these uncertainties in the CDF review appraisals where final decisions on routine commissioning were made.

Methods The first twenty-four drugs exiting the 2016 CDF were included in this review. The information about uncertainty and the use of newly collected data were extracted from the original appraisals and the CDF review appraisals. The additional data used in the CDF review appraisals, distinguishing between clinical trial data and real-world data (RWD), were reviewed to assess the extent to which the additional data were able to reduce the original uncertainties.

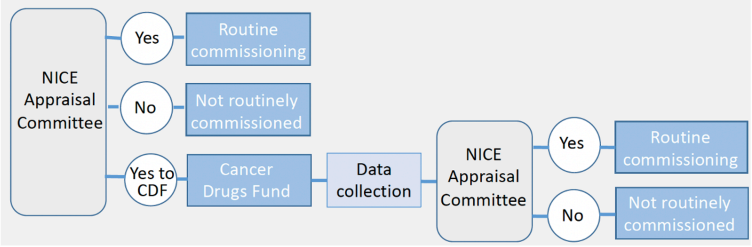
Results The recommendation that the drug be routinely commissioned was made in 87.5 per cent of re-appraisals. Uncertainty stemming from immaturity of the survival data in clinical trials was frequently found in appraisals. Later follow-up of clinical trials was used to address this uncertainty whereas limited use was made of RWD. The systemic anti-cancer therapy (SACT) dataset is the most frequently used source of RWD. SACT data were mostly used in review appraisals to support the clinical outcomes based on later follow-up of trial participants and to inform modelling of subsequent treatments or treatment duration.

Conclusions While additionally collected RWD attracted attention when the 2016 CDF was introduced, RWD have not been widely used in CDF review appraisals and (to date) have done little to reduce uncertainty. Experience with these appraisals has highlighted the importance of longer follow-up of clinical trials and the relatively limited role of RWD, in general, and of SACT data in particular.

6.1 Introduction

New oncology drugs receive special treatment in England. Since January 2009, differential valuation of the health benefits of many cancer drugs has been implemented by adopting a higher cost-effectiveness threshold for life-extending, end-of-life treatments within the National Health Service (NHS) (203). In 2010-2011, the Cancer Drugs Fund (CDF) was introduced to provide cancer patients in England with access to drugs that either had not been appraised by the National Institute for Health and Care Excellence (NICE) or had not been recommended for routine commissioning (204). In the original model of the CDF, there was an absence of clear entry and exit criteria for drugs. This created unsustainable financial pressure without evidence of patient benefit (26,28,205). In 2016, the CDF was revised to provide a more sustainable approach to funding promising new drugs and to collecting additional clinical data (27).

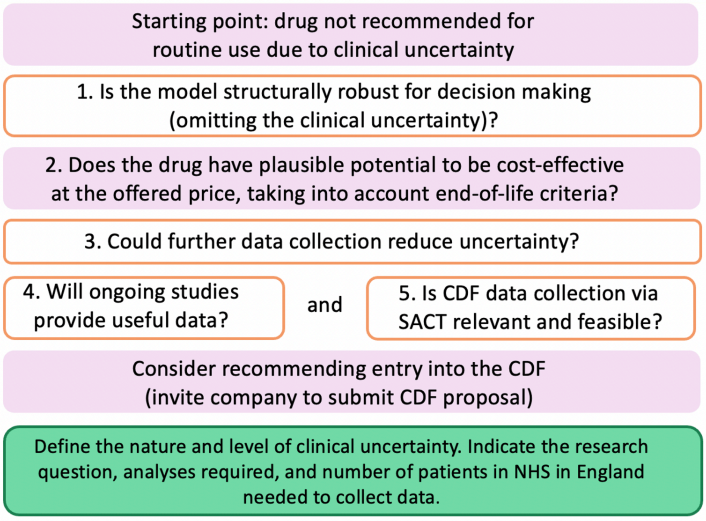
Since the reform of the CDF (from here 2016 CDF), all new oncology drugs are appraised by NICE. The 2016 CDF offers a mechanism for conditional approval. Figure 6.1 shows possible NICE recommendation options. If uncertainties regarding a drug are too great for it to be recommended for routine commissioning, a recommendation for use within the CDF can be considered (206). The appraisal committee uses the criteria in Figure 6.2 to decide which drugs are eligible to be used within the 2016 CDF (28). One of these is whether the clinical uncertainty can be addressed with additional data collected while the drug is provided through the CDF. If the appraisal committee recommends use within the CDF, a data collection arrangement (DCA) working group is formed with representation from NICE and NHS England. The DCA working group reviews the data collection proposal to translate the committee's uncertainties related to clinical outcome into defined data collection questions (29). Additional data are collected in line with the DCA and form the basis for the review appraisal of the case for routine commissioning which is expected to happen normally within two years (207).



(Based on NHS England Board Paper: PB.25.02.2016/04 Appendix 2)

* NICE: National Institute for Health and Care Excellence, CDF: Cancer Drugs Fund

Figure 6.1 Managed access scheme for new cancer drugs in 2016 Cancer Drugs Fund



CDF: Cancer Drugs Fund, SACT: Systemic Anti-Cancer Therapy

Source: NICE website (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/cancer-drugs-fund)

Figure 6.2 2016 Cancer Drugs Fund entry criteria

During this period, more evidence would be collected on the clinical effectiveness of the drug to resolve the key areas of uncertainty. The CDF review appraisal considers the data that have become available since the original appraisal, together with any change to the patient access scheme or commercial access arrangement proposed by the company. However, changes to the scope of the appraisal such as the population and comparators are not considered during CDF reviews (208). There are two main options for data collection, ongoing and new clinical trials and “real-world data” (RWD) from the Systemic Anti-Cancer Therapy (SACT) dataset (29). Other established cancer registries are also potential data sources for further review.

When introducing the 2016 CDF, a role for RWD, particularly SACT data was highlighted (29). SACT data are routinely collected, for patients receiving anti-cancer therapies from NHS England providers,

as a mandatory collection under the responsibility of Public Health England (PHE) (165). The SACT dataset is preferred for any data collection of routine chemotherapy practice in England because the infrastructure (including data protection and information governance) is already established, data are already being collected and progress can easily be monitored (29). PHE, through a cancer data partnership with the NHS, initially reported SACT data to support the re-appraisal of treatments provided by the 2016 CDF. NHS Digital took over this responsibility from PHE, on 1 October 2021, when the latter was replaced by the UK Health Security Agency and the Office for Health Improvement and Disparities.

The additional data collection is expected to address areas of uncertainty highlighted by the appraisal committee. In 2020, PHE indicated that "Real-world data reported by PHE is the primary information used to answer NICE uncertainty for 25% of CDF treatments" (30). This report did not say how RWD had been used as primary evidence to address the uncertainty issues. Understanding how such data are used is more important than simply counting appraisals reporting SACT data. Moreover, given the increasing interest in RWD, 25% of CDF treatments is a relatively low proportion, and the reasons for this low utilisation of RWD need to be reviewed.

Managed access agreements (MAAs) give opportunities to gather additional evidence which could help to reduce uncertainty when making a final decision. A review of the twenty-four CDF review appraisals completed to date can document the extent to which this objective has been met by collecting RWD. Moreover, it can identify the challenges and opportunities for use of RWD by NICE. It is timely to review experience with the 2016 CDF, because twenty-four drugs have now completed their re-appraisal, and a broadly similar fund entitled the Innovative Medicines Fund (IMF) has recently been introduced. This paper reviews the committee's recommendations following re-appraisal in order to obtain insight into the performance of the 2016 CDF. It focuses particularly on the uncertainties which led to drugs being provided through the CDF and on how clinical and cost-effectiveness evidence considered at the re-appraisal differed from that in the original appraisal.

We find that re-appraisals have largely resulted in recommendations for the routine commissioning of these drugs, which might suggest “Don’t think twice, it’s all right” as a maxim for these decision-makers. However, a detailed review of each re-appraisal indicates quite limited success in reducing the uncertainties which led to these drugs not being recommended for routine commissioning in the original appraisal. It also highlights the relative importance of longer follow-up of trial participants, compared to the contribution of RWD, in addressing some of the original uncertainties. Among different types of additionally collected data, a particular focus of this study is on the use of SACT data, which was highlighted when the 2016 CDF was introduced.

6.2 Methods

NICE technology appraisals (https://www.nice.org.uk/guidance) were determined to be eligible if they met the following criteria: 1) the drug was provided for the specific indication through the 2016 CDF following a managed access agreement made between NHS England and the manufacturer and 2) the NICE CDF review appraisal had been completed before 16 August 2022. As a result, twenty-four appraisals were identified for this review. The terminated appraisal (TA674 pembrolizumab) was included in this review. In this appraisal, the company decided not to make a case after the CDF review started, and consequently there was sufficient data available to include it in the review.

Data were extracted following a protocol developed to extract information about how RWD has been used in NICE appraisals of oncology medicines (117). This protocol enables a more comprehensive understanding of the use of RWD in CDF review appraisals by identifying non-parametric and parametric use of RWD in both the base-case and sensitivity analyses. Parametric use of RWD is where such data provide the numerical value of a specific variable in the economic model, whereas non-parametric use is where the data are used to develop the model structure or to support, corroborate or validate assumptions and/or choice of data used to parameterise the model. The

distinction is made to facilitate more consistent and comprehensive data extraction and to provide a means of measuring the intensity of the use of RWD in an appraisal.

As this data extraction tool was developed for a more general purpose, a few additional variables were required for the specific purposes of this study (Appendix 1.3). These variables capture references to additional data especially the SACT data in the CDF review appraisals and the uncertainties identified in the original appraisals and in the review appraisals. Information about uncertainties was extracted from the final appraisal determinations (FADs) in both the original and CDF review appraisals. Uncertainties were classified as either a 'key uncertainty', or 'other uncertainty', following Morrell et al. who reviewed the common types of uncertainty addressed in appraisals of drugs which entered the original CDF and discussed the potential for RWD to resolve these uncertainties (157). If an uncertainty was described in a section heading or highlighted in the conclusion or in the CDF consideration, this uncertainty was considered as a 'key uncertainty'. Any other uncertainty addressed across the appraisal was recorded as 'other uncertainty'. The uncertainty in CDF review appraisals was reviewed to assess how much additional data helped to reduce uncertainty. Three categories were used (still uncertain, uncertainty resolved, newly added uncertainty) by comparing the FADs from the original and the subsequent appraisal. Any comments about uncertainty made by the committee were recorded. Given that CDF review appraisals highlighted resolving uncertainty identified in the original appraisals, remaining uncertainties were usually addressed in review appraisals. If an uncertainty was not mentioned in the FAD of the review appraisal, it was classified as 'resolved.'

The original and CDF review appraisals were compared in terms of the data used, with particular emphasis on where the additional data came from to address the originally identified uncertainties. RWD was of particular interest because one of the arguments for having MAAs was that they provided opportunities to collect additional data, particularly from routine clinical use of the drug. Data were extracted from the main appraisal documents (final scope, company submission, evidence

review group (ERG) report, and FAD). Although most evidence used in decision-making were available in these documents, some parts of the evidence in CDF review appraisals were not fully described. When the assumptions made in original appraisals were followed in CDF reviews, the evidence for these assumptions were not fully described in review appraisal. This was often the case with resource use. In this research, evidence not mentioned in any of four main documents of the CDF review was assumed to be the same as that in the original appraisal. While this is a reasonable assumption, without access to the underlying economic evaluation models, it cannot be guaranteed that the evidence has not changed. Since this research was restricted to the analysis of data in the public domain, this was a potential limitation.

Another research question was whether the pattern of use of RWD changes or not. While drugs were provided through the CDF, companies could collect their own RWD. Additionally collected data could be used not only to reduce uncertainty but also to support their models with more recent evidence. While it might be anticipated that provision through the CDF would increase the opportunities to use RWD in assessing cost-effectiveness, it was possible that the availability of additional trial data reduced reliance on RWD. Hence, the pattern and intensity of use of RWD were reviewed to see whether these changed over the CDF process. Following the data extraction protocol, patterns were identified from both original and review appraisals. Use of RWD in three specific components of an economic evaluation are defined as major use of RWD (uses of RWD in estimating overall survival (OS) for either intervention and comparators, volume of treatment for either intervention and comparators and the choice of comparators). These components are likely to have a major impact on the estimated incremental cost-effectiveness ratio. Along with reviewing use of additional data in addressing identified uncertainties, this pattern and intensity review can give a more comprehensive picture of how NICE has used newly collected RWD in CDF reviews.

6.3 Results

6.3.1 Cancer Drugs Fund review recommendations

The recommendations made by the committee, reported in Table 6.1, were unchanged following re-appraisal in eighteen cases. In three further re-appraisals changes were minor. In the case of atezolizumab (TA739) the change was as a result of a changed marketing authorisation, and in two nivolumab appraisals (TA655 and TA713) guidance was further optimised by requiring no prior PD-1 or PD-L1 inhibitor treatment (reflecting changes in clinical practice). There were three cases where the treatments were not recommended for routine commissioning (TA674, TA692 and TA795). Thus, 87.5 per cent of re-appraisals resulted in recommendation that the treatment be routinely commissioned.

Table 6.1 Summary of characteristics of Cancer Drugs Fund review appraisals

	TA524	TA531	TA629	TA653	TA655
Technology	Brentuximab vedotin	Pembrolizumab	Obinutuzumab + bendamustine	Osimertinib	Nivolumab
Conditions	CD30-positive Hodgkin lymphoma	Untreated PD-L1 positive metastatic non-small-cell lung cancer	Follicular lymphoma refractory after rituximab	EGFR T790M mutation-positive advanced non-small-cell lung cancer	Advanced squamous non-small-cell lung cancer after chemotherapy
Recommendation for routine commissioning	Optimised	Recommended	Recommended	Optimised	Optimised
Date of FAD	04/2018	03/2018	03/2020	09/2020	09/2020
Original appraisal					
TA number	TA446	TA447	TA472	TA416	TA483
Date of FAD	04/2017	05/2017	07/2017	09/2016	09/2017
Key uncertainty in FAD	· Post-treatment stem cell transplant rates	· Immature survival data	· Immature survival data	· Immature survival data	· Immature survival data
Other uncertainties in FAD	· Overall survival and progression-free survival following stem cell transplant	· Duration of continued treatment effect · Appropraite HRQoL values	· Duration of continued treatment effect	· Appropraite HRQoL values · Indirect treatment comparison · Small number of participants after restricting the population	· Appropraite HRQoL values · Stopping rule · Duration of continued treatment effect · Methods of extrapolation for OS & PFS
Month between FADs	12	10	32	48	36
Evidence in CDF reviews					
Clinical effectiveness evidence	New observational data & SACT data	Clinical trial (KEYNOTE-024)	Clinical trial (GADOLIN)	Clinical trial (AURA pooled, AURA3*)	Clinical trial (CheckMate 017, CheckMate 003)
SACT					
PHE report in guidance	No	No	Yes, attached	Yes, attached	Yes, attached
Median follow-up (Mo)	.	.	12.4	Minimum 4 months	3.1
Number (n)	.	.	92	357	389
Use of SACT in economic evaluation					
Base case analysis (Parametric use)	· Estimate of stem cell transplant rates after treatment	.	.	.	.
Base case analysis (Non-parametric use)	.	.	.	.	· Treatment duration · Support trial data
Scenario/sensitivity analysis	.	.	.	.	.

* New data not submitted in original appraisal

** "Optimised" is a recommendation for a smaller group of patients than originally stated by the marketing authorisation. TA recommendations are aligned with NICE reports on (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations)

CDF: Cancer Drugs Fund, FAD: Final Appraisal Determination, HRQoL: Health related Quality of life, OS: Overall survival, PFS: Progression-free survival, SACT: Systemic Anti-Cancer Therapy, TA: Technology appraisal

(Continued, table 6.1 Summary of characteristics of Cancer Drugs Fund review appraisals in this review)

	TA674	TA683	TA684	TA687	TA691
Technology	Pembrolizumab	Pembrolizumab + pemetrexed & platinum chemotherapy	Nivolumab	Ribociclib + fulvestrant	Avelumab
Conditions	Untreated PD-L1 positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable	Untreated, metastatic, non-squamous non-small cell lung cancer	Adjuvant treatment of completely resected melanoma with lymph node involvement or metastatic disease	Hormone receptor-positive, HER2-negative advanced breast cancer after endocrine therapy	Untreated metastatic Merkel cell carcinoma
Recommendation for routine commissioning	Terminated**	Recommended	Recommended	Recommended	Recommended
Date of FAD	02/2021[a]	01/2021	02/2021	02/2021	03/2021
Original appraisal					
TA number	TA522	TA557	TA558	TA593	TA517
Date of FAD	04/2018	11/2018	11/2018	06/2019	03/2018
Key uncertainty in FAD	· Immature survival data · Indirect treatment comparison	· Immature survival data	· Immature survival data · Recurrence-free survival data	· Indirect treatment comparison (results of network meta-analysis) · Extrapolation of PFS	· Immature survival data · Indirect treatment comparison
Other uncertainties in FAD	· Duration of continued treatment effect	· Duration of treatment effect · Indirect treatment comparison	· Subsequent treatment · Clinical effectiveness of PD-1 inhibitors as adjuvant treatments	· Immature survival data · Statistical power as relevant population was subgroup in the trial. · Time-to-treatment stopping · Post-progression survival	· Small number of patients in the trial
Months between FADs	34	26	27	20	36
Evidence in CDF reviews					
Clinical effectiveness evidence	Clinical trial (KEYNOTE-361)*	Clinical trial (KEYNOTE-189)	Clinical trial (CheckMate 238)	Clinical trial (MONALEESA-3)	Clinical trial (JAVELIN MERKEL 200 trial: Part B)
SACT					
PHE report in guidance	Yes, attached	No	Yes, attached	Yes, attached	Yes, attached
Median follow-up (Mo)	4.2	.	5.0	3.7	6.0
Number (n)	61	.	299	221	52
Use of SACT in economic evaluation					
Base case analysis (Parametric use)	.	.	.	.	.
Base case analysis (Non-parametric use)	.	.	.	.	· Treatment duration
Scenario/sensitivity analysis	.	.	· Subsequent treatment · Time gap between treatments	.	.

* New data not submitted in original appraisal, **Unable to make recommendation due to the company's withdrawal
CDF: Cancer Drugs Fund, FAD: Final Appraisal Determination, PFS: Progression-free survival, SACT: Systemic Anti-Cancer Therapy, TA: Technology appraisal

(Continued, table 6.1 Summary of characteristics of Cancer Drugs Fund review appraisals in this review)

	TA692	TA713	TA725	TA736	TA739
Technology	Pembrolizumab	Nivolumab	Abemaciclib +fulvestrant	Nivolumab	Atezolizumab
Conditions	Locally advanced or metastatic urothelial cancer after platinum-containing chemotherapy	Advanced non-squamous non-small-cell lung cancer after chemotherapy	Hormone receptor-positive, HER2-negative advanced breast cancer after endocrine therapy	Recurrent or metastatic squamous cell carcinoma of the head and neck after platinum-based chemotherapy	Untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable
Recommendation for routine commissioning	Not recommended	Optimised**	Recommended	Recommended	Recommended***
Date of FAD	03/2021	05/2021	08/2021	09/2021	09/2021
Original appraisal					
TA number	TA519	TA484	TA579	TA490	TA492
Date of FAD	02/2018	09/2017	03/2019	09/2017	10/2017
Key uncertainty in FAD	· Immature survival data	· Methods of extrapolation for OS & PFS	· Immature survival data	· Effectiveness according to PD-L1 expression · Long-term OS	· Indirect treatment comparison (single-arm)
Other uncertainties in FAD	· Duration of continued treatment effect	· Appropriate HRQoL values · Duration of continued treatment benefit · Clinical stopping rule · Absence of indirect treatment comparison with relevant comparator (BSC)	· Time on treatment · Heterogeneity of network analysis due to absence of direct treatment comparison	· Generalisability of the trial to the UK practice · Appropriate time point to extrapolate the trial data · Underestimated 2yr PFS · Time on treatment · Duration of continued treatment effect · Stopping rule · Appropraite HRQoL values	· Immature survival data · Small number of patients · Duration of continued treatment benefit · Time on Treatment · Appropraite HRQoL values · Effectiveness for PD-L1 subgroup
Months between FADs	37	44	29	48	47
Evidence in CDF reviews					
Clinical effectiveness evidence	Clinical trial (KEYNOTE-045)	Clinical trial (CheckMate 057, CheckMate 003)	Clinical trial (MONARCH 2)	Clinical trial (CheckMate 141)	Clinical trial (Imvigor130*)
SACT					
PHE report in guidance	Yes, attached	Yes, attached	Yes, attached	No	Yes, attached
Median follow-up (Mo)	Not available	4.2	4.4	.	9.6
Number (n)	102	43	298	.	64
Use of SACT in economic evaluation					
Base case analysis (Parametric use)	.	.	.	.	.
Base case analysis (Non-parametric use)	.	· Support trial data	· Treatment duration · Support trial data	· Support trial data	· Supporting a choice of survival curve
Scenario/sensitivity analysis	.	.	.	.	.

* New data not submitted in original appraisal
** "Optimised" is a recommendation for a smaller group of patients than originally stated by the marketing authorisation. TA recommendations are aligned with NICE reports on (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations)
***Indication has been changed from original
CDF: Cancer Drugs Fund, FAD: Final Appraisal Determination, HRQoL: Health related Quality of life, OS: Overall survival, PFS: Progression-free survival, SACT: Systemic Anti-Cancer Therapy, TA: Technology appraisal

(Continued, table 6.1 Summary of characteristics of Cancer Drugs Fund review appraisals in this review)

	TA766	TA770	TA780	TA783	TA784
Technology	Pembrolizumab	Pembrolizumab	Nivolumab + ipilimumab	Daratumumab	Niraparib
Conditions	Adjuvant treatment of completely resected stage 3 melanoma with lymph node involvement	Untreated metastatic squamous non-small-cell lung cancer	Untreated advanced renal cell carcinoma	Relapsed and refractory multiple myeloma	Relapsed, platinum-sensitive high-grade serous epithelial ovarian, fallopian tube or primary peritoneal cancer
Recommendation on routine commissioning	Recommended	Optimised*	Recommended	Optimised*	Optimised*
Date of FAD	12/2021	12/2021	01/2022	01/2022	02/2022
Original appraisal					
TA number	TA553	TA600	TA581	TA510	TA528
Date of FAD	11/2018	05/2019	03/2019	12/2017	05/2018
Key uncertainty in FAD	· Immature survival data · Distant metastases-free survival	· Immature survival data	· Immature survival data	· Indirect treatment comparison	· Immature survival data · More complete understanding of who would benefit most from treatment using somatic and other testing
Other uncertainties in FAD	· Duration of continued treatment effect · Reuse of pembrolizumab after use in the adjuvant setting	· Time to treatment discontinuation · Subsequent treatment	· Proportion of people with intermediate- and poor-risk disease	· Long-term survival benefit · Subsequent treatment	· Duration of treatment
Months between FADs	37	28	34	49	39
Evidence in CDF reviews					
Clinical effectiveness evidence	Clinical trial (KEYNOTE-054)	Clinical trial (KEYNOTE-407)	Clinical trial (CheckMate 214)	Clinical trial (MMY2002)	Clinical trial (NOVA)
SACT					
PHE report in guidance	Yes, attached	No	Yes, attached	Yes, attached	Yes, attached
Median follow-up (months)	15.7	.	10.8	4.3	13.7 (germline BRCA mutation), 5.7 (no germline BRCA mutation)
Number (n)	1,324	.	814	2,301	157 (germline BRCA mutation), 859 (no germline BRCA mutation)
Use of SACT in economic evaluation					
Base case analysis (Parametric use)	· Volume of subsequent treatment	.	.	· Volume of subsequent treatment	.
Base case analysis (Non-parametric use)	· Validation of OS projection · Treatment sequence (subsequent treatment)	.	.	· Time-to-discontinuation · Support trial data	.
Scenario/sensitivity analysis	· Characteristics of population	.	· Subsequent treatment	· OS	· Time to discontinuation

* "Optimised" is a recommendation for a smaller group of patients than originally stated by the marketing authorisation. TA recommendations are aligned with NICE reports on (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations)

CDF: Cancer Drugs Fund, FAD: Final Appraisal Determination, HRQoL: Health related Quality of life, OS: Overall survival, SACT: Systemic Anti-Cancer Therapy, TA: Technology appraisal

(Continued, table 6.1 Summary of characteristics of Cancer Drugs Fund review appraisals in this review)

	TA795	TA796	TA798	TA802
Technology	Ibrutinib	Venetoclax	Durvalumab	Cemiplimab
Conditions	Waldenstrom's macroblobulinaemia	Chronic lymphocytic leukaemia	Maintenance treatment of unresectable non-small-cell lung cancer after platinum-based chemoradiation	Advanced cutaneous squamous cell carcinoma
Recommendation on routine commissioning	Not recommended	Recommended	Recommended	Optimised **
Date of FAD	04/2022	04/2022	05/2022	05/2022
Original appraisal				
TA number	TA491	TA487	TA578	TA592
Date of FAD	09/2017	09/2017	03/2019	05/2019
Key uncertainty in FAD	· Pre-progression mortality	· Patients characteristics · Indirect treatment comparison	· Immature survival data	· Immature survival data · Indirect treatment comparison
Other uncertainties in FAD	· Long-term survival benefit · Indirect treatment comparison	· Long term survival after B-cell receptor pathway inhibitors	· Duration of continued treatment effect	· Dosing regimens
Months between FADs	52	56	38	36
Evidence in CDF reviews				
Clinical effectiveness evidence	SACT dataset	SACT dataset	Clinical trial (PACIFIC)	Clinical trial (Study 1423, EMPOWER CSCC-1)
SACT				
PHE report in guidance	Yes, attached	Yes, attached	Yes, attached	Yes, attached
Median follow-up (months)	12.9	20.6	7.3	10.2
Number (n)	859	153 (Non-del(17p)/TP53 mutation), 218 (Del(17p)/TP53 mutation)	591	352
Use of SACT in economic evaluation				
Base case analysis (Parametric use)	· Overall survival (OS) · PFS (using treatment duration as a surrogate)	· Overall survival (OS) · PFS (using treatment duration as a surrogate)	.	.
Base case analysis (Non-parametric use)	· Characteristics of population	· Characteristics of population	.	. Support trial data
Scenario/sensitivity analysis	· Pre-progression mortality	.	.	· Characteristics of population

* New data not submitted in original appraisal
** "Optimised" is a recommendation for a smaller group of patients than originally stated by the marketing authorisation. TA recommendations are aligned with NICE reports on (https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/data/appraisal-recommendations)
CDF: Cancer Drugs Fund, FAD: Final Appraisal Determination, SACT: Systemic Anti-Cancer Therapy, TA: Technology appraisal

6.3.2 Key uncertainties addressed in the original appraisals

The uncertainties reported in the FADs of the original appraisals are shown in Table 1. Immature survival data in the clinical trials were identified as a common source of uncertainty in most appraisals (67% of appraisals; as key uncertainty, 10% of appraisals; as other uncertainty). These data increased uncertainty around the size of clinical benefits or long-term benefits. In ten appraisals (42%), an indirect treatment comparison was a source of uncertainty when assessing the clinical benefits. Among them, six appraisals identified this as a key uncertainty. Indirect treatment comparisons were made because of an absence of RCTs or because relevant comparators were not included in a single RCT. Another source of uncertainty was how clinical effectiveness varied across subgroups defined by the expression of PD-L1. Duration of treatment effect, time on treatment and health related quality of life (HRQoL) values were frequently noted as sources of uncertainty but were not identified as key uncertainties in many appraisals.

6.3.3 Use of additional data in economic evaluation in Cancer Drugs Fund reviews

Additional data from clinical trials

The average time gap between the publication of the FADs for the original appraisal and the CDF review was 35.6 months (median 36 months). The main evidence for economic evaluation in CDF review appraisals was from clinical trials. The additional data in seventeen CDF review appraisals came from further follow up of patients in the trials featured in the original appraisal. Two CDF review appraisals (TA674, TA739) used data from clinical trials which were not presented in the original appraisals. Another appraisal (TA653) used both later follow-up of a trial and new clinical trial data. Three appraisals (TA524, TA795, TA796) used SACT data along with previously used data as the main evidence for the economic evaluation model.

Since the information about median follow-up was redacted in a few appraisals, five appraisals were excluded to estimate the increase in the duration of follow-up. The average increase in median

follow-up between the original appraisal and the CDF review appraisal was 22.2 months (median 22.2 months). The longest increase in median follow-up was 50 months in the appraisal of niraparib (TA784). Two review appraisals (TA629, TA770) reported increases in median follow-up of 6 months.

Additional real-world data

SACT data were the predominant type of RWD used in CDF review appraisals. The use of SACT is reviewed in a separate section below. Here, RWD, *other than* SACT data are reviewed. There were three cases where RWD, *other than* SACT data were used in the CDF review but not in the original appraisal. In the CDF review appraisal of pembrolizumab (TA766), the company used the registry data (The Surveillance, Epidemiology, and End Results; SEER and American Joint Committee on Cancer; AJCC) as well as SACT data to validate the survival distribution. This appraisal also used the market share data for subsequent treatment lines in a scenario analysis. The appraisal of niraparib (TA784) used a chart review study for clinical outcomes with the comparator, routine surveillance data, in a scenario analysis. One of the uncertainties in the original appraisal derived from an indirect treatment comparison. The company used RWD in a scenario analysis to investigate the uncertainty around the indirect comparison, whereas they used data from another clinical trial for the base case analysis. The appraisal of cemiplimab (TA802) used a new retrospective chart review study for comparative evidence as the lack of comparative evidence was highlighted during original appraisal. However, the comparative effectiveness of cemiplimab remained highly uncertain due to the chart review lacking validity.

The CDF review appraisal of pembrolizumab (TA770) stopped using RWD when extrapolating OS. The company's original model was criticised due to missing information about the second line treatments. In the review appraisal, the company dropped these data and used more recent clinical trial data. The appraisal of ibrutinib (TA795) substituted the RWD used in the original appraisal with UK-based registry data, to help address the data gap (progression free survival; PFS) which SACT

couldn’t provide. Also, these data were used to estimate a rate of pre-progression mortality in a scenario analysis, which was one of the key uncertainties in the original appraisal.

Patterns of use of RWD in the original appraisals and in the CDF review appraisals were compared (Appendix 4.2). Although there were changes in use of RWD, limited use was made of RWD collected during the CDF period. Substantial changes in patterns of RWD use were not found. Consequently, the intensity of the use of RWD has not changed. In the CDF review of pembrolizumab (TA766), RWD were used more broadly for supporting diverse assumptions in the model such as validating survival extrapolation, informing subsequent treatment line and baseline age of population in the model. However, the intensity of use of RWD has not changed much as only one additional component (volume of subsequent treatment) was informed by RWD (in this case SACT data).

6.3.4 Use of SACT data in Cancer Drugs Fund review economic evaluations

This study focused on the use of SACT data in CDF review appraisals. SACT data were the most commonly used form of RWD in CDF reviews. Since data collection via SACT was a part of the MAAs, the primary source of additional RWD was substantially the SACT database. Although the SACT dataset was the major vehicle to collect RWD, its overall use was limited. SACT data were not used to update the economic evaluation model in 9 out of 24 CDF review appraisals (Table 6.2). The remaining fifteen appraisals made limited use of SACT data. SACT data, newly collected from CDF patients, were used more for non-parametric purposes (eleven appraisals) such as validation or corroboration of the model than for parametric purposes (five appraisals). SACT data featured in both non-parametric and parametric uses in four appraisals (TA766, TA783, TA795, TA796).

Table 6.2 Summary of use of systemic anti-cancer therapy data in Cancer Drugs Fund review appraisals

Type of use	Drug	TA umber
Parametric use	Brentuximab vedotin	TA524
	Pembrolizumab	TA766
	Daratumumab	TA783
	Ibrutinib	TA795
	Venetoclax	TA796
Non-parametric use	Nivolumab	TA655
	Avelumab	TA691
	Nivolumab	TA713
	Abemaciclib+fulvestrant	TA725
	Nivolumab	TA736
	Atezolizumab	TA739
	Pembrolizumab	TA766
	Daratumumab	TA783
	Cemiplimab	TA802
	Ibrutinib	TA795
	Venetoclax	TA796
Used in sensitivity/scenario analysis	Nivolumab	TA684
	Pembrolizumab	TA766
	Nivolumab+ipilimumab	TA780
	Daratumumab	TA783
	Niraparib	TA784
	Cemiplimab	TA802
	Ibrutinib	TA795
Not used	Pembrolizumab	TA531
	Obinutuzumab+bendamustine	TA629
	Osimertinib	TA653
	Pembrolizumab	TA674
	Pembrolizumab	TA683
	Ribociclib+fulvestrant	TA687
	Pembrolizumab	TA692
	Pembrolizumab	TA770
	Durvalumab	TA798

CDF: Cancer Drugs Fund, SACT: Systemic Anti-Cancer Therapy

Parametric use

Five cases of parametric use were identified. In the CDF review of brentuximab vedotin (TA524), the company used CDF data to inform the rate of subsequent stem cell transplant following treatment with brentuximab vedotin. This was one of the key clinical uncertainties, which was expected to be resolved during the CDF period. A questionnaire sent to consultants identified the rates of stem cell transplant in patients who had brentuximab vedotin as part of the original CDF between April 2013

and March 2016. Another example was the CDF review appraisal of pembrolizumab (TA766). The company used SACT data for the distribution of subsequent treatments administered in the advanced setting for patients in the adjuvant pembrolizumab arm as clinical evidence was incomplete and SACT data were the best available real-world data to reflect the clinical practice observed in the CDF. Similar to TA766, in the CDF review appraisal of daratumumab (TA783), SACT data were used to inform subsequent therapies for all comparators. The appraisal of ibrutinib (TA795) has used SACT data as primary clinical evidence in an economic evaluation model. In the original appraisal, the company used a single-arm trial, Study 1118E for clinical outcome. Longer-term clinical effects were highly uncertain due to the limited long-term data. In the CDF review, the company revised their base-case analysis using SACT data to calibrate OS for the transition probability (post-progression mortality). Since SACT data did not record disease progression data, the company used other source of RWD to estimate progression-free survival (PFS). Here, treatment duration from SACT data was used to adjust the hazard compared with PFS. The appraisal committee concluded that there was considerable uncertainty around the most appropriate approach to estimating PFS of ibrutinib although an indirect approach to estimate PFS was reasonable. The CDF review appraisal of venetoclax (TA796) also used SACT data as primary clinical evidence in the economic model. Parametric models for OS were explored using SACT data. In this review appraisal, the company assumed that PFS was equivalent to the duration of venetoclax treatment. During the appraisals, the committee concluded that the assumption regarding PFS was plausible and that SACT data was the best available and was acceptable to represent venetoclax efficacy.

Non-parametric use

Non-parametric use of the SACT dataset has been made in eleven CDF review appraisals. Five forms of non-parametric use, informing characteristics of the study population, updating the subsequent treatment line, validation of survival outcome, treatment duration and corroboration of survival data, were identified. Two CDF reviews used SACT data to validate the choice of survival curves in the

model (TA739, TA766). In both original appraisals, extrapolation of the survival data was highly uncertain. Updated clinical trial data directly informed the estimates of OS in the economic evaluation model. The clinical plausibility of the survival distribution selected in the model in the CDF review appraisal was checked with SACT data. The duration of treatment was reviewed in four appraisals (TA655, TA691, TA725, TA783) by seeing to what extent SACT data were aligned with the trial data. This informed the discussion of the generalisability of the trial data to routine clinical practice in NHS England but did not inform the estimates of time-on-treatment directly. In six appraisals (TA655, TA713, TA725, TA736, TA783, TA802), SACT data were used to corroborate the clinical trial evidence. Median OS in SACT data and the overlaid survival curves were usually presented to support the trial data. There was one appraisal where SACT data were used to update the subsequent treatment line in the base case analysis (TA766) and two appraisals (TA795, TA796) where SACT data were used to inform the characteristics of the study population.

Use of SACT data in sensitivity/scenario analyses

SACT data were used in six CDF reviews (TA684, TA766, TA780, TA783, TA784, TA802), to explore the impact of alternative assumptions in sensitivity or scenario analyses. In one appraisal (TA784), the company used the time-to-discontinuation in SACT data at the request of NHS England. The company used the SACT data in a scenario analysis but not in the base case economic model, due to limited availability of baseline characteristics in the SACT database.

SACT not used

Evidence from SACT was not used to either update the economic model or support the evidence in nine appraisals. Three patterns of non-use of SACT were identified. In pattern 1, no information on SACT data was reported in the appraisal documentation nor was the PHE report uploaded (TA531, TA683, TA770). In pattern 2, SACT data was attached, but were not reported in the company submission (TA674, TA692). In pattern 3, the company submission reported SACT data and the PHE reports were attached (TA629, TA653, TA687, TA798), but the SACT data were neither used as

corroboration nor used directly in the economic evaluation model. The small number of patients and the limited follow-up periods were given as reasons for not using the SACT data.

6.3.5 Assessment of the extent to which additional data reduced the original uncertainties

In the CDF review, the technical engagement process was important to discuss the methods with which to deal with uncertainties. Technical engagement is a step where companies get a technical report from the NICE technical team and have a chance to mitigate the remaining uncertainties in the evidence base before appraisal committee meetings (209). In this process, discussion between ERGs and companies is also allowed. Companies have an opportunity to improve their evidence through this engagement.

Although the technical engagement could help to reduce the methodological challenges, some uncertainties remained. Data from new trials and later follow-up of existing trials were important when it comes to resolving these uncertainties. Uncertainty around immaturity was addressed by clinical trials which had further follow-up. However, later analysis of clinical trials could not solve all immaturity issues. Committees in three review appraisals (TA531, TA684, TA766) still had concerns about the immaturity of survival data. Although the clinical trial captured survival events over a longer period, choice of parametric model to predict OS was highly uncertain in five appraisals (TA655, TA683, TA687, TA692, TA713).

Uncertainty around survival benefit due to indirect treatment comparison was resolved by clinical trials when new randomised controlled trials (RCTs) were available. When the original appraisal was based on a single-arm trial while RCTs were ongoing, the review appraisal updated the model based on new phase 3 trials (TA492, TA519). However, if the RCTs didn't include all relevant comparators, clinical trials had limited scope to reduce uncertainty coming from indirect treatment comparisons. Commonly unresolved uncertainties in CDF review appraisals were the duration of continued treatment effects and the best utility values to use. It was common to use the assumptions

previously preferred by committee. Also, clinical experts' opinions were often used to discuss these issues.

The SACT dataset has rarely been actively used to deal with uncertainties because SACT data were not regarded as robust enough for use in the economic evaluation. A few review appraisals (TA629, TA691, TA725, TA766, TA784) directly indicated that the SACT data were too immature. As later clinical-trial data were available, SACT data were less relevant to address the uncertainty around immature data. However, SACT data have provided useful information such as time-to-treatment discontinuation and subsequent treatment. For example, in the CDF review appraisal of TA581 (TA780), one of the uncertainties was answered by SACT data. The committee preferred to use the proportions based on SACT data to weight the effectiveness estimates by risk group in the clinical trials as the SACT data were expected to inform the true proportion.

6.4 Discussion

The central findings of this study of experience to date with CDF review appraisals are the limited role played by SACT data and the importance of longer follow-up of the patients in the clinical trials upon which the original appraisals were based. Reasons for these key features of the review appraisals are not hard to find. The additional data available from SACT is limited in several respects - SACT data are not randomised, and survival data are generally immature given the period during which the CDF provided the treatment. The value of the SACT data may be further limited by the number of patients included and the information recorded. The former is also a direct consequence of the timetable chosen for the CDF review.

The use of clinical trial data in preference to SACT data is partly because the latter are not randomised. Comparisons of SACT data with other groups of patients in terms of progression-free survival and overall survival potentially introduces bias because of differences between patient groups in the distribution of effect modifiers (23,210). However, not all the trials used in the original

appraisals were randomised trials. In such cases, this limitation of SACT data is less important. For example, in the recent re-appraisal of ibrutinib for treating Waldenstrom's macroglobulinaemia (TA795) the committee concluded that the SACT data (n=823) were more relevant than updated trial data from the single arm study 1118E (n=63) and the iNNOVATE arm C (n=31).

The number of patients available for analysis is often smaller and the length of patient follow-up is shorter in the SACT database than in the original clinical trial. For example, in the re-appraisal of avelumab (TA691), the number of trial participants exceeded those in the SACT data (n=116 versus n=52), also median follow-up in JAVELIN was 16 months versus 6 months in the SACT database. Also, the data required for the economic model is more often available from the trial rather than from the SACT database. Potentially important model inputs such as PFS, health-related quality of life and response rate are not available in the SACT database (211).

Latimer suggests that the problem lies not just with the SACT database itself but is in part a failure to exploit the analytical opportunities these data offer (212). In reviewing the early entrants to the 2016 CDF, he notes that little information was given as to how the SACT dataset would be analysed. In recent CDF review appraisals, TA795 and TA796, SACT data have been used to a greater extent for OS and PFS estimation through active technical engagements and exploring the plausible ways of using the data. A more coherent analytical plan for assessing comparative effectiveness could facilitate better use of SACT data to support the reduction of uncertainties (212,213).

It is important to stress that this paper reviews experience with the first twenty-four drugs to exit the 2016 CDF. It accurately documents this recent experience. It is not claiming that SACT data (or other RWD) cannot play a major role in resolving the clinical uncertainties which have in turn contributed to uncertainty regarding the cost-effectiveness of many new oncologic drugs. The claim is simply that to date the contribution to resolving clinical uncertainty has been modest. More detailed planning for future analysis and longer periods of data collection might both increase the potential contribution of SACT data. It is noteworthy that the consultation over the Innovative Medicine Fund

(IMF) (a recently introduced sister fund to the CDF for non-oncologic medicines, made reference to provision for a period not exceeding five years (214), as does the recent NICE process and methods manual.

A review of the operation of the 2016 CDF is particularly relevant since NHS England is expanding the use of managed access schemes with the introduction of the IMF. While, it is likely that the IMF will operate in a similar fashion to the CDF, it will support "patients with any condition, including those with rare and genetic diseases, to get early access to the most clinically promising treatments where further data are needed to support NICE make recommendations with respect to routine commissioning by the NHS" (10). Consideration of experience with the CDF can aid understanding of the opportunities and challenges of using additional data to address uncertainties.

Although the use of RWD in CDF review appraisals is an institution-specific issue, the use of RWD in drug appraisals is of more general interest. The Italian Medicines Agency (AIFA) monitoring platform of registries track eligible patients and a complete flow of treatment to evaluate the appropriate use of drugs following their approval in the Italian national health system (215). The data collected are useful sources for verifying the real impact of the initial reimbursement criteria (216). In Dutch HTA reports, RWD have been used for initial decision-making. In conditional financing, a type of MAA, use of RWD to reduce uncertainty has attracted attention (217). However, a detailed analysis of the utilisation of different forms of RWD in different HTA systems is beyond the scope of this paper.

This paper has focused on the uncertainties in the original appraisal and the additional data considered at the re-appraisal. It has not sought to assess the success or otherwise of the CDF. Patients have had access to these twenty-four therapies through the CDF following an initial decision not to recommend routine commissioning. Moreover, following re-appraisal twenty-one have moved to routine commissioning. In addition, while in the CDF the drugs have had a price which is deemed cost-effective given the available evidence. An alternative perspective might be that the original clinical uncertainties do not appear to have been markedly reduced and still the re-appraisals have

been overwhelmingly positive. Possibly suggesting that CDF review appraisals should be regarded as a “review to ensure that the original decision is consistent with the latest evidence”, rather than as a “final chance to make the case”. However, before accepting Bob Dylan’s rejection of re-appraisal and the re-assurance that NICE committees can generally make the correct decision at the first attempt, it is important to recognise that any assessment of the value of the CDF needs to make a judgment regarding the counter-factual, including how the existence of the CDF might be influencing committees’ decision-making and manufacturers’ research activities and pricing decisions.

6.5 Conclusion

While additionally collected RWD attracted attention when the 2016 CDF was introduced, RWD were not widely used in CDF review appraisals and (to date) do little to reduce uncertainty. Experience with these appraisals has highlighted the importance of longer follow-up of clinical trials and the relatively limited role of RWD, in general, and SACT data in particular. Although the 2016 CDF, with its MAAs, is a clear improvement on the original CDF, the extent to which the clinical uncertainties have been resolved by additional data is unclear.

6.6 Key points for decision-makers

· When uncertainties regarding the clinical evidence have been too great for NICE to recommend routine commissioning, managed access agreements have allowed patients to be treated while additional data are collected.

· Immature survival data are an important source of clinical uncertainty which has largely been addressed by later follow-up of patients in clinical trials rather than by additional real-world data.

· Systemic Anti-Cancer Therapy data (an important English source of real-world data) have been used to address a limited number of clinical uncertainties.

Declaration

Funding/support

Jiyeon Kang is supported by the Centre for Cancer Biomarkers, University of Bergen funded by the Research Council of Norway grant number (223250). The funder is not involved in any aspect of the study conduct or the decision to submit the paper for publication.

Author contributions

Both authors contributed to conceptualising and designing the study. Jiyeon Kang analysed the data and drafted the manuscript. John Cairns revised the manuscript for important intellectual content and contributed to the methodology.

Conflicts of interest

None declared.

Ethics approval

This study was approved by the Ethics Committee of the London School of Hygiene and Tropical Medicine on 14 November 2019 (17315).

Consent to participate

This study used publicly available data. The consent to participate is not applicable in this study.

Consent for publication

Not applicable.

Availability of data and material

The data analysed during the current study are available in the National Institute for Health and Care Excellence website:

https://www.nice.org.uk/guidance/published?ngt=Technology%20appraisal%20guidance&ndt=Guidance

Code availability

Not applicable.

Chapter 7. Discussion

7.1 Research objective

Despite recognition as the gold standard of scientific evidence, it is challenging to collect all the evidence required for a drug appraisal from RCTs (85,218). These evidence gaps are an important source of uncertainty in payers' decision-making. In this context, real-world data (RWD) are often regarded as a potential solution to these challenges. However, there is a shortage of comprehensive and systematic discussion about using RWD in drug appraisals. This thesis aimed to contribute an in-depth understanding of the previous use of RWD. The primary focus of this research was to highlight patterns and intensity of use of RWD rather than simply counting instances of its use. The identified patterns and intensity of the use of RWD were analysed in several ways. Mainly, associations were investigated between the use of RWD and a set of factors related to the sources of uncertainty. This thesis also emphasised the sources of uncertainty often found in appraisals of oncology medicines and how patterns/levels of intensity of use of RWD varied across the appraisals. The final paper in the thesis explored the role of RWD, focusing on the Systemic Anti-Cancer Therapy (SACT) data to reduce uncertainties in CDF review appraisals. This concluding section will highlight some findings, summarise the contributions and limitations of this thesis and suggest policy implication and future research.

7.2 Key findings

Interestingly, there is evidence of using RWD in appraisals of oncology medicines since 2011. Also, RWD have been used in diverse parts of the economic models, such as validating the choice of survival distribution and estimating overall survival and resource use. However, there was no dominant pattern of use of RWD. This may indicate lack of agreement over the best ways to use RWD in economic models. When reviewing intensity, using three major uses of RWD (use for choice of comparators, estimating overall survival, and estimating volume of treatment), relatively low levels

of intensity of use were often observed. The potential bias of RWD due to non-randomisation can be one of the reasons that hold back its use in these three major components. Although the use of RWD was restricted for these three components in past appraisals, it appears that RWD are increasingly used over time.

Notably, some sources of uncertainty were associated with the use of RWD. The absence of direct treatment comparison was frequently associated with greater use and a higher level of intensity of use of RWD. When the evidence for comparators does not exist, use of RWD in economic modelling is more likely. This was an intuitive finding as various sources of evidence, such as non-randomised studies, are used in appraisals when RCTs are unavailable. Another source of uncertainty, the maturity of survival data, had a statistical association with using RWD to estimate overall survival. When the survival data were immature, RWD supplemented the information for diverse purposes, including adjusting background mortality and change of the hazard of disease or extrapolating the survival curve.

The sources of uncertainty in appraisals and the use of RWD were reviewed in appraisals highlighting targeted cancer therapies (TCT). While the appraisal of TCT could be generally challenging, the challenges identified were neither new nor distinctive. It appears that the uncertainty in appraisals of TCT was derived from insufficient data rather than the characteristics of the drugs. Although RWD might be expected to play a more active role in appraisals of TCT, the use of RWD has generally been limited.

Cancer Drugs Fund (CDF) review appraisals were investigated to identify the common sources of uncertainty and to what extent RWD were used to reduce the uncertainty. The common source of uncertainty found in CDF appraisals was immature survival data, challenging the estimation of long-term effects. While additionally collected RWD attracted attention, RWD have not been widely used in CDF review appraisals and have done little to reduce uncertainty.

7.3 Contributions of this research

This thesis differs from previously published studies by reviewing STAs published in the last 11 years following a systematically applied extraction tool and by exploring the use of RWD in diverse ways, such as characterising the intensity and analysing the likelihood of increased use of RWD. There are several ways in which this research contributes to an in-depth understanding of the use of RWD in the appraisals of cancer drugs (Figure 7.1).

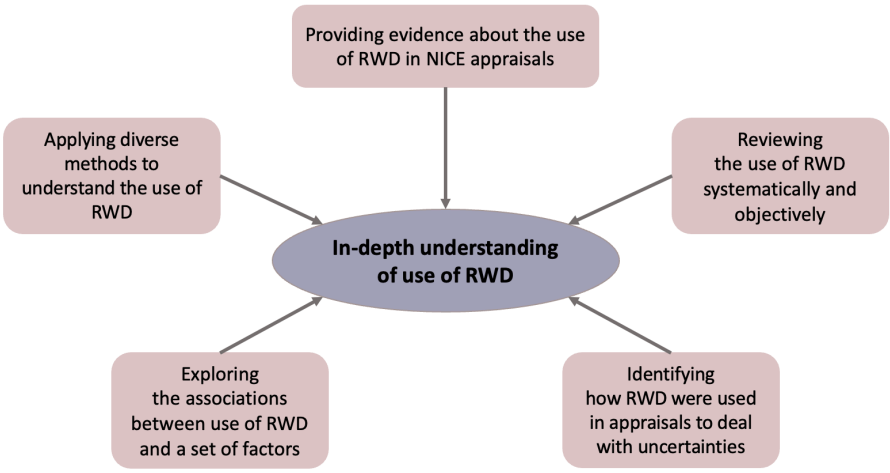


Figure 7.1 Contributions of this research

This research has contributed to establishing comprehensive evidence about the use of RWD in NICE appraisals of oncology medicine. Some research findings were intuitive and empirically established but have not been adequately described in the literature. A review of 229 NICE STAs of oncology medicines brought an opportunity to understand the various use of RWD in NICE decision-making in diverse ways and document the results transparently. This thesis provided confirmatory written evidence about the use of RWD in NICE appraisals.

This research reviewed the use of RWD in NICE appraisals systematically and objectively, following a detailed data extraction protocol. While a few studies have reviewed the use of RWD in healthcare decision-making, few studies review the use of RWD in a systematic manner. The data extraction protocol described in Chapter 2 had two essential features to ensure that data were extracted comprehensively and unbiasedly. First, the protocol covers critical components of economic

evaluation finely divided to improve the reliability of data extraction. It also distinguished non-parametric and parametric use of RWD. This separation allowed the question 'how has NICE incorporated a broad range of evidence in appraisals of oncology medicine' to be answered robustly. Second, the protocol was validated by two independent researchers. The reproducibility of the data extraction and clarity of the instructions were checked. These features facilitate having a dataset systematically and objectively collected.

This thesis undertook new approaches in describing the use of RWD in NICE appraisals using systematically collected data. The two different methods – a review of the patterns and an analysis of the intensity of use of RWD - were applied to improve understanding of current use of RWD in NICE appraisals. The review of patterns provided further information about changes in use over time and by type of cancer. An analysis of the intensity of the use of RWD explored the trends in the use of RWD for three major components. Given the increasing interest in leveraging RWD for comparative treatment effects, this analysis reflects the current interest in RWD and allows a thorough understanding of the present use of RWD.

Along with these descriptive analyses, various regression models were estimated to test hypotheses regarding the associations between the use of RWD/the intensity of use of RWD and a range of factors potentially related to the sources of uncertainty. While some literature (114–116,164) has identified where RWD can be used, this thesis makes a first attempt to explain these associations using regression analysis. This statistical analysis provides more robust evidence about the association with use of RWD than do case studies. Moreover, by considering several different factors related to the sources of uncertainty or availability of data in the analysis, the regression model described the associations more clearly.

In addition, this thesis demonstrated the extent to which RWD can reduce the uncertainties found in economic models during appraisals. No other studies have yet reviewed how additionally collected data helped to reduce uncertainty in the CDF review appraisals. Although the role of RWD in

supplementing evidence gaps was emphasised in some literature (52,92,162,219), there has been little literature about the degree to which RWD reduces uncertainty in decision-making. The analysis of the CDF review appraisals thoroughly explored how additionally collected RWD were used and helped to reduce uncertainty addressed in the original appraisals.

7.4 Limitations

Despite my best efforts, this thesis has several limitations which may impact on the data and analyses used in this thesis (Figure 7.2).

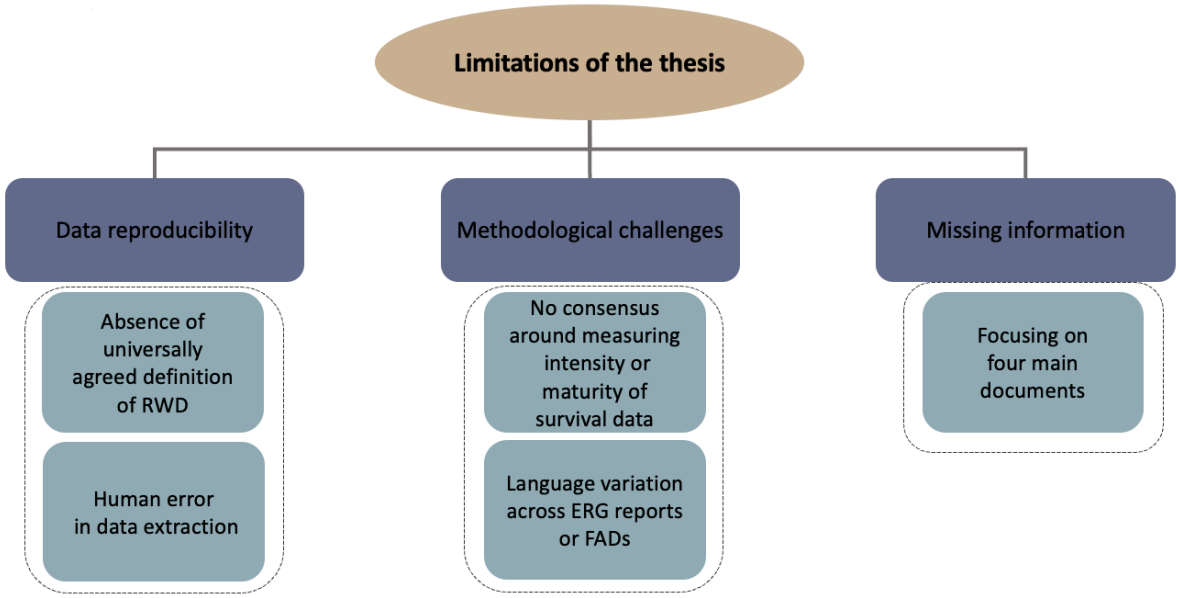


Figure 7.2 Research limitations

The main limitation of this research is related to data reproducibility. This thesis did not use an existing dataset. A new dataset was created following the data extraction protocol. This protocol was specially developed for reproducible data extraction about the use of RWD in the economic model of oncology medicine in NICE appraisals. Although the extraction results were consistent in most parts, the validation identified different results in some variables due to the different recognition and interpretation of RWD, missing information and human error. This result flags a caveat that the findings of this thesis might not be fully reproducible.

The issue of data reproducibility regarding the use of RWD is heightened by the absence of a universally agreed definition of RWD. Although the protocol defined RWD, grey areas remained in determining what is RWD. This increases the risk of different interpretations of the definition of RWD. Also, human error in data extraction was unavoidable. Although the data extraction protocol was developed to minimise biased extraction, human error can be caused without intention. In this thesis, all data extraction was conducted manually. While manual extraction allows one to review whole documents carefully and identify the information that is not clearly stated in some areas, it leaves a chance of erroneous data extraction due to simple mistakes. However, human errors are not expected to have a substantial impact on data quality. These errors were likely reduced over time as the researcher trained more from the repetitive extraction practice.

This thesis has some methodological challenges. First, there is no consensus on classifying the variables used in the analyses. The classifications for the intensity of use of RWD and for the maturity of survival data were newly established. They were based on a review of the literature and on interviews with key stakeholders in NICE appraisals, but different researchers could reasonably differ as to the most appropriate classification. The research findings might vary depending on the classifications used. Another challenge is a variation which can be possibly observed across evidence review groups (ERGs) or appraisal committees. With respect to the sources of uncertainty, the data were mainly extracted from ERG reports and Final Appraisal Determinations (FADs). Each ERG and appraisal committee follows the appraisal guide and shares common grounds on appraisals of new technology. While the ERG documents and FADs are substantially coherent, their language can differ slightly. However, this research did not have a process to capture potential differences in language and perspective between committees/ERGs in appraisals. Differences in language might result in different classifications of the uncertainties.

Another limitation is missing information. Due to human error, some data could be missing. Also, some data about use of RWD might not be available in the four main documents. The data on the use

of RWD were extracted from four main NICE documents (final scope, company submission, ERG report, FAD). Although most evidence used in appraisals should be available in these documents, other documents, such as technical engagement documents or patient group, professional group and NHS organisation submissions were excluded from data extraction. Although new evidence is less likely to be introduced in these documents, and additional data will be presented in the FAD if these data impact the decision-making, there may be a slim chance of missing information about the use of RWD to justify the assumption in economic models.

7.5 Policy implication

This section outlines the policy context of RWD at the time of writing (2022) and briefly comments on some potential implications of the research findings for the role of RWD in healthcare decision-making. Interest in RWD can be found in several health policy documents. Life science industry actors have claimed that RWD can provide opportunities to discover new drugs and bring innovation into the market early (220). An Association of British Pharmaceutical Industry (ABPI) guidance document indicated that RWD could demonstrate value of any medicine to a relevant UK population more comprehensively rather than simply relying on clinical expert opinions (38). Macmillan Cancer Support, a patient organisation, believes that RWD from health registries could help inform NHS planning and policy decisions (221). The European Medicines Agency (EMA) outlined its vision to enable use of RWD in the EMA network strategy to 2025 (222) and recently initiated a coordination centre for Data Analysis and Real-World Interrogation (DARWIN EU) to develop and manage a network of RWD sources across the EU (223). While these policy documents highlight the potential use of RWD and even some plans for data collection, how this use of RWD could answer current problems in the regulatory process or payer's decision-making has not been well-described.

NICE has also shown its ambition for RWD with more detailed guidance. In *The NICE strategy 2021 to 2026: Dynamic, Collaborative, Excellent* (109), one of its visions for the future was to be "a scientific

leader driving the research agenda and developing innovative and data-driven methods, using real-world data to resolve issues of uncertainty and improve access to innovations for patients." In January 2022, NICE updated their methods guide adopting new approaches to the evidence (167). Here, potential areas where RWD may be used were addressed, such as showing long-term effectiveness. More practically, NICE published a real-world evidence framework (32) that included where RWD can reduce uncertainties and improve guidance. The framework goes beyond describing the roles for RWD, to providing guidance on methods for assessing comparative treatment effects using RWD.

In many policy papers, active management of health data is expected to take the lead in health science and fast uptake of technology by providing helpful information. Increasing the development of RWD databases can, to some extent, assure the great opportunity for the use of RWD in the decision-making process, and more broadly in health science research. However, RWD are not new data. Data have been routinely collected from clinical practice for a long time, but they have not been used. For example, despite substantial interest, registry data for rare diseases have not been fully exploited. The discussion arising during past decision-making processes should be fully explored to expand the strategy for the use of RWD. This thesis clarifies what types of real-world information need to be collected to help payer's decision-making. The components of the economic model where RWD could be useful were identified in the protocol. The datasets such as hospital data, SACT or national audit data can be reviewed to determine to what extent these components can be obtained and the challenges for obtaining this information from these data. This can assist the inclusion of useful data variables in patient registries and improve management of RWD to further its use in future appraisals.

The NICE real-world evidence framework identified several areas where RWD can help to reduce uncertainty. However, this thesis found that the most frequently reported sources of uncertainty, such as the long-term effects and the absence of direct treatment comparison, were hardly

answered by RWD. This thesis found that in CDF review appraisals, the role of RWD was very limited, although SACT data were available. This result indicates the need for more detailed methods guides, and further discussion for future versions of the real-world evidence framework.

7.6 Areas for future research

There are several areas of future research that have emerged as priorities resulting from this thesis:

1) Explore the impacts of the NICE real-world evidence framework on the use of RWD
2) Compare use of RWD in the appraisal of non-oncology medicines with that of oncology medicines
3) Analyse different views on the use of RWD among key HTA stakeholders in NICE appraisals

A well-developed RWD collection tool and guidance on its use can maximise the use of existing RWD and collect good quality RWD for decision-making (224). The NICE real-world evidence framework was issued in June 2022. This guidance for using real-world studies in NICE guidance can contribute to increasing the understanding of RWD in HTA. It would be interesting to measure the impacts of this framework in terms of greater use of RWD or change of the patterns/intensity of use of RWD.

In this thesis, the cancer drug appraisals were reviewed to explore to what extent RWD were used in appraisals. This research question can be expanded to all appraisals. Future study can ask what the common sources of uncertainty are in appraisals of non-cancer drug. Also, an innovative medicines fund (IMF) has been recently introduced by the NHS. IMF is a managed access process for non-cancer drugs, which has a similar structure to that of the CDF. It would be interesting to explore different patterns of using RWD and examine to what extent RWD reduced the uncertainties identified in the appraisals of drugs entering the IMF, although it will be a number of years before we have enough appraisals for such an analysis.

Another area for future research is to investigate the different views on RWD. Appraisal committees can have different judgments regarding the quality or suitability of RWD for answering questions in

appraisals. By systematically comparing willingness to use RWD and identifying if for some uses, appraisal committees and manufacturers more likely to agree or less likely to agree, the different views on RWD between key players in NICE appraisals can be identified. This could contribute a fuller, more balanced understanding of use of RWD in NICE appraisals.

References

1. Kang J. Real-World Data in Health Technology Assessment: Do We Know It Well Enough? Hum Perspect Heal Sci Technol [Internet]. 2022;5:187–203. Available from: https://link.springer.com/chapter/10.1007/978-3-030-92612-0_12
2. De Brún C. The Information Standard Guide Finding the Evidence-7th. 2013.
3. WHO. WHO | HTA Definitions. WHO [Internet]. 2015; Available from: https://www.who.int/health-technology-assessment/about/Defining/en/
4. Potter BK, Avard D, Entwistle V, Kennedy C, Chakraborty P, McGuire M, et al. Ethical, Legal, and Social Issues in Health Technology Assessment for Prenatal/Preconceptional and Newborn Screening: A Workshop Report. Public Health Genomics [Internet]. 2008;12(1):4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2790790/
5. Henshall C, Oortwijn W, Stevens A, Granados A, Banta D. Priority setting for health technology assessment. Theoretical considerations and practical approaches. Priority setting Subgroup of the EUR-ASSESS Project. Int J Technol Assess Health Care [Internet]. 1997;13(2):144–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9194351
6. Velasco-Garrido M, World Health Organization., European Observatory on Health Systems and Policies. Health technology assessment and health policy-making in Europe : current status, challenges and potential. World Health Organization on behalf of the European Observatory on Health Systems and Policies; 2008. 181 p.
7. National Institute for Health and Care Excellence (NICE). About NICE [Internet]. Available from: https://www.nice.org.uk/about/what-we-do
8. Culyer AJ. NICE's use of cost effectiveness as an exemplar of a deliberative process. Heal Econ Policy Law [Internet]. 2006;1(3):299–318. Available from: https://www.cambridge.org/core/journals/health-economics-policy-and-law/article/nices-use-of-cost-effectiveness-as-an-exemplar-of-a-deliberative-process/3A0EF9C864D942039C80C8D48CF50824

9. National Institute for Health and Care Excellence (NICE). About | NICE [Internet]. NICE; Available from: https://www.nice.org.uk/about
10. National Insitute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual [Internet]. 2022. Available from: https://www.nice.org.uk/process/pmg36/
11. National Institute fore Health and Care Excellence (NICE). Cost comparison - Addendum to the Guide to the methods of technology appraisal [Internet]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/methods-guide-addendum-cost-comparison.pdf
12. Guide to the multiple technology appraisal process [Internet]. Available from: www.nice.org.uk
13. National Institute for Health and Care Excellence (NICE). Guide to the single technology appraisal process [Internet]. 2009. Available from: https://www.nice.org.uk/media/default/about/what-we-do/nice-guidance/nice-technology-appraisals/guide-to-the-single-technology-appraisal-process.pdf
14. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011 Jul;128(1):305–10.
15. Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: Challenges and opportunities. J Child Neurol [Internet]. 2013 Sep;28(9):1142–50. Available from: /pmc/articles/PMC3964003/?report=abstract
16. OECD. Addressing Challenges in Access to Oncology Medicines. 2020.
17. Walton MJ, O'Connor J, Carroll C, Claxton L, Hodgson R. A Review of Issues Affecting the Efficiency of Decision Making in the NICE Single Technology Appraisal Process. PharmacoEconomics - Open [Internet]. 2019 Sep 1;3(3):403–10. Available from: https://link.springer.com/article/10.1007/s41669-018-0113-0
18. Dickson R, Boland A, Duarte R, Kotas E, Woolacott N, Hodgson R, et al. EMA and NICE Appraisal Processes for Cancer Drugs: Current Status and Uncertainties. Appl Health Econ

Health Policy [Internet]. 2018 Aug 1;16(4):429. Available from: /pmc/articles/PMC6028888/

19. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, et al. Generalisability in economic evaluation studies in healthcare: A review and case studies. Vol. 8, Health Technology Assessment. NIHR Journals Library; 2004.

20. Latimer N. Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation with Patient-Level Data - PubMed - NCBI. Med Decis Mak [Internet]. 2013;33(6):743–54. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27905716

21. Tai TA, Latimer NR, Benedict A, Kiss Z, Nikolaou A. Prevalence of Immature Survival Data for Anti-Cancer Drugs Presented to the National Institute for Health and Care Excellence and Impact on Decision Making. Value Heal. 2020 Dec 8;

22. Vokinger KN, Kesselheim AS. Characteristics of trials and regulatory pathways leading to US approval of innovative vs. non-innovative oncology drugs. Health Policy (New York). 2019 Aug 1;123(8):721–7.

23. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. Med Decis Mak [Internet]. 2018 Feb 1;38(2):200. Available from: /pmc/articles/PMC5774635/

24. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research. Value Heal. 2012 Sep 1;15(6):940–7.

25. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ [Internet]. 2009 Apr 3;338(apr03 1):b1147–b1147. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19346285

26. Claxton K. Pharmaceutical Pricing: Early Access, The Cancer Drugs Fund and the Role of NICE. Cent Heal Econ Univ York [Internet]. 2016; Available from: http://www.york.ac.uk/media/che/documents/policybriefing/Drug_prices.pdf

27. Doolub N, Gupta S, Lohmann S, Khogeer B, Grosvenor A. PCN296 OUT WITH THE OLD AND IN WITH THE NEW: THE EVOLUTION OF THE CANCER DRUGS FUND. Value Heal. 2019 Nov;22:S493.

28. National Institute for Health and Care Excellence (NICE). Cancer Drugs Fund [Internet]. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/cancer-drugs-fund

29. National Institute for Health and Care Excellence (NICE). Specification for CDF data collection arrangements [Internet]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/cancer-drugs-fund/data-collection-specification.pdf

30. Public Health England. The National Cancer Registration and Analysis Service A guide to cancer data and working with us. 2020; Available from: www.facebook.com/PublicHealthEngland

31. National Insitute for Health and Care Excellence (NICE). The NICE methods of health technology evaluation: the case for change [Internet]. 2020. Available from: https://www.nice.org.uk/news/article/nice-s-methods-of-technology-evaluation-presenting-a-case-for-change

32. National Institute for Health and Care Excellence (NICE). NICE real-world evidence framework [Internet]. 2022. Available from: www.nice.org.uk/corporate/ecd9

33. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value Heal [Internet]. 2017;20(7):858–65. Available from: http://dx.doi.org/10.1016/j.jval.2017.03.008

34. Murphy LA, Mestre-Ferrandiz J, Solà-Morales O, de Pouvourville G, Cunningham D, Corry S, et al. PMU112 Context Matters — a Taxonomy of Real-World DATA and Evidence to Support HTA Decision-Making. Value Heal. 2020 Dec;23:S622.

35. National Institute for Health and Care Excellence (NICE). Evidence Standards Framework for

Digital Health Technologies Contents. 2019.

36. FDA. Framework for FDA's Real-World Evidence Program [Internet]. 2018. Available from: www.fda.gov

37. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. Value Heal [Internet]. 2007 Sep 1;10(5):326–35. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1524-4733.2007.00186.x

38. Association of the British Pharmaceutical Industry. Demonstrating value with Real World data | ABPI [Internet]. 2011. Available from: https://www.abpi.org.uk/publications/real-world-data/

39. Flynn R, Plueschke K, Quinten C, Strassmann V, Duijnhoven RG, Gordillo-Marañon M, et al. Marketing Authorization Applications Made to the European Medicines Agency in 2018–2019: What was the Contribution of Real-World Evidence? Clin Pharmacol Ther. 2022 Jan 1;111(1):90–7.

40. Berger M, Daniel G, Frank K, Hernandez A, McClellan M, Okun S, et al. A framework for regulatory use of real-world evidence [Internet]. 2017. Available from: https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe_white_paper_2017.09.06.pdf

41. Grimberg F, Asprion PM, Schneider B, Miho E, Babrak L, Habbabeh A. The Real-World Data Challenges Radar: A Review on the Challenges and Risks regarding the Use of Real-World Data. Digit Biomarkers [Internet]. 2021 Jun 24;5(2):148–57. Available from: https://www.karger.com/Article/FullText/516178

42. EMA. Compassionate use [Internet]. Available from: https://www.ema.europa.eu/en/human-regulatory/research-development/compassionate-use

43. National Institute for Health and Care Excellence (NICE). TA487 Venetoclax for treating chronic lymphocytic leukaemia . NICE; 2017.

44. EMA. Compassionate use [Internet]. Available from: https://www.ema.europa.eu/en/human-

regulatory/research-development/compassionate-use

45. National Institute for Health and Care Excellence (NICE). TA802 Cemiplimab for treating advanced cutaneous squamous cell carcinoma | Guidance | NICE. NICE; 2022.

46. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. Br J Cancer. 2006 Sep 18;95(6):683–90.

47. NHS Digital. Health Survey for England [Internet]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england

48. Makady A, Ham R ten, de Boer A, Hillege H, Klungel O, Goettsch W. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. Value Heal [Internet]. 2017 Apr 1;20(4):520–32. Available from: http://dx.doi.org/10.1016/j.jval.2016.12.003

49. Allen G, Summers DM, Evans J. PBI43 REAL-WORLD EVIDENCE RECOMMENDATIONS FOR GENE THERAPIES: REGULATORY VERSUS HEALTH TECHNOLOGY ASSESSMENT. Value Heal. 2020 May;23:S21.

50. Anderson M, Naci H, Morrison D, Osipenko L, Mossialos E. A review of NICE appraisals of pharmaceuticals 2000–2016 found variation in establishing comparative clinical effectiveness. J Clin Epidemiol [Internet]. 2019 Jan 1;105:50–9. Available from: http://www.jclinepi.com/article/S0895435618302786/fulltext

51. Rizzo M, Garcia A, Appiah K, Hernandez L, Sarri G. HTA51 Unravelling the Rationale for the Source of External Control Arm in Single-Arm Oncology Submissions: What Is the Justification behind the Maze of Options? Value Heal. 2022 Jul;25(7):S513.

52. Jaksa A, Louder A, Maksymiuk C, Vondeling GT, Martin L, Gatto N, et al. A Comparison of Four Oncology External Control Arm Case Studies: Critiques From Regulatory and Health Technology Assessment Agencies. Value Health [Internet]. 2022 Jun 24; Available from: http://www.ncbi.nlm.nih.gov/pubmed/35760714

53. National Institute for Health and Care Excellence (NICE). TA559 Axicabtagene ciloleucel for

treating diffuse large B-cell lymphoma and primary mediastinal large B-cell lymphoma after 2 or more systemic therapies. NICE; 2019.

54. EMA. Orphan medicine [Internet]. Available from: https://www.ema.europa.eu/en/glossary/orphan-medicine

55. Stavropoulou A, Teale C, Paydas L, Harrow G, Roldan-Gomendio A. POSA258 Leveraging RWE and Integrating Multi-Source Data to Build the Case for Funding and Access for Rare Disease Drugs. Value Heal. 2022 Jan;25(1):S170.

56. Mickle A, Harwood M, Deighton A, Qian C, Szabo S. RW2 THE USE OF REAL-WORLD DATA IN HEALTH TECHNOLOGY ASSESSMENT OF MEDICATIONS FOR RARE DISEASES IN CANADA AND THE UK. Value Heal. 2020 May;23:S384.

57. Harricharan S, Gurjar K, Nguyen K, Forsythe A. PCN172 Evaluation of Real-World Evidence (RWE) Used within Health Technology Assessments (HTA) in Oncology with Pivotal Single-Arm Trials: A Comparative Study of Six HTA Agencies. Value Heal. 2021 Jun;24:S51.

58. Falla E, Gaultney J. POSB267 Influence of Real-World Evidence to Support Single Arm Trials Submitted to NICE for Orphan Oncology Products. Value Heal. 2022 Jan;25(1):S179.

59. National Institute for Health and Care Excellence (NICE). TA604 Idelalisib for treating refractory follicular lymphoma. NICE; 2019.

60. NICE. 7 Assessing cost effectiveness.

61. Pittell H, Kent S, Groves B, Mpofu P, Baxi S, Copeland A, et al. P14 Exploring the Potential for EHR-Derived Real-World Data to Reduce Uncertainty in HTA Decision-Making: A Case Study of Long-Term Survival Outcomes. Value Heal [Internet]. 2022 Jan 1;25(1):S3–4. Available from: http://www.valueinhealthjournal.com/article/S1098301521018088/fulltext

62. National Institute for Health and Care Excellence (NICE). TA600 Pembrolizumab with carboplatin and paclitaxel for untreated metastatic squamous non-small-cell lung cancer. NICE; 2019.

63. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, et al. Bridging the

efficacyg-effectiveness gap: A regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov. 2011;10(7):495–506.

64. NICE. TA310 Afatinib for treating epidermal growth factor receptor mutation-positive locally advanced or metastatic non-small-cell lung cancer. NICE; 2014.

65. Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. Int J Med Inform [Internet]. 2017 Jul 1;103:32–41. Available from: https://pubmed.ncbi.nlm.nih.gov/28550999/

66. National Institute for Health and Care Excellence (NICE). TA491 Ibrutinib for treating Waldenstrom's macroglobulinaemia. NICE; 2017.

67. Sievers H, Joos A, Hiligsmann M. Real-world evidence: Perspectives on challenges, value, and alignment of regulatory and national health technology assessment data collection requirements. Int J Technol Assess Health Care. 2021;

68. McDonald JH. Handbook of Biological Statistics [Internet]. Sparky House Publishing. 2009. 313 p. Available from: https://books.google.co.uk/books/about/Handbook_of_Biological_Statistics.html?id=AsRTywAACAAJ&redir_esc=y

69. Grieve R, Abrams K, Claxton K, Goldacre B, James N, Nicholl J, et al. Cancer Drugs Fund requires further reform. Vol. 354, BMJ (Online). BMJ Publishing Group; 2016.

70. Evans SR. Clinical trial structures. Vol. 3, Journal of Experimental Stroke and Translational Medicine. Society for Experimental Stroke; 2010. p. 8–18.

71. NICE. TA502 Ibrutinib for treating relapsed or refractory mantle cell lymphoma. NICE; 2018.

72. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ, et al. NICE DSU TECHNICAL SUPPORT DOCUMENT 18: METHODS FOR POPULATION-ADJUSTED INDIRECT COMPARISONS IN SUBMISSIONS TO NICE REPORT BY THE DECISION SUPPORT UNIT [Internet]. 2016. Available from: www.nicedsu.org.uk

73. NICE. TA592 Cemiplimab for treating metastatic or locally advanced cutaneous squamous cell carcinoma. 2019;

74. Jarkowski A, Hare R, Loud P, Skitzki JJ, Kane JM, May KS, et al. Systemic Therapy in Advanced Cutaneous Squamous Cell Carcinoma (CSCC): The Roswell Park Experience and a Review of the Literature. Am J Clin Oncol Cancer Clin Trials [Internet]. 2016;39(6):545–8. Available from: https://pubmed.ncbi.nlm.nih.gov/24879468/

75. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value Heal. 2017;20(7):858–65.

76. National Institute for Health and Care Excellence (NICE). TA474 Sorafenib for treating advanced hepatocellular carcinoma. NICE; 2017.

77. NICE. TA395 Ceritinib for previously treated anaplastic lymphoma kinase positive non-small-cell lung cancer. NICE; 2016.

78. National Institute for Health and Care Excellence (NICE). TA462 Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma. NICE; 2017.

79. Cheah CY, Chihara D, Horowitz S, Sevin A, Oki Y, Zhou S, et al. Patients with classical Hodgkin lymphoma experiencing disease progression after treatment with brentuximab vedotin have poor outcomes. Ann Oncol Off J Eur Soc Med Oncol. 2016 Jul;27(7):1317–23.

80. Bullement A, Podkonjak T, Robinson MJ, Benson E, Selby R, Hatswell AJ, et al. Real-world evidence use in assessments of cancer drugs by NICE. Int J Technol Assess Health Care [Internet]. 2020 Aug 1;36(4):388–94. Available from: https://doi.org/10.1017/S0266462320000434

81. York Health Economics Consortium. Health Technology Assessment (online) [Internet]. 2016. Available from: https://yhec.co.uk/glossary/health-technology-assessment/

82. Health technologies. 2021 Dec 10; Available from: https://www.euro.who.int/en/health-topics/Health-systems/health-technologies-and-medicines/health-technologies

83. Chan K, Nam S, Evans B, Deoliveira C, Chambers A, Gavura S, et al. Developing a framework to

incorporate real-world evidence in cancer drug funding decisions: The Canadian Real-world Evidence for Value of Cancer Drugs (CanREValue) collaboration. BMJ Open. 2020 Jan 7;10(1).

84. Hilgers R. Design and analysis of clinical trials for small rare disease populations. J Rare Dis Res Treat. 2016 Nov 1;1(3):53–60.

85. Averitt AJ, Weng C, Ryan P, Perotte A. Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations. npj Digit Med 2020 31 [Internet]. 2020 May 11;3(1):1–10. Available from: https://www.nature.com/articles/s41746-020-0277-8

86. Davies C, Briggs A, Lorgelly P, Garellick G, Malchau H. The “hazards” of extrapolating survival curves. Med Decis Mak [Internet]. 2013 Apr;33(3):369–80. Available from: https://pubmed.ncbi.nlm.nih.gov/23457025/

87. Katkade VB, Sanders KN, Zou KH. Real world data: An opportunity to supplement existing evidence for the use of long-established medicines in health care decision making. J Multidiscip Healthc. 2018;11:295–304.

88. Littlejohns P, Chalkidou K, Culyer AJ, Weale A, Rid A, Kieslich K, et al. National Institute for Health and Care Excellence, social values and healthcare priority setting [Internet]. Vol. 112, Journal of the Royal Society of Medicine. SAGE Publications Ltd; 2019. p. 173–9. Available from: https://journals.sagepub.com/doi/full/10.1177/0141076819842846

89. Schlander M. Health Technology Assessments by the National Institute for Health and Clinical Excellence. Health Technology Assessments by the National Institute for Health and Clinical Excellence. Springer New York; 2007.

90. Broader types of data to be used in development of NICE guidance | News and features | News | NICE [Internet]. Available from: https://www.nice.org.uk/news/article/broader-types-of-data-to-be-used-in-development-of-nice-guidance

91. Wang S, Goring SM, Lozano-Ortega G. Inclusion of real-world evidence in submission packages to health technology assessment bodies: What do current guidelines indicate? Value Heal

[Internet]. 2016 May 1;19(3):A287. Available from: http://www.valueinhealthjournal.com/article/S1098301516008317/fulltext

92. Pietri G, Masoura P. Market Access and Reimbursement: The Increasing Role of Real-World Evidence. Value Heal [Internet]. 2014 Nov 1;17(7):A450–1. Available from: http://www.valueinhealthjournal.com/article/S1098301514031465/fulltext

93. Harwood M, Deighton A, Mickle A, Qian C, Szabo S. The use of real-world data in health technology assessment of medications for rare diseases. Value Heal [Internet]. 2019 Nov 1;22:S863. Available from: http://www.valueinhealthjournal.com/article/S1098301519348296/fulltext

94. Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. PharmacoEconomics - Open [Internet]. 2021 Mar 18;5(1):3–11. Available from: https://doi.org/10.1007/s41669-020-00221-y

95. FDA. Real-World Evidence [Internet]. 2020. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

96. National Institute for Health and Care Excllence. TA620 Olaparib for maintenance treatment of relapsed platinum-sensitive ovarian, fallopian tube or peritoneal cancer. NICE; 2020.

97. National Institute for Health and Care Excellence (NICE). TA495 Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. NICE;

98. Singh G, Schulthess D, Hughes N, Vannieuwenhuyse B, Kalra D. Real world big data for clinical research and drug development. Vol. 23, Drug Discovery Today. Elsevier Ltd; 2018. p. 652–60.

99. NICE. NICE signs up to join the GetReal Institute | News and features [Internet]. NICE; Available from: https://www.nice.org.uk/news/article/nice-signs-up-to-join-the-getreal-institute

100. Batra A, Cheung WY. Role of real-world evidence in informing cancer care: lessons from colorectal cancer. Curr Oncol [Internet]. 2019;26(Suppl 1):S53. Available from:

/pmc/articles/PMC6878934/

101. Mahendraratnam N, Mercon K, Gill M, Benzing L, McClellan MB. Understanding Use of Real-World Data and Real-World Evidence to Support Regulatory Decisions on Medical Product Effectiveness. Clin Pharmacol Ther [Internet]. 2022 Jan 1;111(1):150–4. Available from: https://pubmed.ncbi.nlm.nih.gov/33891318/

102. Hogervorst MA, Pontén J, Vreman RA, Mantel-Teeuwisse AK, Goettsch WG. Real World Data in Health Technology Assessment of Complex Health Technologies. Front Pharmacol [Internet]. 2022 Feb 10;13. Available from: /pmc/articles/PMC8866967/

103. Griffiths EA, Macaulay R, Vadlamudi NK, Uddin J, Samuels ER. The Role of Noncomparative Evidence in Health Technology Assessment Decisions. Value Health [Internet]. 2017 Dec 1;20(10):1245–51. Available from: https://pubmed.ncbi.nlm.nih.gov/29241883/

104. Jao R, Jaksa A, Pontynen A, Wang X. Health Techonolgy Assessment (HTA) Agencies Consideration of Real World Evidence (RWE). Value Heal [Internet]. 2018 May 1;21:S7. Available from: http://www.valueinhealthjournal.com/article/S1098301518303322/fulltext

105. Orsini LS, Berger M, Crown W, Daniel G, Eichler HG, Goettsch W, et al. Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—Why, What, and How: Recommendations and a Road Map from the Real-World Evidence Transparency Initiative. Value Heal [Internet]. 2020 Sep 1;23(9):1128–36. Available from: https://doi.org/10.1016/j.jval.2020.04.002

106. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices - Overview: A report of the ISPOR-SMDM modeling good research practices task force-1. Value Heal. 2012;15(6):796–803.

107. EUnetHTA. Publication of a Ja2 methodological guideline (2015). Internal Validity of Non-randomised Studies (NRS) on Interventions [Internet]. Available from: https://www.eunethta.eu/publication-of-a-ja2-methodological-guideline-internal-validity-of-non-randomised-studies-nrs-on-interventions/

108. NICE. Increasing use of health and social care data in guidance development [Internet]. NICE; Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/how-we-develop-nice-guidelines/data-and-analytics-statement-of-intent

109. National Institute for Health and Care Excellence (NICE). NICE strategy 2021 to 2026: Dynamic, Collaborative, Excellent. 2021;

110. Bell H, Wailoo AJ, 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 [Internet]. 2016. Available from: www.nicedsu.org.uk

111. National Institute for Health and Care Excellence (NICE). TA627 Lenalidomide with rituximab for previously treated follicular lymphoma. NICE; 2020.

112. National Institute for Health and Care Excellence (NICE). TA795 Ibrutinib for treating Waldenstrom's macroglobulinaemia. NICE; 2022.

113. Glen F, Pulfer A. The Use of Real World Data for Nice Decision-Making: A Review of Submissions To Three Treatment Effectiveness Evaluation Programmes In 2015 And 2016. Value Heal. 2017 Oct 1;20(9):A693.

114. Makady A, van Veelen A, Jonsson P, Moseley O, D'Andon A, de Boer A, et al. Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies. Pharmacoeconomics [Internet]. 2018 Mar 1;36(3):359–68. Available from: https://pubmed.ncbi.nlm.nih.gov/29214389/

115. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: Opportunities and challenges. J Comp Eff Res. 2018 Dec 1;7(12):1133–43.

116. Boyle JM, Hegarty G, Frampton C, Harvey-Jones E, Dodkins J, Beyer K, et al. Real-world outcomes associated with new cancer medicines approved by the Food and Drug Administration and European Medicines Agency: A retrospective cohort study. Eur J Cancer [Internet]. 2021 Sep 1;155:136–44. Available from:

https://linkinghub.elsevier.com/retrieve/pii/S0959804921004226

117. Kang J, Cairns J. Protocol for data extraction: how real-world data have been used in the National Institute for Health and Care Excellence appraisals of cancer therapy. BMJ Open [Internet]. 2022 Jan 1;12(1):e055985. Available from: https://bmjopen.bmj.com/content/12/1/e055985

118. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good Practices for Real-World Data Studies of Treatment and/or Comparative Effectiveness: Recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. Value Heal. 2017 Sep 1;20(8):1003–8.

119. Peterson B, Harrell FE. Partial Proportional Odds Models for Ordinal Response Variables. Appl Stat. 1990;39(2):205.

120. Liu X. Applied ordinal logistic regression using Stata : from single-level to multilevel modeling. SAGE Publications; 2015. 552 p.

121. National Institute for Health and Care Excellence (NICE). TA707 Nivolumab for previously treated unresectable advanced or recurrent oesophageal cancer . NICE; 2021.

122. George E. How real-world data compensate for scarce evidence in HTA. Z Evid Fortbild Qual Gesundhwes [Internet]. 2016;112:S23–6. Available from: https://pubmed.ncbi.nlm.nih.gov/27320024/

123. Faria R, Alava MH, Manca A, Wailoo AJ, Uk WWNO. NICE DSU TECHNICAL SUPPORT DOCUMENT 17: THE USE OF OBSERVATIONAL DATA TO INFORM ESTIMATES OF TREATMENT EFFECTIVENESS IN TECHNOLOGY APPRAISAL: METHODS FOR COMPARATIVE INDIVIDUAL PATIENT DATA REPORT BY THE DECISION SUPPORT UNIT [Internet]. 2015. Available from: www.nicedsu.org.uk

124. Latimer NR. Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data. Med Decis Mak [Internet]. 2013 Aug 22;33(6):743–54. Available from: http://journals.sagepub.com/doi/10.1177/0272989X12472398

125. Hunger T, Schnell-Inderst P, Sahakyan N, Siebert U. USING EXPERT OPINION IN HEALTH TECHNOLOGY ASSESSMENT: A GUIDELINE REVIEW. Int J Technol Assess Health Care [Internet]. 2016;32(3):131–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27502308/

126. Grigore B, Peters J, Hyde C, Stein K. Methods to Elicit Probability Distributions from Experts: A Systematic Review of Reported Practice in Health Technology Assessment. PharmacoEconomics 2013 3111 [Internet]. 2013 Oct 9;31(11):991–1003. Available from: https://link.springer.com/article/10.1007/s40273-013-0092-z

127. National Institute for Health and Care Excellence (NICE). Developing NICE guidelines: the manual - appendix H: Appraisal checklists, evidence tables, GRADE and economic profiles [Internet]. 2022. Available from: http://www.bristol.ac.uk/media-library/sites/social-community-medicine/robis/robisguidancedocument.pdfAmstarwww.amstar.caDSUNMAmethodologychecklisthttp://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf

128. The problem of multicollinearity. Underst Regres Anal [Internet]. 1997 Nov 22;176–80. Available from: https://link.springer.com/chapter/10.1007/978-0-585-25657-3_37

129. Heemstra L, Sweeney N, Engen A Van. Do Nice Evidence Review Groups (ERG) Focus on Different Aspects of Manufacturer Submissions in Oncology? Value Heal [Internet]. 2013 Nov 1;16(7):A429. Available from: http://www.valueinhealthjournal.com/article/S1098301513025126/fulltext

130. Panageas KS. Clinical trial design for rare cancers - why a less conventional route may be required. Expert Rev Clin Pharmacol [Internet]. 2015 Nov 2;8(6):661. Available from: /pmc/articles/PMC4724195/

131. Agrawal S, Arora S, Vallejo JJ, Gwise T, Chuk MK, Amiri-Kordestani L, et al. Use of single-arm trials to support malignant hematology and oncology drug and biologic approvals: A 20-year FDA experience. https://doi.org/101200/JCO20213915_suppl.e13572. 2021 May

28;39(15_suppl):e13572–e13572.

132. National Institute for Health and Care Excellence (NICE). TA660 Darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer. NICE; 2020.

133. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford: Oxford University Press; 2015.

134. National Institute for Health and Care Excellence. The guidelines manual. NICE; 2012.

135. Latimer NR, Adler AI. Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness. BMJ Med [Internet]. 2022 Mar 1;1(1):e000094. Available from: https://bmjmedicine.bmj.com/content/1/1/e000094

136. Roze S, Bertrand N, Eberst L, Borget I. Projecting overall survival data for health-economic models in oncology: Do maturity levels impact uncertainty? https://doi.org/101200/JCO20193715_suppl.e18350. 2019 May 26;37(15_suppl):e18350–e18350.

137. Hlatky MA, Boothroyd DB, Johnstone IM. Economic evaluation in long-term clinical trials. Stat Med [Internet]. 2002 Oct 15;21(19):2879–88. Available from: https://pubmed.ncbi.nlm.nih.gov/12325104/

138. Pruefert A, Skaltsa K, Maervoet J, Van EA. Immature Survival Data From Early Trial Termination – Theory and Hta Practice. Value Heal. 2014 Nov 1;17(7):A581–2.

139. Facey KM, Rannanheimo P, Batchelor L, Borchardt M, De Cock J. Real-world evidence to support Payer/HTA decisions about highly innovative technologies in the EU - Actions for stakeholders. Int J Technol Assess Health Care [Internet]. 2020 Aug 1;36(4):459–68. Available from: https://doi.org/10.1017/S026646232000063X

140. Arlett P, Kjær J, Broich K, Cooke E. Real-World Evidence in EU Medicines Regulation: Enabling Use and Establishing Value. Clin Pharmacol Ther [Internet]. 2022 Jan 1;111(1):21–3. Available

from: https://onlinelibrary.wiley.com/doi/full/10.1002/cpt.2479

141. Wedam S, Fashoyin-Aje L, Bloomquist E, Tang S, Sridhara R, Goldberg KB, et al. FDA approval summary: Palbociclib for Male patients with metastatic breast cancer. Clin Cancer Res [Internet]. 2020 Mar 15;26(6):1208–12. Available from: https://pubmed.ncbi.nlm.nih.gov/31649043/

142. Wang S V., Schneeweiss S, Gagne JJ, Evers T, Gerlinger C, Desai R, et al. Using Real-World Data to Extrapolate Evidence From Randomized Controlled Trials. Clin Pharmacol Ther [Internet]. 2019 May 1;105(5):1156–63. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cpt.1210

143. Alili H El, Gezin A, Postma M, Ouwens DM, Heeg B. PCN430 USING BAYESIAN METHODOLOGIES TO INFORM IMMATURE SURVIVAL EXTRAPOLATIONS BASED ON RWD AND EXPERT ELICITATION IN NON-SMALL CELL LUNG CANCER (NSCLC). Value Heal [Internet]. 2019 Nov 1;22:S519. Available from: http://www.valueinhealthjournal.com/article/S1098301519330013/fulltext

144. Predicting long-term effectiveness with a model-averaging approach – RWE Navigator [Internet]. Available from: https://rwe-navigator.eu/use-real-world-evidence/model-effectiveness-in-the-real-world/overview-of-methods-for-predicting-outcomes/model-averaging-approach-to-predicting-long-term-effectiveness/

145. Gill JL, Jahnz-Rozyk, Schramm K, Spandonaro W, Thomas F, Kanavos M. The use of Real World Evidence in the European context An analysis of key expert opinion.

146. Kang J, Cairns J. “Don’t Think Twice, It’s All Right”: Using Additional Data to Reduce Uncertainty Regarding Oncologic Drugs Provided Through Managed Access Agreements in England. PharmacoEconomics - Open 2022 [Internet]. 2022 Sep 20;1–15. Available from: https://link.springer.com/article/10.1007/s41669-022-00369-9

147. U.S. Food & Drug Administration (FDA). Accelerated Approval [Internet]. Available from: https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-

priority-review/accelerated-approval

148. European Medicines Agency (EMA). Support for early access [Internet]. Available from: https://www.ema.europa.eu/en/human-regulatory/overview/support-early-access

149. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: Retrospective cohort study of drug approvals 2009-13. BMJ [Internet]. 2017 Oct 4;359:4530. Available from: http://dx.doi.org/10.1136/bmj.j4530

150. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. Value Heal [Internet]. 2017 Mar 1;20(3):487–95. Available from: https://pubmed.ncbi.nlm.nih.gov/28292495/

151. Quinn C, Garrison LP, Pownell AK, Atkins MB, De Pouvourville G, Harrington K, et al. Current challenges for assessing the long-term clinical benefit of cancer immunotherapy: A multi-stakeholder perspective [Internet]. Vol. 8, Journal for ImmunoTherapy of Cancer. BMJ Publishing Group; 2020. p. 648. Available from: http://jitc.bmj.com/

152. Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. EClinicalMedicine [Internet]. 2020 Apr 1;21:100332. Available from: http://creativecommons.org/licenses/by-nc-nd/4.0/

153. Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. EClinicalMedicine [Internet]. 2020 Apr 1;21:100332. Available from: http://www.thelancet.com/article/S2589537020300766/fulltext

154. Cherla A, Naci H, Kesselheim AS, Gyawali B, Mossialos E. Assessment of Coverage in England of Cancer Drugs Qualifying for US Food and Drug Administration Accelerated Approval. JAMA Intern Med [Internet]. 2021 Apr 1;181(4):490–8. Available from:

https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2776284

155. Cheema PK, Burkes RL. Overall survival should be the primary endpoint in clinical trials for advanced non-small-cell lung cancer. Curr Oncol [Internet]. 2013;20(2):e150. Available from: /pmc/articles/PMC3615866/

156. NICE task and finish group. CHTE methods review: Sources and synthesis of evidence task and finish group report. 2020.

157. 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 Serv Res [Internet]. 2018 May 31;18(1):1–9. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-018-3162-2

158. Van Oostrum I, Ouwens M, Remiro-Azócar A, Baio G, Postma MJ, Buskens E, et al. Comparison of Parametric Survival Extrapolation Approaches Incorporating General Population Mortality for Adequate Health Technology Assessment of New Oncology Drugs. 2021; Available from: https://doi.org/10.1016/j.jval.2021.03.008

159. Rutherford MJ, Lambert PC, Sweeting MJ, Pennington B, Crowther MJ, Abrams KR, et al. NICE DSU TECHNICAL SUPPORT DOCUMENT 21: Flexible Methods for Survival Analysis REPORT BY THE DECISION SUPPORT UNIT. 2020; Available from: www.nicedsu.org.uk

160. Ramagopalan S V., Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? [Internet]. Vol. 18, BMC Medicine. BioMed Central; 2020. p. 13. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1481-8

161. Patel D, Grimson F, Mihaylova E, Wagner P, Warren J, van Engen A, et al. Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. Value Heal. 2021 Aug 1;24(8):1118–25.

162. Burcu M, Dreyer NA, Franklin JM, Blum MD, Critchlow CW, Perfetto EM, et al. Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. Pharmacoepidemiol Drug Saf. 2020 Oct 1;29(10):1228–35.

163. Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. Clin Pharmacol Ther [Internet]. 2019 Jul 1;106(1):36. Available from: /pmc/articles/PMC6617710/

164. Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. PharmacoEconomics - Open [Internet]. 2021 Mar 1;5(1):3–11. Available from: https://link.springer.com/article/10.1007/s41669-020-00221-y

165. SACT Homepage [Internet]. Available from: http://www.chemodataset.nhs.uk/home

166. Bowrin K, Briere JB, Levy P, Millier A, Clay E, Toumi M. Cost-effectiveness analyses using real-world data: an overview of the literature. Vol. 22, Journal of Medical Economics. Taylor and Francis Ltd; 2019. p. 545–53.

167. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual [Internet]. NICE; Available from: https://www.nice.org.uk/process/pmg36/chapter/developing-the-guidance

168. Bullement A, Taylor M, McMordie ST, Waters E, Hatswell AJ. NICE, in Confidence: An Assessment of Redaction to Obscure Confidential Information in Single Technology Appraisals by the National Institute for Health and Care Excellence. Pharmacoeconomics [Internet]. 2019 Nov 1;37(11):1383–90. Available from: https://pubmed.ncbi.nlm.nih.gov/31250397/

169. Charlton V. NICE and Fair? Health Technology Assessment Policy Under the UK's National Institute for Health and Care Excellence, 1999–2018. Heal Care Anal [Internet]. 2020 Sep 1;28(3):193–227. Available from: https://link.springer.com/article/10.1007/s10728-019-00381-x

170. NIH(National Cancer Institute). Targeted Cancer Therapies Fact Sheet - National Cancer Institute [Internet]. Available from: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet#q1

171. Riely GL, Kettering S. What, When, and How of Biomarker Testing in Non–Small Cell Lung Cancer. J Natl Compr Cancer Netw [Internet]. 2017 May 1;15(5S):686–8. Available from:

https://jnccn.org/view/journals/jnccn/15/5S/article-p686.xml

172. Majeed U, Manochakian R, Zhao Y, Lou Y. Targeted therapy in advanced non-small cell lung cancer: current advances and future trends. J Hematol Oncol 2021 141 [Internet]. 2021 Jul 8;14(1):1–20. Available from: https://jhoonline.biomedcentral.com/articles/10.1186/s13045-021-01121-2

173. König IR, Fuchs O, Hansen G, von Mutius E, Kopp M V. What is precision medicine? [Internet]. Vol. 50, The European respiratory journal. European Respiratory Society; 2017. p. 1700391. Available from: https://doi.org/10.1183/13993003.00391-2017].

174. Marshall DA, Grazziotin LR, Regier DA, Wordsworth S, Buchanan J, Phillips K, et al. Addressing Challenges of Economic Evaluation in Precision Medicine Using Dynamic Simulation Modeling. Value Health [Internet]. 2020 May 1;23(5):566–73. Available from: https://pubmed.ncbi.nlm.nih.gov/32389221/

175. Phillips KA, Deverka PA, Marshall DA, Wordsworth S, Regier DA, Christensen KD, et al. Methodological Issues in Assessing the Economic Value of Next-Generation Sequencing Tests: Many Challenges and Not Enough Solutions. Value Health [Internet]. 2018 Sep 1;21(9):1033–42. Available from: https://pubmed.ncbi.nlm.nih.gov/30224106/

176. Husereau D, Marshall DA, Levy AR, Peacock S, Hoch JS. Health technology assessment and personalized medicine: Are economic evaluation guidelines sufficient to support decision making? Int J Technol Assess Health Care [Internet]. 2014;30(2):179–87. Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/abs/health-technology-assessment-and-personalized-medicine-are-economic-evaluation-guidelines-sufficient-to-support-decision-making/FAF941E6848923E17271851354463E70

177. Cleary M. A Review of Precision Medicine, Companion Diagnostics, and the Challenges Surrounding Targeted Therapy. Value & Outcomes Spotlight. 2019;

178. Yap TA, Jacobs I, Baumfeld Andre E, Lee LJ, Beaupre D, Azoulay L. Application of Real-World

Data to External Control Groups in Oncology Clinical Trial Drug Development. Front Oncol. 2022 Jan 6;11:5569.

179. Faulkner E, Holtorf AP, Walton S, Liu CY, Lin H, Biltaj E, et al. Being Precise About Precision Medicine: What Should Value Frameworks Incorporate to Address Precision Medicine? A Report of the Personalized Precision Medicine Special Interest Group. Value Heal [Internet]. 2020 May 1;23(5):529–39. Available from: www.elsevier.com/locate/jval

180. Becla L, Lunshof JE, Gurwitz D, Schulte In Den Bäumen T, Westerhoff H V., Lange BMH, et al. Health technology assessment in the era of personalized health care. Int J Technol Assess Health Care [Internet]. 2011 Apr;27(2):118–26. Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/abs/health-technology-assessment-in-the-era-of-personalized-health-care/77EE92D3548D71887F321004D896F0D6

181. Cancer Drugs Fund | Technology appraisal guidance | NICE guidance | Our programmes | What we do | About | NICE. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/cancer-drugs-fund

182. Peters GJ. From 'Targeted Therapy' to Targeted Therapy. Anticancer Res [Internet]. 2019 Jul 1;39(7):3341–5. Available from: https://ar.iiarjournals.org/content/39/7/3341

183. Khorsan R, Crawford C. How to Assess the External Validity and Model Validity of Therapeutic Trials: A Conceptual Approach to Systematic Review Methodology. Evid Based Complement Alternat Med [Internet]. 2014;2014. Available from: /pmc/articles/PMC3963220/

184. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" Lancet (London, England) [Internet]. 2005 Jan 1;365(9453):82–93. Available from: https://pubmed.ncbi.nlm.nih.gov/15639683/

185. National Institute for Health and Care Excellence (NICE). TA530 Nivolumab for treating locally advanced unresectable or metastatic urothelial cancer after platinum-containing chemotherapy. NICE; 2018.

186. National Institute for Health and Care Excellence (NICE). TA578 Durvalumab for treating locally advanced unresectable non-small-cell lung cancer after platinum-based chemoradiation. NICE; 2019.

187. Cairns JA. Assessing the Cost-Effectiveness of Molecular Targeted Therapies and Immune Checkpoint Inhibitors. Hum Perspect Heal Sci Technol [Internet]. 2022;5:175–85. Available from: https://link.springer.com/chapter/10.1007/978-3-030-92612-0_11

188. Cancer incidence for common cancers | Cancer Research UK [Internet]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared

189. National Institute for Health and Care Excellence (NICE). TA630 Larotrectinib for treating NTRK fusion-positive solid tumours. NICE; 2020.

190. National Institute for Health and Care Excellence (NICE). TA644 Entrectinib for treating NTRK fusion-positive solid tumours. NICE; 2020.

191. Lengliné E, Peron J, Vanier A, Gueyffier F, Kouzan S, Dufour P, et al. Basket clinical trial design for targeted therapies for cancer: a French National Authority for Health statement for health technology assessment. Lancet Oncol [Internet]. 2021 Oct 1;22(10):e430–4. Available from: http://www.thelancet.com/article/S1470204521003375/fulltext

192. Salgado R, Solit DB, Rimm DL, Bogaerts J, Canetta R, Lively T, et al. Addressing the dichotomy between individual and societal approaches to personalised medicine in oncology. Eur J Cancer. 2019 Jun 1;114:128–36.

193. Beresford L, Murphy P, Dias S, Claxton L, Walton M, Metcalf R, et al. Appraising the Costs of Genomic Testing for Histology-Independent Technologies: An Illustrative Example for NTRK Fusions. Value Heal. 2022 Jul 1;25(7):1133–40.

194. Borcoman E, Kanjanapan Y, Champiat S, Kato S, Servois V, Kurzrock R, et al. Novel patterns of response under immunotherapy. Ann Oncol. 2019 Mar 1;30(3):385–96.

195. National Institute for Health and Care Excellence (NICE). TA692 Pembrolizumab for treating

locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy [Internet]. 2021. Available from: https://www.nice.org.uk/guidance/ta692

196. Love-Koh J, Peel A, Rejon-Parrilla JC, Ennis K, Lovett R, Manca A, et al. The Future of Precision Medicine: Potential Impacts for Health Technology Assessment. Pharmacoeconomics [Internet]. 2018 Dec 1;36(12):1439–51. Available from: https://doi.org/10.1007/s40273-018-0686-6

197. National Institute for Health and Care Excellence (NICE). TA704 Trastuzumab deruxtecan for treating HER2-positive unresectable or metastatic breast cancer after 2 or more anti-HER2 therapies. NICE; 2021.

198. Verweij J, Hendriks HR, Zwierzina H. Innovation in oncology clinical trial design. Cancer Treat Rev [Internet]. 2019 Mar 1;74:15–20. Available from: https://pubmed.ncbi.nlm.nih.gov/30665053/

199. Pruce D., Michel S., Theocharous P. Do payers find value in innovative trial designs? The Evidence Form [Internet]. 2017. Available from: https://www.evidera.com/wp-content/uploads/2017/05/How-Health-Technology-Assessment-Bodies-See-New-Trial-Design.pdf

200. Larson L., Walker A., Parow D. Innovative clinical trial designs Welcomed by regulators but what about the payers? ISPOR Glasgow. [Internet]. 2017. Available from: https://www.ispor.org/docs/default-source/presentations/1330.pdf?sfvrsn=56ed4398_1

201. Hogervorst MA, Vreman RA, Mantel-Teeuwisse AK, Goettsch WG. Reported Challenges in Health Technology Assessment of Complex Health Technologies. Value Heal [Internet]. 2021;0(0). Available from: http://www.valueinhealthjournal.com/article/S1098301521031879/fulltext

202. Dagenais S, Russo L, Madsen A, Webster J, Becnel L. Use of Real-World Evidence to Drive Drug Development Strategy and Inform Clinical Trial Design. Clin Pharmacol Ther [Internet]. 2022 Jan 1;111(1):77–89. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cpt.2480

203. National Institute for Health and Care Excellence (NICE). Appraising life-extending, end of life treatments. 2009; Available from: https://www.nice.org.uk/guidance/gid-tag387/documents/appraising-life-extending-end-of-life-treatments-paper2

204. House of Commons Committee of Public Accounts. Cancer Drugs Fund Twentieth Report of Session 2015-16 HC 583. Available from: https://publications.parliament.uk/pa/cm201516/cmselect/cmpubacc/583/58302.htm

205. Aggarwal A, Fojo T, Chamberlain C, Davis C, Sullivan R. Do patient access schemes for high-cost cancer drugs deliver value to society?-lessons from the NHS Cancer Drugs Fund. Ann Oncol Off J Eur Soc Med Oncol [Internet]. 2017 Aug 1;28(8):1738–50. Available from: https://pubmed.ncbi.nlm.nih.gov/28453615/

206. NHS England. Appraisal and Funding of Cancer Drugs from July 2016 [Internet]. Available from: https://www.england.nhs.uk/publication/cdf-sop-16/

207. NICE. PMG9 addendum – final amendments to the NICE technology appraisal methods guide to support the new Cancer Drugs Fund arrangements [Internet]. 2018. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/process-and-methods-guide-addendum.pdf

208. CONSULTATION ON PROPOSALS FOR A NEW CANCER DRUGS FUND (CDF) OPERATING MODEL FROM 1 ST APRIL 2016. Available from: https://www.engage.england.nhs.uk/consultation/cdf-consultation/supporting_documents/cdfconsultationdoc.pdf

209. National Institute for Health and Care Excellence (NICE). Guide to the processes of technology appraisal [Internet]. 2018. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf

210. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Stat Med

[Internet]. 2020; Available from: https://doi.org/10.1002/sim.8759

211. Jørgensen J, Kefalas P. Upgrading the SACT dataset and EBMT registry to enable outcomes-based reimbursement in oncology in England: a gap analysis and top-level cost estimate. J Mark Access Heal Policy [Internet]. 2019 Jan 1;7(1):1635842. Available from: /pmc/articles/PMC6609347/

212. Latimer NR. PCN317 - THE CANCER DRUGS FUND: KEY UNCERTAINTIES, DATA COLLECTION PLANS, ANALYTICAL METHODS AND USE OF THE SYSTEMATIC ANTI-CANCER THERAPY (SACT) REAL WORLD DATA SET. Value Heal [Internet]. 2018 Oct 1;21:S68. Available from: http://www.valueinhealthjournal.com/article/S1098301518336994/fulltext

213. Macdonald H, Goldacre B. Does The Reformed Cancer Drug Fund Generate Evidence On Effectiveness? A Cross-sectional Analysis On Publicly Accessible Documentation. medRxiv [Internet]. 2020 Mar 8;2020.03.06.19014944. Available from: https://www.medrxiv.org/content/10.1101/2020.03.06.19014944v1

214. The Innovative Medicines Fund: engagement on proposals. 2021; Available from: www.engage.england.nhs.uk,

215. Montilla S, Xoxi E, Russo P, Cicchetti A, Pani L. MONITORING REGISTRIES AT ITALIAN MEDICINES AGENCY: FOSTERING ACCESS, GUARANTEEING SUSTAINABILITY. Int J Technol Assess Health Care [Internet]. 2015 Dec 8;31(4):210–3. Available from: https://pubmed.ncbi.nlm.nih.gov/26646859/

216. Sultana J, Addis A, Braga M, Campomori A, Capuano A, Corrao G, et al. What can real world evidence contribute to regulatory science in pre and post marketing setting? Pharmadvances. 2020 Jul;02(02).

217. Makady A, van Acker S, Nijmeijer H, de Boer A, Hillege H, Klungel O, et al. Conditional Financing of Drugs in the Netherlands: Past, Present, and Future—Results From Stakeholder Interviews. Value Heal [Internet]. 2019 Apr 1;22(4):399–407. Available from: http://www.valueinhealthjournal.com/article/S1098301519300749/fulltext

218. Opel M, McDonell A, Woolmore A, Hiller J, Deitch S, Berger K. How to ensure access to innovative cancer treatment in times of health technology assessment and comparative effectiveness research? J Clin Oncol. 2015 May 20;33(15_suppl):e17501–e17501.

219. Hatogai K, Kato Y, Hirase C. Efficacy evaluation of anticancer agents in single-arm clinical trials: analysis of review reports from Pharmaceuticals and Medical Devices Agency. https://doi.org/101080/0284186X20211871946 [Internet]. 2021;60(2):143–8. Available from: https://www.tandfonline.com/doi/abs/10.1080/0284186X.2021.1871946

220. Real World Evidence Enabling the life sciences industry to transform patient care.

221. Cancer Research UK, Macmillan Cancer support. Improving awareness of the English cancer registry amongst patients, health professionals and the public Review of Informed Choice for Cancer Registration. 2016.

222. European Medicines Agency, Heads of Medicines Agencies. European medicines agencies network strategy to 2025 - Protecting public health at a time of rapid change. 2020;

223. Initiation of DARWIN EU® Coordination Centre advances integration of real-world evidence into assessment of medicines in the EU | European Medicines Agency [Internet]. Available from: https://www.ema.europa.eu/en/news/initiation-darwin-eur-coordination-centre-advances-integration-real-world-evidence-assessment

224. Allen A, Patrick H, Ruof J, Buchberger B, Varela-Lema L, Kirschner J, et al. Development and Pilot Test of the Registry Evaluation and Quality Standards Tool: An Information Technology-Based Tool to Support and Review Registries. Value Health [Internet]. 2022 Mar 8; Available from: http://www.ncbi.nlm.nih.gov/pubmed/35277336

225. NICE. TA638 Atezolizumab with carboplatin and etoposide for untreated extensive-stage small-cell lung cancer. NICE; 2020.

Appendices

Appendix 1. Glossary of variables used in extraction template

A.1.1 Variables in the extraction protocol

General information		
Variable	**Explanation**	**Coding**
Type of cancer	The NICE classification of the cancer (website: https://www.nice.org.uk/guidance/conditions-and-diseases/cancer)	Bladder cancer=1, Blood and bone marrow cancer =2, Breast cancer=3, Colorectal=4, Neuroblastoma=5, Head and neck=6, Liver=7, Lung=8, Oesophageal=9, Ovarian=10, Pancreatic=11, Prostate=12, Renal=13, Skin=14, Stomach=15, Sarcoma=16
Technology of interest	The name of drug in the current appraisal. If it is combination therapy, the key technology which manufacturer focuses on will be taken here.	Narrative description
Indication	Clinical indications which are addressed in Final Appraisal Determination (FAD) document	Narrative description
TA number	the reference number of the technology guidance	Narrative description
Replace	Whether TA guidance has replaced or not. Appraisals can be replaced after rapid reviews/reviews/updates of previous appraisals or CDF reviews. Regardless of reasons of replacement, TA reference number which is replaced by this appraisal of interest will be recorded.	None= 0 If current appraisal replaces previous appraisal, the replaced TA reference number is recorded here.
• Pre-2016 CDF reconsideration	Before April 2016, the drug which was not reviewed or not recommended for routine commissioning by NICE can be used using the previous model of CDF. When new CDF was introduced in April 2016, these drugs in the old CDF were appraised by NICE to transit the model of CDF. This variable describe whether the appraisal of interest is an appraisal of the CDF reconsideration for the drug used in the old model of CDF before 2016.	No, it is not pre-2016 CDF reconsideration =0 Yes, it is a appraisal of pre-2016 CDF reconsideration =1
• 2016 CDF review	In April 2016, a new model of CDF was introduced. In the new model, an additional recommendation, recommended for use within the CDF is available when NICE appraising cancer drugs. The drug available via the CDF has to collect the data for further review for the routine commissioning after a certain period. As this mandated data collection can impact on the use of RWD, this variable allows to distinguish the appraisals, which RWD is more likely to be used.	No, it is not 2016 CDF review =0 Yes, it is 2016 CF review=1

(Continued Table A.1.1)

Targeted cancer therapy	Treatment that uses drugs or other substances to identify and attack specific types of cancer cells	Non-targeted therapy = 0, targeted therapy = 1, not sure = Narrative description
Recommendation	the classification of recommendations made by the NICE committee in FAD document - Not recommended: 0 - Recommended (in line with marketing authorisation): 1 - Recommended (in line with marketing authorisation) in CDF:2 - Optimised: 3 - Optimised in CDF: 4 - Recommended in research: 5	Not recommended=0, recommended=1, recommended (cdf)=2, optimised=3, optimised (cdf)=4, recommended in research=5
number of comparators	Count the number of comparators in each manufacturer submission or FAD document. The information in manufacturer submission and FAD is recorded in the separated rows (manufacturer row/committee row).	Number in the manufacturer's submission
name of comparators	Record the name of comparators in manufacturer submission or FAD document	Narrative description
name of manufacturer	the name of manufacturer in manufacturer submission	Narrative description
name of the ERG	the name of the ERG (evidence review group)/AG (assessment group) in ERG critiques or AG reports	Narrative description
published date of final scope	the date of final scope as MM/YYYY	Date (MM/YYYY)
published date of manufacturer	the date of manufacturer submission as MM/YYYY.	Date (MM/YYYY)
published date of FAD guidance	the date of FAD document as MM/YYYY	Date (MM/YYYY)

(continued Table A.1.1)

Explanatory variables		
Variable	**Explanation**	**Coding**
Incidence (rate, year)	The rate would be recorded as it is in the appraisal. Incidence rate could be found in the final scope document or in manufacturer submission document. If the figures are not identical in each document, the latest rate is recorded. Most appraisals present the annual estimate of the number of patients who are eligible for the treatment in the "Budget Impact" section of company submission. This number is mainly used for the incidence. If this information is not available in the appraisal, the number in previous appraisal for similar indication is used instead.	Number
H2H	Whether the head-to-head clinical trial of a technology of interest exists or not, which compares with agreed comparators. The information is most likely to be found in the section: Identification and selection of relevant studies in clinical effectiveness part.	no=0, yes=1, yes but some comparators missing =2
• ITC	ITC (indirect treatment comparison). The information could be found in the section: Indirect and mixed treatment comparisons in clinical effectiveness part.	no=0, yes=1
• RCT (technology of interest)	Main RCT used in the appraisal: the name of the H2H RCT, if it exists. Unless there is an H2H, RCT refers to the clinical trial of technology of interest in the ITC.	no=0, yes=1
- Name of RCT	The name of the aforementioned RCT	Narrative description
- Intervention in RCT	The name of the intervention used in the aforementioned RCT. This variable helps to identify the main technology in RCT when technology is appraised as combination therapy.	Narrative description
- Comparators in RCT	The comparator of the aforementioned RCT	Narrative description
- Size of RCT	The number of participants in the aforementioned RCT	Number
- Median duration of follow-up	The median duration of follow-up in the aforementioned RCT. If it is not reported, record as NR (not reported).	Unit: month Not reported = ..
• Anchored/unanchored	"Anchored" means that RCT of technology of interest exists, and the RCT has been linked to any other studies which evaluate the drug's effectiveness. "Unanchored" means that the clinical outcome study doesn't have any comparators which connect to other studies. For example, comparing a single-arm study with a single-arm study is "unanchored". Also, RCTs compared without common comparators in ITC is "unanchored".	Not anchored=0, Anchored =1
• MAIC/STC	Matching adjusted indirect comparison (MAIC), Simulated Treatment Comparison (STC). A methodology of making adjustment to increase the comparability of two distinct populations mostly among unanchored studies. But it could be used in anchored studies in case where the two populations in ITC is starkly different from each other.	Naive=0, MAIC=1 STC=2 Other methods=3

(continued Table A.1.1)

Risk of bias (RoB) of RCT (direct quotation)	In order to evaluate the internal validity of RCTs, the risk of bias, which was reported in the ERG report, will be recorded here. Information is available at the quality assessment part of the ERG report. The ERG assesses the risk of bias of the included study using quality assessment tools. The ERG statement is directly quoted. The ERG often addresses the issue of quality of study narratively. Moreover, the ERG uses different terminology, whereas the domain of assessment is consistent. Therefore, the risk of bias would be narratively recorded. Prior to analysis, it will be scored by looking at the number of factors about which the ERG has expressed concern.	Direct quotation from ERG documents
• Risk of bias in RCT (grade)	In order to conduct statistical analysis, a set of codes will be used here. The direct quotation will be classified into four groups following the number of risk factors.	High/good quality without mentioned weakness= 0, risk factor 1 (low) =1, risk factor 2-3 (moderate)=2, risk factor 4 (high) =3
External validity of RCT	As narrative accounts, generalisability of RCT is reported in the ERG report whether the population of RCT properly represents the UK general population in terms of aging structure, health status and health care practice (practice-dose, subsequent treatment, etc.).	Direct quotation from ERG documents
• External validity in RCT (grade)	In order to conduct statistical analysis, a set of codes will be used here. The direct quotation will be classified into three groups following the severity of generalisability assess by ERG.	Representative without mentioned weakness= 0, Representative but minor concerns =1, Questionable generalisability =2
Previously recommended in other indication	Whether the technology has been recommended for other types of cancers besides the current indication of the technology.	No =0, Yes including all recommend, CDF, Optimised, Optimised (cdf) =1
• TA number & date of appraisal in other indication	If it was recommended for other indications, record the TA number and the date of the FAD documents (MM/YYYY).	Narrative description of date
Previous recommended treatment in the same cancer	Whether the technology has been recommended for other treatment lines in the same type of cancer.	No =0, Yes including all recommend, CDF, Optimised, Optimised (cdf) =1
• TA number & date of appraisal in the same cancer	If it was recommended for other treatment lines in the same cancer category, record the TA number and the date of the FAD documents (MM/YYYY).	Narrative description of date
Maturity of survival data in clinical trial	The data maturity is examined by looking at the number of events (deaths) of intervention arm in clinical trials. In published appraisal document, some of the information is redacted due to confidentiality. If the information is not available, the article of clinical trial published in journals is searched in order to check how many events are observed during the trial. Nonetheless, data are still not available in some cases. Since manufacturer is likely to redact the OS information when median OS was not reached. Hence, the survival data in this case are regarded as immature.	Direct quote from manufacturer submission
• Maturity (grade)	The direct quotation will be classified into three groups following the data cut point, 20% and 50 % of the number of events. This protocol adapts the criterion for measuring maturity of survival data in Tai et al. which investigates data maturity in STAs by looking at the proportion of death in pivotal trials. In the study, 20, 50 and 70 % of proportion of number of deaths are used to discuss the maturity of survival data (21). This protocol only uses 20% and 50% to assess the maturity without the category “unclear.”	Immature (number of events < 20%) =1, Relatively immature (20%≤number of events≤50%)=2 Mature (number of events < 50%) =3

(Continued Table A.1.1)

Outcome variables			
Variable	**Explanation**	**Coding**	**Example**
characteristic of population	Whether RWD are used to determine the characteristic of population, including the initiation age and health performance status (ECOG) or not. - Soft use: when RWD are supplementary evidence to decide the population characteristics - Hard use: when RWD determine the characteristics of population in economic evaluation	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Pomalidomide, in combination with low-dose dexamethasone, for treating multiple myeloma in adults at third or subsequent relapse (NICE TA427): baseline patient characteristics were obtained from RWD collected from a hospital population since the majority of the trial populations were previously untreated, which was different from target population.
treatment sequence	Whether RWD are used to determine the subsequent treatment option or not. After the disease progression onto the later stages of cancer treatments, patients are likely to receive idiosyncratic subsequent treatments. The pattern of subsequent treatment for cost-effectiveness analysis could be observed by RCT or RWD.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer (NICE TA495): a study of medical records was used to determine the treatment sequence.
choice of comparator	Whether RWD are used to choose the comparators in economic evaluation or not. Although comparators are chosen based on the current clinical guideline, drug utilisation data or clinical expert opinion are frequently referred to find the most relevant comparators in evaluation.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Ixazomib with lenalidomide and dexamethasone for treating relapsed or refractory multiple myeloma (NICE TA505): the manufacturer considered that lenalidomide was appropriate comparator based on IMS market research data (lenalidomide, 69% market share and panobinostat, 7%).
structure (health state)	Whether RWD are used to determine the health state such as stable, progression, and death in a given model. Information is available at health state in the model of cost-effectiveness analysis in manufacturer submission documents.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer (NICE TA495): the model health state of post-progression was specified based on a retrospective patient medical record review study.
structure (model cycle)	Whether RWD are used to determine model cycle or not. Model cycle, hereby, means that the duration between different health states, which can be influenced by the severity of conditions.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A **

(Continued Table A.1.1)

Variable	Explanation	Coding	Example
Structure (survival distribution of intervention)	Whether RWD are used to decide the survival distribution of intervention or not. Since survival rate observed in RCTs is immature, it is necessary to extrapolate the survival rate for analysis. In order to choose proper survival distribution, the goodness of fit is tested (AIC, BIC). Also, the clinical plausibility is asked to validate the distribution. In this case, the alternative data can be utilized. - If RWD is utilised for choosing distribution, mark as “hard use”. - If RWD is utilised as supplementary evidence for the chosen distribution, mark as “soft use”.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Larotrectinib for treating advanced solid tumours with NTRK fusions (NICE TA630): UK all-cause mortality data were used to assess the clinical acceptability of distributions whether patient overall survival exceeded current UK life expectancy
Structure (survival distribution of comparator)	Whether RWD are used to validate the feasibility of survival distribution of comparator or not. As survival distributions of intervention and comparators are separately determined, the extraction tool approach it independently. Apply the abovementioned description on survival distribution of intervention to comparator in this row.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	
Structure (Time to discontinuation of intervention)	Whether RWD are used to decide the time to discontinuation of intervention or not. The time to discontinuation is likely to be decided by 1) simply adopting discontinuation rule in trials, 2) formulating distribution of discontinuation, or 3) clinical experts’ opinion. - If RWD are used for designating the time to discontinuation, mark as “hard use” - If RWD are used as supplementary evidence for designating the time to discontinuation, mark as “soft use”. - If clinical experts’ opinions are used for designating the time to discontinuation, it is not regarded as RWD.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Lorlatinib for previously treated ALK-positive advanced non-small-cell lung cancer (NICE TA628): The plausibility of the extrapolation of time on treatment was validated by UK RWD, hospital network data.
Structure (time to discontinuation of comparator)	Whether RWD are used to decide the time to discontinuation of comparator or not. Apply the above-mentioned description on time to discontinuation of intervention to comparator in this row.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	

(Continued Table A.1.1)

Variable	Explanation	Coding	Example
Clinical outcome (OS) intervention	Whether RWD give the figure for overall survival (OS) of intervention or not. In order to measure the Quality Adjusted Life-Years (QALYs), it is necessary to extrapolate overall survival based on observed data on survival. The survival data could come from RCT or RWD.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Nivolumab for adjuvant treatment of completely resected melanoma with lymph node involvement or metastatic disease (NICE TA558): the survival model applied the registry data (American Joint Committee on Cancer; AJCC) to both treatment arms after a certain time point.
Clinical outcome (PFS) intervention	Whether RWD give the figure for progression free survival (PFS) of intervention or not. The progression of disease is important for economic evaluation model in terms of health state transitions and treatment switching. The survival data could come from RCT or RWD.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (RR) intervention	Whether RWD provides the response rate (RR) for the intervention or not. The effectiveness of cancer treatment is often shown by responses of tumour cells, which is evaluated by the RECIST criteria or other criteria. The response rate data would be collected in RCT or other type of data.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (TTP) intervention	Whether RWD give the figure for time-to-progression (TTP) of intervention or not. Some cancer treatments show their clinical effectiveness not through the progression free survival (PFS), but alternatively through time-to-progression.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (AE) intervention	Whether RWD give the figure of adverse event (AE) of intervention or not. Adverse events are crucial information for the estimation of the QALYs. The adverse events are collected in RCT. However, RWD, including cohort studies, retrospective studies, or other type of studies, also provide the information of adverse events, which cannot be found in RCT.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Blinatumomab for treating acute lymphoblastic leukaemia in remission with minimal residual disease activity (NICE TA589): retrospective non-interventional cohort study collected from 2000 to 2017 was used to inform the clinical outcome of comparators as well as adverse event.

(Continued Table A.1.1)

Variable	Explanation	Coding	Example
Clinical outcome (OS) comparators	Whether RWD give the figure of overall survival (OS) of comparators or not.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	Refer to the variable, clinical outcome (OS) intervention
Clinical outcome (PFS) comparators	Whether RWD give the figure for the progression free survival (PFS) of comparators or not.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (RR) comparators	Whether RWD provide the response rate (RR) of comparators or not.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (TTP) comparators	Whether RWD provide the time-to-progression (TTP) of comparators or not.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Clinical outcome (AE) comparators	Whether RWD provide the figure adverse events (AE) for the comparators or not.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	Refer to the variable, clinical outcome (AE) intervention
Transition probability	Whether RWD provide the transition probability from one state to other state, if it is applicable.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Pembrolizumab for treating melanoma with high risk of recurrence (NICE TA553): electronic health records (Flatiron database) collected by cancer care providers in the US was used to model transition from the "locoregional recurrence (LR)" state to the "distant metastases" and life tables for transition from the LR to "death" state.
Health utility of health state (generic)	Whether health state utility survey of generic measurement is done in RWD or RCT. Health state utility is necessary information for the estimation of the QALYs. Generic health utility measurement, EQ-5D, is frequently used. There is national tariff of EQ-5D to get the scores. Hereby, the way of collecting survey (RWD or RCT) is highlighted.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Health utility of health state (condition-specific)	Whether health state utility survey of condition-specific measurement is done in RWD or RCT. In cancer treatment, condition-specific measurement is commonly adopted. Similar to the previous row, the way of collecting survey (RWD or RCT) is highlighted.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Disutility of adverse events	Whether survey of collecting disutility data is done in RWD or RCT. As adverse events are likely to reduce the patient's quality of life, the disutility of adverse events is included in estimates. The way of collecting survey (RWD or RCT) is drawn to attention.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**

(Continued Table A.1.1)

Variable	Explanation	Coding	Example
Resource use (Health state cost) common	Whether resource use for estimating health state cost is derived from RWD or RCT. In economic evaluation, the unit cost mostly comes from the national reference cost. The total cost is calculated by the total resource use (volume of technology and health care services) multiplied by the reference cost. Here, the only resource use is focused in data extraction. Resource use for estimating health state cost includes all activity like monitoring, GP visits, pharmacy cost etc. Health state resource use could be aggregated or individually listed. Here, the difference of describing health state cost is not separately considered.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Axicabtagene ciloleucel for treating diffuse large B-cell lymphoma and primary mediastinal large B-cell lymphoma after 2 or more systemic therapies (NICE TA559): RWD was used for estimating the cost of inpatient admission (data: Hospital Episode Statistics), the cost of home care and hospice (data: National Audit Office), and GP time (data: Personal Social Services Research Unit; PSSRU).
Resource use (end-of-life care) common	Whether resource use for estimating end-of-life care is derived from RWD or RCT. Resource use of terminal cancer patients is not frequently reported in the RCT providing the treatment effect. Therefore, other data resources, including RCTs of other technologies, provide the information of resource use in the end-of-life care.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Resource use (Managing AE) intervention	Whether resource use for managing adverse events of intervention is derived from RWD or RCT. Resource use of managing adverse events is reported in RCTs as well as in other types of researches which can provide alternative perspectives.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Resource use (volume of treatment) intervention	Whether resource use for volume of treatment of intervention is derived from RWD or RCT. In this study, scope of the volume of treatment is limited to the frequency of treatment, frequency of administration, and amount of subsequent treatment.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	- Fulvestrant for treating untreated locally advanced or metastatic oestrogen-receptor positive breast cancer (NICE TA503): a medical chart review study was used to determine the proportion of patient using subsequent treatment for cost calculation.
Resource use (Dose adjustment) intervention	Whether resource use for dose adjustment of intervention is derived from RWD or RCT. There are several reasons for adjusting dose such as adverse events (AEs). The dose of cancer treatments is calculated by BSA (body surface area). This study focuses only on BSA and dose adjustment due to AEs, because these information are commonly reported in NICE appraisals.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**

(Continued Table A.1.1)

Variable	Explanation	Coding	Example
Resource use (Managing AE) comparators	Whether resource use for managing adverse events of comparators is derived from RWD or RCT.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**
Resource use (volume of treatment) comparators	Whether resource use for volume of treatment of comparators is derived from RWD or RCT.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	Refer to the variable, resource use (volume of treatment) intervention
Resource use (Dose adjustment) comparators	Whether resource use for dose adjustment of comparators is derived from RWD or RCT. Since the intervention is a novel technology, RCTs provide less information on the adjustment. RWD could be utilised to provide more relevant information regarding dose adjustment of existing technologies which have been used in routine clinical practice.	No RWD = 0 Yes, data from RWD = 1 Not clear = 9	N/A**

* In order to detect the use of RWD in sensitivity analysis, the parametric part is duplicated.
** As data extraction is not conducted, all of examples are not available at this stage. In this case, it marked as N/A.
*** Benefits/challenges of the use of RWD are collected in outcome variables.
**** In cases where trials have more than two arms, only the arms considered as relevant for decision problem in evidence submission are included. If there are two intervention arms and these arms are separately used for different indications in appraisals, the data extraction is carried out separately. When two arms are relevant as comparators for same indication, the data are recorded without distinguishing these arms.

A.1.2 Variable additionally included in Chapter 4

Variable	Description	Code
Stage of cancer	In this study, the information about the stage of cancer was extracted from the indication. The stage of a cancer is a clinical term, describing the size of a tumour and how far it has spread from where it originated. There are several types of staging systems used for different types of cancer. When drugs are approved, the indication usually includes the rough information about the stage of cancer such as early stage of cancer, advanced cancer or metastatic cancer. For example, an indication of sacituzumab govitecan is for treating unresectable triple-negative advanced breast cancer 2 or more therapies (NICE TA819). Although this information is rough, it is summarised information based on staging systems. Three categories are used for classification. If the indication is for metastatic cancer, it was recorded as 3. If the indication is for advanced, relapsed or recurrent cancer, it was recorded as 2. If there is no specific comment, it was recorded as 1.	Early =1 Advanced = 2 Metastatic = 3

A.1.3 Variables additionally included in Chapter 6

Variables	Description	Code
Change in indication	This change includes change both in the indication (in the scope) and in the recommendation. The indication could be changed if regulatory body changed the approval indication. Also, the	No=0, Yes=1
Uncertainty in FAD (Original appraisal)		
• Key uncertainty	The uncertainty highlighted by appraisal committees in headline or conclusion/CDF sections is considered as a 'key uncertainty.'	Direct quotation
• Any uncertainty in FAD	Any uncertainty addressed by appraisal committees in FADs	Direct quotation
Uncertainty in FAD (CDF review)		
• Still uncertain	If the committee addresses that the uncertainty in the original appraisal is still highly uncertain in the CDF review, the uncertainty is recorded here. Although committee accepts the assumptions, it would be recorded as a "still uncertain" if they note that the remaining uncertainty is high.	Direct quotation
• Resolved uncertainty	The committee agrees that the updated models in review appraisals reduces uncertainty, the uncertainty is recorded here.	Direct quotation
• Newly commented uncertainty	When new evidence was used, the comments on the evidence could be made. The newly addressed uncertainty would be recorded here.	Direct quotation
SACT		
• PHE report attached in guidance	Public Health England (PHE) reports the outcome of SACT cohort study in CDF review appraisals. The report can be either attached or not attached in guidance. Regardless use of SACT in economic evaluation, this variable focuses on only presence of PHE report in guidance.	No=0, Yes, attached=1
• Median follow-up period	Median follow-up period in SACT cohort. Median follow-up of the study is routinely reported. However, if the information is not available, median follow-up time with which overall survival (OS) benefit is presented is recorded.	Number (unit: month)
• Number of patients in SACT cohort study	The number of patients identified in SACT dataset for main analysis cohort	Number
• Reason for non-use of SACT dataset in economic evaluation	Although companies do not always give a reason of not using SACT data, a few companies provide the reasons. If the reasons of not using the dataset is presented in company submissions, it will be recorded here.	Direct quotation from company submission

Appendix 2. Interview: Understanding different views on use of real-world data in health technology assessment from stakeholder interviews

A.2.1 Interview summary

1) Research questions

- What is the limitation of using FDA's definition of RWD in a context of NICE appraisal?
- To what extent RWD are able and available to provide useful information in NICE appraisals?
- How will use of RWD change in economic modelling in future NICE appraisals?

2) Methods

Data for this study were collected through interviews. The sampling of stakeholders and interview protocol were developed and approved by Ethics Committee at the London School of Hygiene and Tropical Medicine (LSHTM). Sixteen representatives from each five stakeholders participated in the interviews (Table A2.1). Information for identifying representatives was retrieved from website and/or the authors' professional network. All representatives were approached by email using a standardised invitation. Due to lower response from patient organisation, only two interviewees were included in this study. The interview was a part of a research with individual groups participating in a NICE appraisal to understand practical issues they have faced in the use of RWD in the NICE appraisal. Sixteen people were invited for the interview. Ten experts from consultancy, ERGs and appraisal committees were selected for the question of grading.

Table A.2.1 Interview participants

Group	Number (n=16)
Company/Consultancy	3
NICE technical advisor	4
ERG	3
Appraisal Committee	4
Patient organisation	2

An interview guide was developed on the basis of literature on use of RWD in HTA (Table A2.2).

Table A.2.2 Summary of interview topic

Topic	Content
Definition of RWD for NICE appraisals	· Appropriate definition of RWD for NICE appraisal · Relevant sources of RWD for NICE
Value of use of RWD in NICE appraisals	· Benefit from additional data in four areas* · Technical ability of RWD in four areas · Practical ability of RWD in four areas · Critical components in economic modelling
Use of RWD in future NICE appraisals	· Prerequisite to the reliable use of RWD in NICE · Future trend of use of RWD
* Four areas: supplement information on survival distribution, information for comparators, generalisability, information about rare disease	

The interview consisted of three parts. First, the opinions on the definition of RWD were collected. One of the commonly used definition is a FDA's definition, data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources (95). Interviewees were asked about how well this definition captures relevant sources of real-world data in a context of NICE appraisals. Also, relevant sources of RWD for NICE economic models were asked during the interview. Second part of interview was about the value of use of RWD. In this part, the value of RWD was measured by using survey type of question in three domains: areas where additional benefits are expected from additional RWD, areas where RWD are technically able to provide the information and areas where RWD are practically able to provide the information. It is followed by asking experts about how they grade the importance of elements in economic evaluation to get feedback on the intensity scale. In the last part of the interview, the prerequisite to reliable use of RWD in NICE decision making and use of RWD in future NICE appraisals were asked to understand the vision of RWD by different stakeholders.

Interview was conducted by author from December 2020 to July 2021 on online video conferencing programme Zoom. All interviews were carried out as a one-to-one interview for an hour. The

interviews were digitally recorded and transcribed on Zoom. The transcripts were entered into NVivo for qualitative data management.

3) Result

Part 1. Definition of RWD

In the interview, most interviewees were satisfied with the FDA's definition of RWD.

"Definition is not important. How much useful information we can get is important." (Company 1)

Various types of RWD were presented to discuss about relevant data for the HTA process. Claims data are not particularly useful for NHS context. However, SACT data could be seen as claims data in a way. There were mixed opinions on the interview data or data which collected from routine practice setting, however only collected one time. Some people interpret the definition that the data collected from routine practice are RWD whereas other think that routinely collected data can provide more meaningful information.

Part 2. Value of RWD

· Area where can get benefits from additional information

<Priority>			
	H	M	L
Corroborating survival distribution	80%	10%	.
Comparator information including treatment effects	60%	20%	10%
Provide information for generalisability	30%	30%	30%
Information for rare disease	40%	40%	10%

Figure A.2.1 Priority in drug appraisals by different stakeholders

Overall, participants agreed that the issues presented during the interviews are important and key issues in HTA. There was a consensus that the more data the better. Since assumptions about survival distribution models are uncertain, it would be of benefit if more information is available during the time of appraisals. However, the amount of benefit appraisals could have might be different from each appraisal since what happens in RCTs is likely to be different. Companies are less incentivized if additional data do not have good impacts. For example, there is less incentive to show the results with additional data such as RWD in sensitivity analysis if there is no clear additional benefit to analyse the data.

· Technical ability of real-world data

<Technical ability>			
	H	M	L
Corroborating survival distribution	30%	60%	.
Comparator information including treatment effects	40%	50%	10%
Provide information for generalisability	30%	70%	.
Information for rare disease	20%	70%	.

Figure A.2.2 Technical ability of RWD to provide information

This interview question asked whether RWD are technically able to provide the information about corroborating survival distribution, comparator information, providing information for generalisability, information about rare disease if the data are good enough to use. Some positivity was observed, however, it was moderate level. About providing information for rare disease, several accounts were found that rare disease patients were more likely to participate in clinical study. In the rare disease registry, most patients participate in randomised controlled trials; therefore, less people are available for RWD. The quality of rare disease registry might not be tenable.

· Practical availability

<Practical availability>			
	H	M	L
Corroborating survival distribution	10%	60%	10%
Comparator information including treatment effects	10%	60%	20%
Provide information for generalisability	10%	60%	20%
Information for rare disease	10%	50%	30%

Figure A.2.3 Practical availability of RWD to provide information

Participants, in general, acknowledged the practical issues which RWD cannot solve addressed issues. Top challenge in using RWD for economic modelling is confounding and potential bias. Also, quality of the data was addressed a challenge. The number of patients and follow-up might not be enough for the decision-making. One of ERGs pointed out an issue with respect to using systemic anti-cancer therapy (SACT) dataset in the ERG review that the data were too immature to be used in the appraisal.

· Major components for ordering intensity of use of RWD

The variables in extraction template were presented to experts. They were asked to pick the three the most critical factors in economic evaluation. Most interviewees selected overall survival, volume of treatment and choice of comparators, which have the potential having a major impact on the result of the cost-effectiveness analysis while all components of economic evaluation indirectly impact on cost-effectiveness estimates.

Table A.2.3 Major components in the use of RWD

Components	Reasons
Overall survival	In appraisals of oncology medicine, OS is a critical piece of evidence as a main clinical outcome. It is directly linked to quality-adjusted life years (QALYs), and thus to the ICER.
Volume of treatment	As the drug price tends to be expensive, the cost is more likely to be changed depending on the quantity of the drug used.
Choice of comparators	The choice of comparators is also important axis of cost-effectiveness analysis. Depending on the comparators, the result of cost-effectiveness could be different. Therefore, selecting proper comparator which reflects practice is important.

Part 3. Vision of RWD

· Prerequisite

Maturity of the data are commonly addressed. Some interview participants emphasised that details about disease characteristics were essential to provide right information to decision makers. In line with the emphasis on the quality of the data, the equipoise in the use of RWD was also emphasised. Although there is interest in the use of RWD, the data should not be used when the quality is not guaranteed.

"If RCT is not good, it should not be used. The use of RWD should not be based on company's willingness to use it." (ERG 3)

· Foresee in the future

Overall, more use of RWD was expected. However, the attitudes toward RWD are not always positive. Some participants were cautious about using RWD even though they anticipated that RWD would be used more and more over time.

"RWD is a buzz-word." (committee 2)

"We cannot see the future, but RWD helps." (ERG 2)

A.2.2 Interview script

1) Introduction

Hi *name of interviewee*, how are you? It is pretty cold these days.

(Inform the record state)

• *When interviewee refuse the recording:* Before start, I inform that this interview is not recorded as you requested.

• *When interviewee agree with recording:* Before start, I inform that this interview is being recorded.

Thank you for participating in this interview. As I already told you in the email, I will ask you about your opinion on RWD in health technology assessment, HTA. Today's your answer to the following questions will help me understand diverse perspectives on real-world data in the context of HTA and tackle the key challenges in incorporating RWD in evidence synthesis. The scope of this interview is limited to English context, which is mainly NICE appraisal of oncology medicine. This interview should take one hour. If at any point of the interview, you need a break, please let me know. Okay. Do you have any question so far?

• *If interviewee has question:* Thank you for your question. (answer) I hope it help. If you don't have any other question, can I start the interview? Thank you.

• *If interviewee does not have any question:* Okay, good. Let's start the interview.

2) Interview part 1: definition of RWD in HTA

In the first part of the interview, I will share FDA definition of RWD and various types of RWD with you and ask you some questions. I will share my screen.

(Share the screen – slide 1)

The slide shows the definition of RWD by FDA. As an umbrella term, FDA defined that real-world data are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. How do you find this definition? Is it too narrow or too broad to be used in HTA?

• *If the answer needs more details:* Could you explain little bit more why you think so?

There are various types of RWD. *(Click the button)* Electronic medical records, registry and claims data are commonly mentioned in HTA. What do you think of these data? Are they particularly useful or not useful in NICE context?

• *If the interviewee says yes:* Could you tell me why they are useful in HTA?

• *If the interviewee says no:* Could you tell me why they are not useful in HTA?

Do you think there is other types of RWD useful in HTA?

I will give you a case to listen what you think of these studies. *(Click the button)*

Some studies are conducted in general population and done one time outside clinical trials. Do you think these studies are eligible to be categorised in RWD?

• *If the interviewee says yes:* Could you explain why you think it is RWD?
• *If the interviewee says no:* Could you explain why you don't think it is RWD?

(Stop sharing the screen)

As we see, RWD includes rather broad range of data. Each dataset is likely to have different information. Do you think a certain type of RWD can provide more information in HTA?

• *If the interviewee says yes:* Which type of RWD can do more and could you explain why you think it is RWD?

• *If the interviewee says no:*

- Do you think there is no big differences among types of RWD?

- If so, can you tell me which common traits of RWD should be considered when being used in HTA?

3) Interview part 2: value of use of RWD

Thank you for your answers. Now we will move to the second part of the interview. I will share the screen again. *(Share the screen – slide 3)*

RWD is expected to help to assess the cost-effectiveness of new health technology by filling the evidence gap. The screen shows the area where additional data can add more value.

Could you tell me how much it would be of benefit to have additional information in these areas? *(Click the button)* If it is rather critical and top priority to have additional information, please answer high. If it is important, but not a top priority, medium. If it is less important, please answer low. For supplementing information of survival model, how important is it to having additional information? (answer)

How about giving information of comparators? (answer)

and for increasing generalisability of appraisal? (answer)

Last, for giving more information on rare disease. (answer)

(Stop sharing the slide)

You already quickly mentioned reasons. Can you explain little bit more why you think *selected area* is important?

Apart from above cases, is there any particular circumstance where RWD can fill the evidence gap?

I am going to share the screen again.

(Share the screen – slide 4)

How much do you think RWD is technically able to fill these gaps? *(Click the button)* Please answer it using the scale highly able, moderate, low on the screen. If you think RWD is able to answer the question fully, please answer highly able. If you think RWD is partially able, moderate. If you think RWD is not able to do currently, please answer not able.

For supplementing information of survival model, how much is RWD able to do? (answer)

How about giving information of comparators? (answer)

and for increasing generalisability of appraisal? (answer)

Last, for giving more information on rare disease. (answer)

(Stop sharing the slide)

• *When answer is highly able:* You answered that RWD is highly able to do in *selected area*.

- What is the strength of using RWD in that area?

• *When answer is moderate:* You answered that RWD is moderately able to do in *selected area*.

- What is the big challenge to give full information?

- Do you think it could be overcome?

• *When answer is low:* You answered that RWD is barely able to do in *selected area*.

- Why do you think RWD is not able to answer it?

- What is the big challenge to use RWD in the area?

This question has similar structure of previous question.

(Share the screen – slide 5)

I will ask about practical availability of RWD. Is RWD currently available, which is needed to provide the information? Please use the scale highly, moderate, not available on the screen. *(Click the button)*

For supplementing information of survival model, is relevant RWD available to do so? (answer)

How about giving information of comparators? (answer)

and for increasing generalisability of appraisal? (answer)

Finally, for giving more information on rare disease. (answer)

(Stop sharing the slide)

• *When answer is highly available:* You said that RWD is highly available to do in *selected area*.

- Could you specify which type of RWD, for example EMR, registry or claims data is most suitable to do so?

• *When answer is moderate:* You answered that RWD is moderately available to do in *selected area*.

- Why do you think RWD is partially available?

• *When answer is low:* You answered that RWD is not available to do in *selected area*.

- What do you think the biggest challenge of having RWD to fill the gap?

Across the questions, is there any advantage or disadvantage of using RWD you want to emphasise? It is okay to talk about either overall use or specific case.

Thank you for sharing your view with me. I will move to the next question.

(Share the screen – slide 6)

On this slide, I identify the components of economic evaluation where RWD is likely to be used. I classified two big groups, non-parametric use and parametric use. Non-parametric use is to use RWD to establish the model structure of economic evaluation. On the other hand, parametric use is to use RWD to determine the value of parameters in the model.

Could you pick three the most critical components in economic evaluation?

• *Sub-question 1*: Why are they important?

• *Sub-question 2*: Can you say RWD is intensively used in appraisal when RWD being used in these components?

• *Sub-question 3*: Is there any benefit or concern to use RWD in these components?

Thank you. Now, please pick three least critical components in economic evaluation.

• *Sub-question 1*: Why do you think they are less important?

• *Sub-question 2*: Can you say it is minor use of RWD when RWD are used for the components in the appraisal?

Apart from these components on the slide, is there any other part of economic evaluation in appraisal where RWD can be used?

• *If yes:* Could you give me little more explanation how RWD could be used in *selected area*?

Thank you for sharing your opinion. Move on. I am going to show you another slide.

(Move to the next slide – slide 7)

On this slide, you can see five hypotheses about the greater use of RWD.

What do you think of the hypotheses? Do you agree that these factors are likely to be associated with greater use of RWD?

Is there other factor potentially associated with the greater use of RWD?

Thank you for your answer.

Using the hypotheses, I conducted regression analysis in hundred sixty-three appraisal. I am going to share some of statistical results from this analysis.

(Move to the next slide – slide 8)

In the regression, the data were aggregated to any use of RWD, parametric use of RWD and non-parametric use of RWD. The slide shows the positive association in the regression with aggregated data. There were association between presence of previous TA guidance of same technology and any use of RWD and association between absence of head-to head trial and less use of non-parametric use of RWD. In addition, the association between longer time horizon of economic evaluation model and more use of RWD in non-parametric way was found. How do you find these results?

• *Sub-question 1*: Do these results surprise you or do you a sort of expect it?

• *Sub-question 2*: Could you little bit more explain why you think so?

I will share the results which comes after breaking down the level.

(Move to the next slide – slide 9)

I've found positive association especially between long time horizon of economic evaluation model and choosing survival distribution of intervention and comparators.

How do you find this result?

• *Sub-question 1*: Why do you think this result comes out?

(Move to the next slide – slide 10)

The association between external validity of RCT and using RWD for choosing survival distribution was also found. When external validity of RCT is low, RWD is more likely to be used in choosing survival distribution of intervention. What do you think of this result?

And in estimating volume of treatment of intervention and dose of adjustment of intervention and comparators, there are associations with moderate external validity of RCT not low. Why do you think these results come?

(Move to the next slide – slide 11)

When testing hypotheses, I found some negative relationships. When the time horizon of economic evaluation gets longer, RWD is less likely to be used in determining characteristics of population of economic evaluation, time-to-discontinuation of intervention, and estimating dose-adjustment of intervention and comparators. How do you find these results? Is it surprising or could be expected?

• *Sub-question 1*: Why do you think this result comes out?

(Move to the next slide – slide 11)

Another negative result. When the incidence of disease is low, RWD is less likely to be used in estimating overall survival of intervention. Also, when internal validity of RCT of intervention is low, RWD is less likely to be used in estimating overall survival and progression-free survival. What do you think of this result? Again, is it surprising or being expected?

• *Sub-question 1*: Could you little bit explain more why you think so?

• *Sub-question 2*: Do you think the preference to RCT can be related to this result? For example, if RCT is not good enough, network meta-analysis is used in order to increase the validity rather than using RWD, which is commonly questioned about confounding factors etc?

• *Sub-question 3*: What is the biggest challenge of using RWD in clinical outcome do you think?

(Stop sharing the slide)

Thank you for the answer. Now it is the last part of today's interview.

4) Interview part 3: future use of RWD in HTA

(Share the screen – slide 12)

In early this year, NICE published a statement of intent that NICE would accept a broad range of data resources including 'real-world data', such as registries and clinical audits.

(Stop sharing the screen)

I would like to ask what you think of current interest in RWD. I am not asking this call is good or not, but do you think the current interest in using RWD in NICE appraisal is inflated or still less emphasised?

• *If it is inflated*:

- Could you tell me more in which perspective do you think it is inflated?
- What does make it inflated?

• *If it is less emphasised*:

- Could you tell me more which part of using RWD in HTA should be more emphasised?
- What does make it less emphasised?

What is the prerequisite to incorporate RWD reliably in appraisal?

How do you foresee the use of RWD in NICE appraisal in future?

5) Conclusion

I think that is it. Thank you for your time. It was great time for me to understand the benefit and challenges of RWD in NICE appraisal from an expertise' view. I wonder how you find this interview. (answer & quick chat)

If you want to provide additional information, which would be helpful for the research or you are curious about this interview after this, please email me. I am very happy to share more with you. Alright. Thank you very much, and Merry Christmas in advance. Have a lovely day. Bye.

Appendix 3. Use of real-world data in each component in data extraction tool

1. Any use of RWD in economic models

Figure A.3.1 shows any form of use of RWD in economic models in manufacturer submissions over time. The number of appraisals has increased despite small drops in 2019 and 2020. Since the early year, more than half of single technology appraisals (STAs) have used RWD in economic models. One hundred and ninety-two appraisals (84%) used RWD in some parts of the submitted cost-effectiveness analysis. In 2020, 89% of included STAs used RWD in economic models in manufacturer submission.

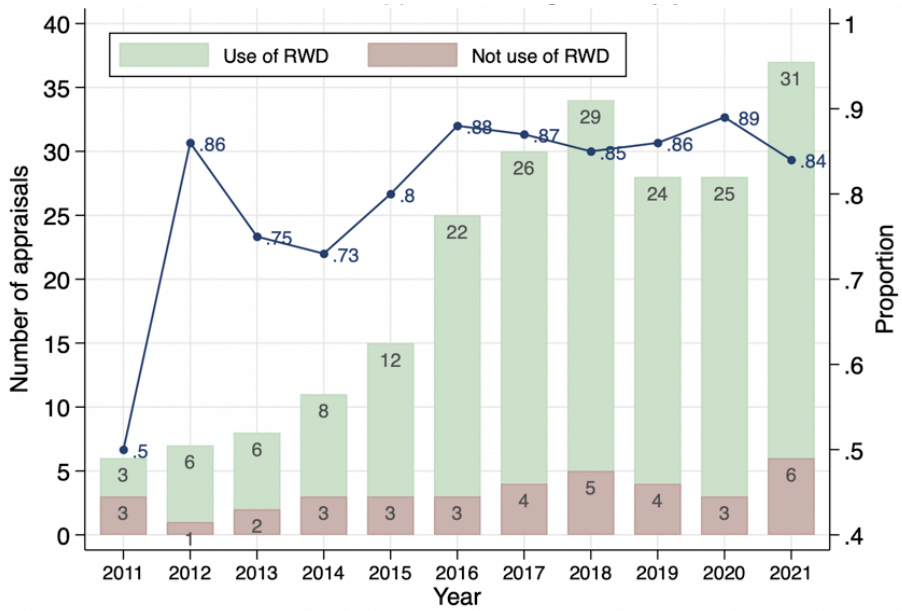


Figure A.3.1 The number of appraisals making any use of RWD in economic models

Figure A.3.2 shows the number of STAs using RWD in economic models by the type of cancer. RWD was used in any parts of economic models in STAs of eight types of cancer (colorectal cancer, neuroblastoma, liver cancer, oesophageal cancer, pancreatic cancer, skin cancer stomach cancer, genomic biomarker-based cancer and thyroid cancer), although number of appraisals for these cancers was relatively small. In blood and bone marrow cancer, the largest number of appraisals have used RWD in economic model is (50 appraisals, 80% of appraisals of blood and bone marrow cancer). Second largest number was lung cancer, thirty-nine appraisals (91.4% of appraisals of lung cancer). As

a proportion of using RWD in economic models, appraisals of lung cancer show the highest proportion except for small number of cancers.

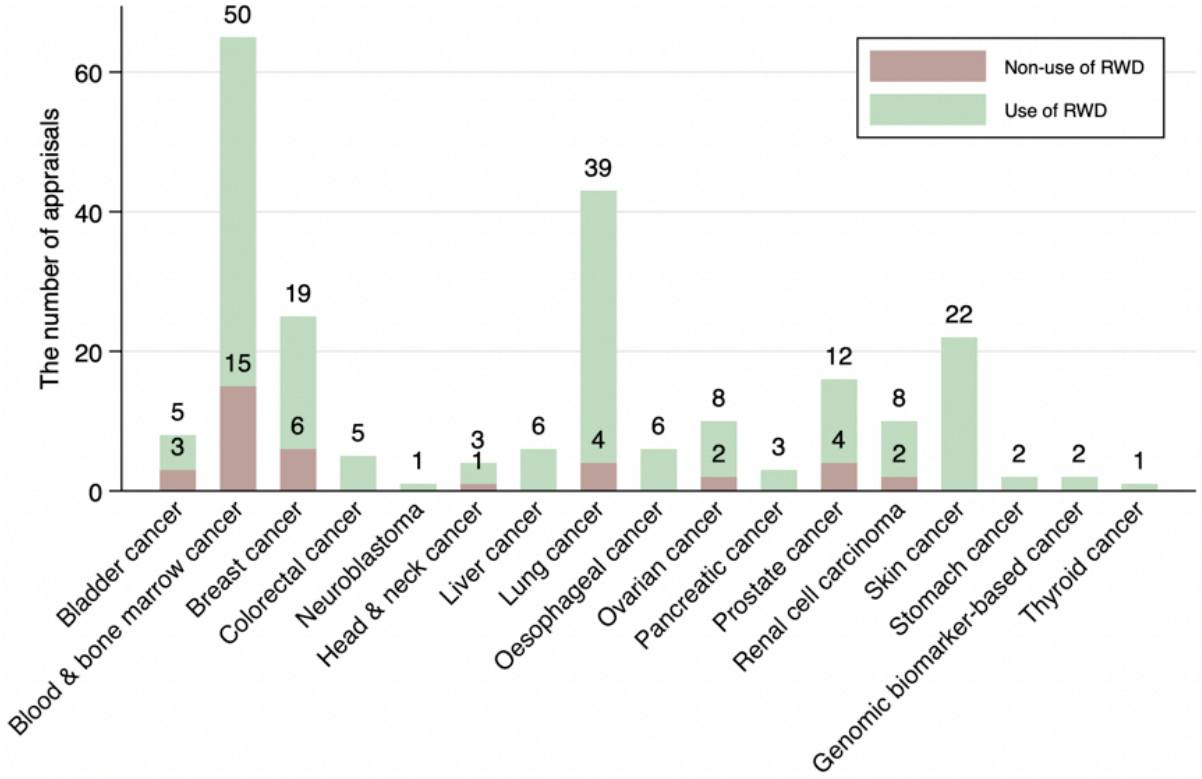


Figure A.3.2 Use of RWD by type of cancers

2. Parametric and non-parametric use of RWD

Although reviewing any use of RWD can provide a broad picture of use of RWD in appraisals, it is too blunt to tell where and how RWD is used. Distinguishing between parametric and non-parametric use of RWD provides a more detailed picture of where RWD has been used in appraisals. Parametric uses of RWD (Figure A.3.3) and non-parametric use of RWD (Figure A.3.4) are separately presented.

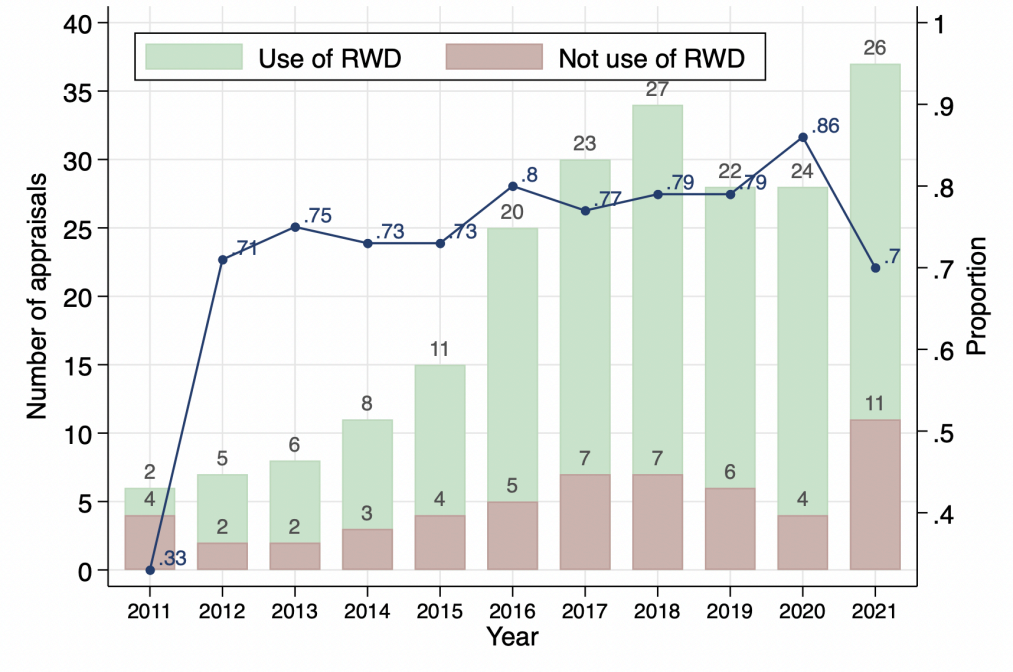


Figure A.3.3 The number of appraisals in parametric use over year

Overall, RWD have been used more often for parametric purposes. One hundred seventy-four appraisals included RWD to determine one or more parametric components (76%). Parametric use of RWD indicates the use of RWD to inform the value of parameters in the economic model. Parametric use includes clinical outcome, health utility value and resource use which are essential input for cost-effectiveness analysis. Surprisingly, RWD have been used since early appraisals to estimate the parameters in economic evaluation. The parametric use of RWD were found in more than 70% of appraisals each year apart from 2011. In very early time appraisals, in 2011, only one third of STAs made parametric uses in economic model. In 2020, the highest proportion of parametric use of RWD was observed (86%). Although the parametric use was observed quite steadily over time, the small drop was observed in 2021.

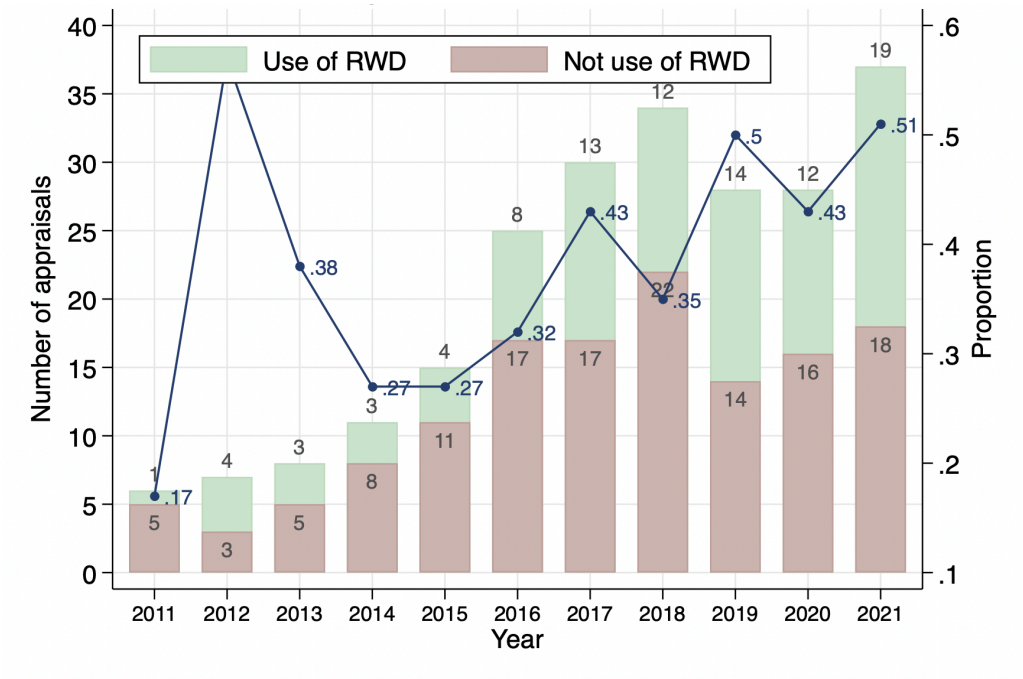


Figure A.3.4 The number of appraisals in non-parametric use over year

Non-parametric use of RWD was less frequently found in economic models. Non-parametric use of RWD refers that RWD are used to justify or validate the model assumptions. Non-parametric use includes use of RWD to support the choice of comparators and validate the survival distribution. Ninety-three appraisals used RWD to inform non-parametric components in the economic models (40%). The highest proportion of non-parametric use was in 2012. However, the number was quite limited as the total number of STAs in 2017 was only seven. In 2021, the number of non-parametric use was increased (n=19, 51%). The fluctuation of non-parametric use in economic models was found. Despite the fluctuation, the direction of non-parametric use indicates the increasing uses.

Figure A.3.5 presents a matrix by non-parametric and parametric uses. The x-axis records the number of parametric uses of RWD, and the y-axis records non-parametric uses of RWD. The matrix shows thirty different combinations of non-parametric use and parametric use by number of elements included in each type. The largest group are the 37 appraisals where no use of RWD was made. Of the remaining appraisals, twenty-two estimate a single parameter using RWD, and twenty-seven base two parameters on RWD.

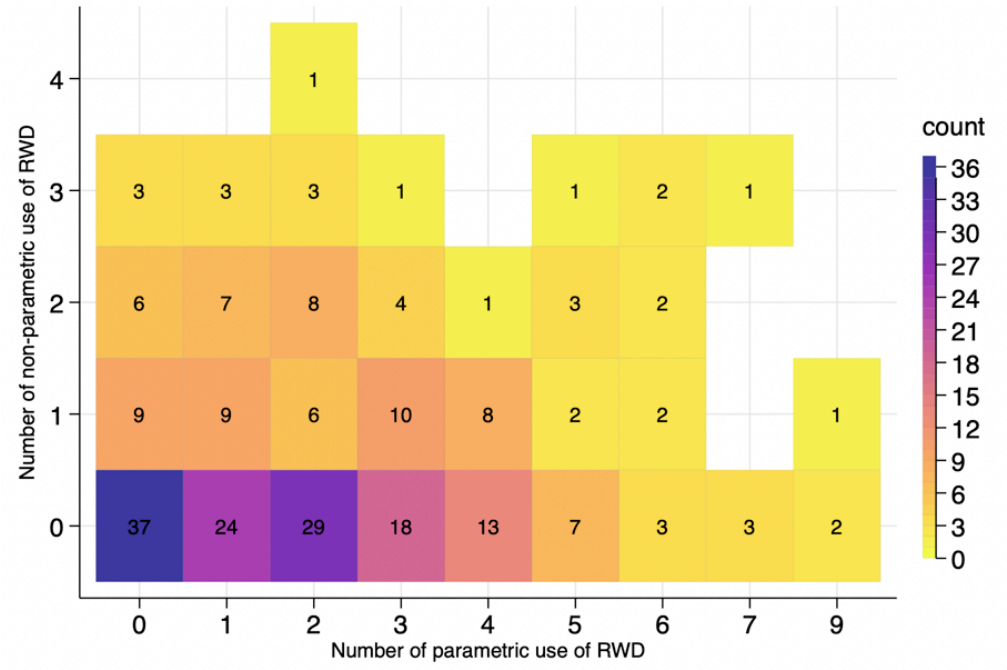
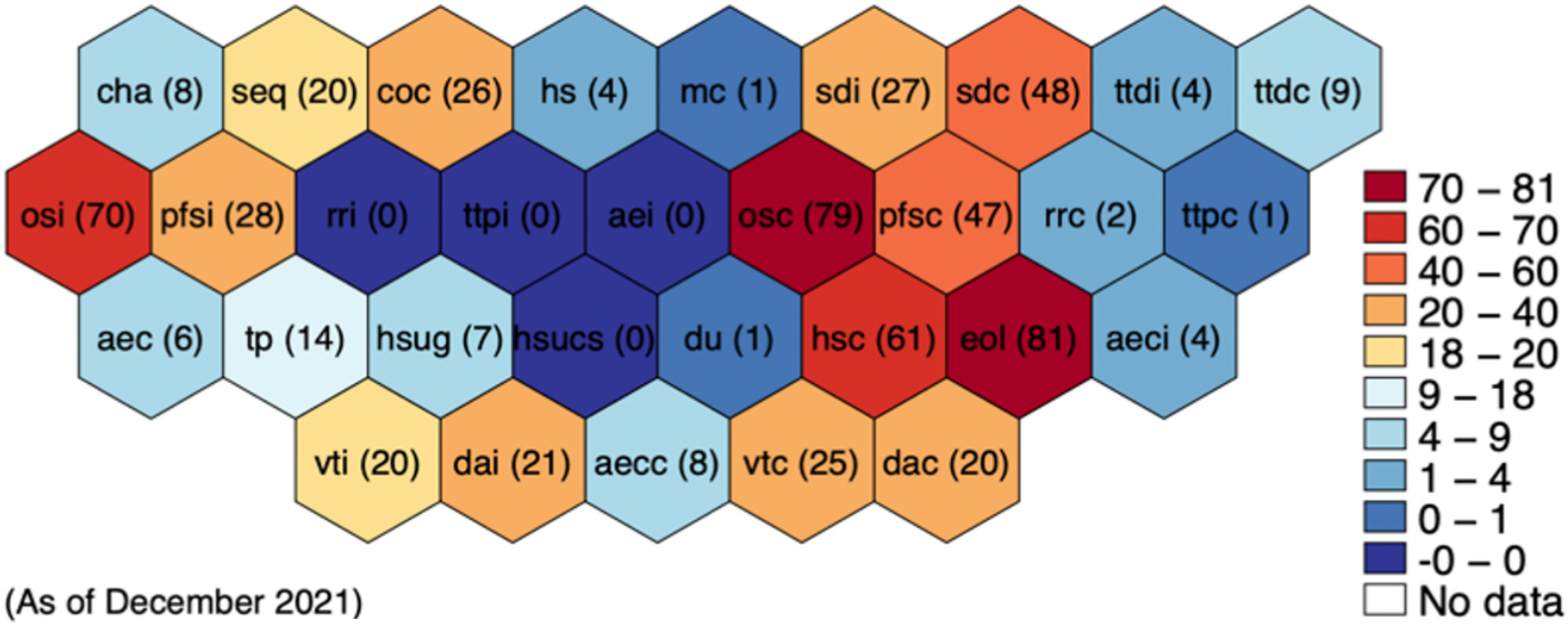


Figure A.3.5 Matrix of use of RWD by types of use

3. Single use of RWD in economic model

Figure A.3.6 shows where RWD were used in single components in economic models by breaking down the level. The components in parametric use show the high number of the use of RWD. The resource uses for health state cost and end-of-life cost are the main use of RWD in resource use. Different from the expectation that less use of RWD in treatment effect and more use of RWD for resource use, large number of STAs used RWD to estimate the overall survival in economic model (30% of included STAs).



cha	Characteristic of population	osi	Overall survival (intervention)	aec	Adverse event (comparators)	dai	Dose adjustment (intervention)
seq	Treatment sequence	pfsi	Progression-free-survival (intervention)	tp	Transition probability	aecc	Resource use of adverse event cost (comparators)
coc	Choice of comparators	rri	Response rate (intervention)	hsug	Health state utility (generic measure)	vtc	Volume of treatment (comparators)
hs	Health state	ttpi	Time-to-progress (intervention)	hsucs	Health state utility (cancer-specific)	dac	Dose adjustment (comparators)
mc	Model cycle	aei	Adverse event (intervention)	du	Disutility		
sdi	Survival distribution (intervention)	osc	Overall survival (comparators)	hsc	Resource use of health state cost		
sdc	Survival distribution (comparators)	pfsc	Progression-free-survival (comparators)	eol	Resource use of end-of-life treatment		
ttdi	Time-to-discontinuation (intervention)	rrc	Response rate (comparators)	aeci	Resource use of adverse event cost (intervention)		
ttdc	Time-to-discontinuation (comparators)	ttpc	Time-to-progress (comparators)	vti	Volume of treatment (intervention)		

Figure A.3.6 Frequency of use RWD for different components

RWD are frequently used to inform the medical resource use such as GP visits and hospitalisation. In general, resource utilisation requires supplementary data to reflect the resource use in the routine clinical practice such as regular check-up exams and frequently used subsequent treatment. The resource use in routine practices is likely to vary from that in clinical trials where the patient monitoring is tightly controlled. As a consequence, there is greater scope for the use of supplementary data, which reflect current practice.

An example of the use of RWD in resource use of health state cost is the appraisal of darolutamide with androgen deprivation therapy for treating hormone-relapsed non-metastatic prostate cancer (NICE TA660). In this appraisal, the company used RWD to estimate the frequencies of resource utilisation such as GP visit in the base-case since there was no healthcare resource use frequencies were reported for the study population (132). The applied RWD was longitudinal retrospective cohort study data from a large National Health Service (NHS) trust, which used both structured data from hospital electronic medical records and unstructured information derived from clinical notes. The primary outcome of this study was per cycle of frequencies of different monitoring events. The company expected that it reflected the current practice as selected NHS trust covered large area of over 750,000 for secondary care and more than 5 million patients in tertiary care. Here, RWD took part in the appraisals.

Use of RWD for overall survival was made in one third of included STAs. RWD were used for not only treatment effect but also adjusting background mortality or disease hazard. Another way to use of RWD was to extrapolate survival curve after clinical data cut. An example is the appraisal of nivolumab for adjuvant treatment of resected stage III and IV melanoma (TA588, replaced by ta684). In this appraisal, RWD, Melanoma registry data were used for transition probabilities on both arms from 10 years.

RWD were also used for corroborating or validating chosen survival curve. For example, an appraisal of atezolizumab with carboplatin and etoposide for untreated extensive-stage small-cell lung cancer

(TA638) used RWD to pick the survival distribution and validate the robustness of survival model (225). The company argued that the best fit survival models according to Akaike Information criterion and Bayesian Information criterion are too conservative to be considered clinically plausible. Instead, the log-logistic extrapolation, the next best fit was selected, which from a visual fit gave the closest estimate of long-term survival to the real-world data.

4. Use of RWD in sensitivity analyses

The study also reviewed the use of RWD in sensitivity analysis, separating from base-case analysis. In economic evaluation, a sensitivity analysis is an approach to deal with parameter uncertainty and methodological assumptions in the analysis (133). In company submission, alternative input resources for the efficacy and cost inputs, which can be possible driver of uncertainty are tested in a section of sensitivity analysis. It usually represents uncertainty by varying parameter values by some specified amount. In line with it, different efficacy and cost scenarios are also evaluated. The study analysed how frequently RWD was used in sensitivity analysis as supplementary evidence. Although the more use of RWD in sensitivity analysis was expected, the limited use of RWD was made. Only twenty-three appraisals used RWD in manufacturer's submission. The analysis of parametric use of RWD in sensitivity analysis found that small number of appraisals using RWD in a sensitivity analysis and the number was less than the number of appraisals using RWD in the base-case analysis.

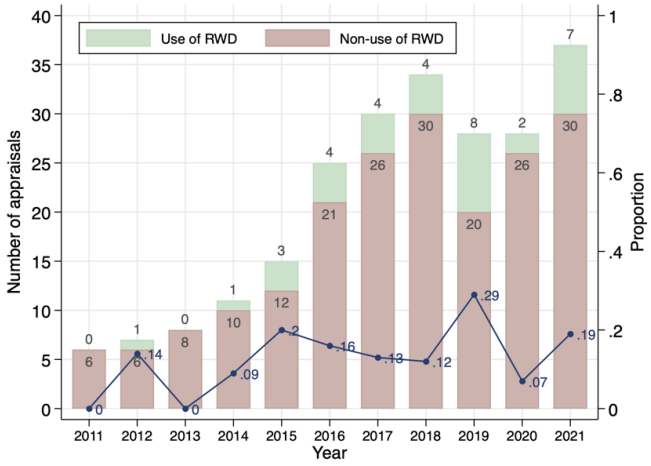


Figure A.3.7 The number of appraisals making any use of RWD in economic models in sensitivity analysis

In lung cancer, the highest proportion of use of RWD in sensitivity analysis was observed (23%, 10 STAs). In the STAs of blood and bone marrow cancer, 20% of STAs made the use of RWD in sensitivity analysis (n=13).

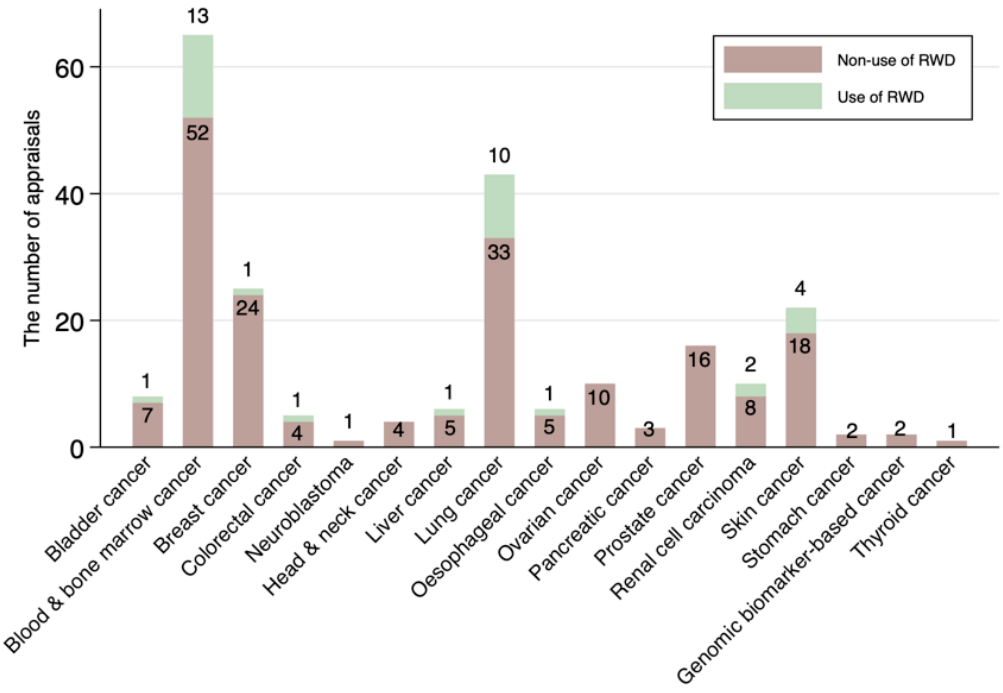


Figure A.3.8 Use of RWD by type of cancers in sensitivity analysis

Case study – use of RWD in sensitivity analysis in lung cancer appraisal

While clinical trials are dominantly used in sensitivity analysis, RWD is used in a few appraisals in different ways. The appraisal of pembrolizumab with carboplatin and paclitaxel for untreated metastatic squamous non-small-cell lung cancer (NICE TA600) used market share data to estimate the volume of subsequent treatments in the sensitivity analysis. Registry data was used to estimate the discontinuation of subsequent treatment in a sensitivity analysis in the appraisal of pomalidomide for relapsed and refractory multiple myeloma previously treated with lenalidomide and bortezomib (NICE TA338). In the appraisal of atezolizumab with carboplatin and etoposide for untreated extensive-stage small-cell lung cancer (NICE TA638), a registry is used for OS of comparators in a scenario analysis. Registry data (Flatiron Health) is incorporated into the extrapolation model for long-term survival. Another example of using RWD is the appraisal of polatuzumab vedotin with rituximab and

bendamustine for treating relapsed or refractory diffuse large B-cell lymphoma (NICE TA649). In this appraisal, the comparator is rituximab in combination with one or more chemotherapy agents. The company chose the combination of bendamustine and rituximab regimen in the base case as it was not feasible to conduct a robust treatment comparison with other comparator regimen due to the limited evidence available. In a scenario analysis, the company included an additional comparator, rituximab, gemcitabine and oxaliplatin regimen and assumed that the efficacy of additional comparator is equivalent with comparator in base-case. This was supported by recent real-world data, US Veterans Health Association database, demonstrating no OS difference between target population treated with either regimen. Although ERG found that this scenario analyses to be uninformative, this example shows where RWD can be used in economic evaluation.

Results of interview to understand the use of RWD in sensitivity analysis

The interview with key players in NICE appraisals helps to draw an implication. First, the manufacturer is less likely to present results of analysis of RWD in sensitivity analysis if the data hardly provide the additional benefit in appraisals. Processing RWD required lots of resource in terms of collection and analysis. If there is no absolute motivation to use RWD in sensitivity analysis, manufacturers rather use other published RCTs to explore the uncertainty in their model input and survival distribution.

During the evidence synthesis, the best evidence for the most suitable assumptions of model is selected in a main analysis. Then, remaining uncertainty around parameters is explored in the sensitivity analysis using the alternative evidence which is reviewed but not used in the base-case analysis. Clinical trials which have similar decision problem are more frequently used in scenario analysis than RWD. As each appraisal has unique issues which inherit the uncertainty in the decision making, parameters chosen for the sensitivity analysis are likely to be different. Among parameters, the uncertainty around the survival outcome and subsequent treatment is mainly highlighted. The extrapolating survival rate is one of the main sources of uncertainty. It is a standard practice to

investigate results of survival distributions, which are not chosen in base-case analysis. Not only testing the survival distribution, but also different source of clinical effectiveness is used in sensitivity analysis. For example, different hazard ration is applied as alternative clinical trial of comparators is used for the effectiveness of comparators. The uncertainty subsequent treatment is also tested using mainly clinical trials.

Appendix 4. Patterns of the use of real-world data

A.4.1 Patterns of the use of real-world data in appraisals of targeted cancer therapy and non-targeted cancer therapy

a. Pattern of use of RWD (Any use without considering non-parametric/parametric use)

Patterns	All appraisals	Non-targeted	Targeted
No use of RWD	37 (16.16%)	25 (17.12%)	12 (14.46%)
Estimating OS of intervention and comparators	13 (5.68%)	12 (8.22%)	1 (1.20%)
Estimating end-of-life resource use	12 (5.24%)	4 (2.74%)	8 (9.64%)
Estimating end-of-life resource use & resource use of health state costs	8 (3.49%)	4 (2.74%)	4 (4.82%)
Estimating resource use of health state costs	7 (3.06%)	5 (3.42%)	2 (2.41%)
Estimating OS of intervention and comparators and end-of-life resource use & resource use of health state costs	6 (2.62%)	4 (2.74%)	2 (2.41%)
Estimating OS and PFS of intervention and comparators and resource use of health state costs	5 (2.18%)	3 (2.05%)	2 (2.41%)
Validating survival distribution of intervention and comparators and estimating end-of-life resource use	5 (2.40%)	1 (0.68%)	4 (4.82%)
Estimating OS and PFS of intervention and comparators	5 (2.40%)	1 (0.68%)	4 (4.82%)
Estimating end-of-life resource use and dose adjustment of intervention and comparators	4 (1.75%)	4 (2.74%)	0 (0%)
Estimating volume of treatment for intervention and comparators	3 (1.31%)	2 (1.37%)	1 (1.20%)
Estimating OS of intervention and comparators and resource use of health state costs	3 (1.31%)	3 (2.05%)	0 (0%)
Validating survival distribution of intervention and comparators	3 (1.31%)	1 (0.68%)	2 (2.41%)
Choosing comparators	3 (1.31%)	3 (2.05%)	0 (0%)
Choosing comparators and estimating resource use of health state costs	3 (1.31%)	2 (1.37%)	1 (1.20%)
Others*	115 (48.91%)	72 (49.33%)	43 (48.21%)
Total	229 (100%)	146 (100%)	83 (100%)

* The patterns are not listed because each represent a particular pattern of use of RWD only observed once or twice.

OS: Overall survival, PFS: Progression-free survival

b. Pattern of use of RWD (Non-parametric use)

Pattern	All appraisals n (%)	Non-targeted n (%)	Targeted n (%)
No use of RWD	136 (59.39%)	91 (62.33%)	45 (54.22%)
Validating survival distribution of intervention and comparators	20 (8.73%)	9 (6.16%)	11 (13.25%)
Choice of comparators	14 (6.11%)	9 (6.16%)	5 (6.02%)
Validating survival distribution of comparators	13 (5.68%)	8 (5.48%)	5 (6.02%)
Treatment sequence	7 (3.06%)	3 (2.05%)	4 (4.82%)
Characteristics of population	7 (3.06%)	5 (3.42%)	2 (2.41%)
Validating survival distribution of intervention	4 (1.75%)	3 (2.05%)	1 (1.20%)
Treatment sequence & validating survival distribution of intervention and comparators	4 (1.75%)	3 (2.05%)	1 (1.20%)
Choice of comparator & validating survival distribution of intervention and comparators	3 (1.31%)	2 (1.37%)	1 (1.20%)
Choice of comparator & validating survival distribution of comparators & time-to-discontinuation of comparators	2 (0.87%)	2 (1.37%)	.
Treatment sequence & time-to-discontinuation of intervention and comparators	2 (0.87%)	2 (1.37%)	.
Treatment sequence & validating survival distribution of comparators	2 (0.87%)	.	2 (2.41%)
Other*	15 (6.55%)	9 (6.16%)	6 (7.25%)
Total	229 (100%)	146 (100%)	83 (100%)

* The patterns are not listed because each represent a particular pattern of use of RWD only observed once.

c. Pattern of use of RWD (Parametric use)

Patterns	All appraisals	Non-targeted	Targeted
No use of RWD	55 (24.02%)	37 (25.34%)	18 (21.69%)
Estimating end-of-life resource use	23 (10.04%)	7 (4.79%)	16 (19.28%)
Estimating OS of intervention and comparators	17 (7.42%)	15 (10.27%)	2 (2.41%)
Estimating end-of-life resource use & resource use of health state costs	14 (6.11%)	7 (4.79%)	7 (8.43%)
Estimating resource use of health state costs	13 (5.68%)	8 (5.48%)	5 (6.02%)
Estimating OS & PFS of intervention and comparators	10 (4.37%)	4 (2.74%)	6 (7.23%)
Estimating end-of-life resource use & dose adjustment of intervention and comparators	9 (3.93%)	7 (4.79%)	2 (2.41%)
Estimating OS of intervention and comparators, end-of-life resource use & resource use of health state costs	6 (2.62%)	4 (2.74%)	2 (2.41%)
Estimating OS & PFS of intervention and comparators & resource use of health state costs	6 (2.62%)	4 (2.74%)	2 (2.41%)
Estimating volume of treatment for intervention and comparators	4 (1.75%)	3 (2.05%)	1 (1.20%)
Estimating OS & PFS of comparators	3 (1.31%)	2 (1.37%)	1 (1.20%)
Estimating OS of intervention and comparators & end-of-life resource use	3 (1.31%)	2 (1.37%)	1 (1.20%)
Estimating OS of intervention and comparators & resource use of health state costs	3 (1.31%)	3 (2.05%)	.
Estimating OS & PFS of intervention and comparators, end-of-life resource use & resource use of health state costs	3 (1.31%)	2 (1.37%)	1 (1.20%)
Other*	60 (26.20%)	41 (28.08%)	19 (22.89%)
Total	229 (100%)	146 (100%)	83 (100%)

* The patterns are not listed because each represent a particular pattern of use of RWD only observed once or twice.

OS: Overall survival, PFS: Progression-free survival

A.4.2 Patterns of use of real-world data in the original appraisals and Cancer Drugs Fund review appraisals

	TA number	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae
Original	TA416							ν								ν	ν									ν						
CDF	TA653							ν								ν	ν									ν						
Original	TA446		ν								ν	ν				ν	ν															
CDF	TA524		ν								ν	ν				ν	ν															
Original	TA447																								ν	ν						
CDF	TA531																								ν	ν						
Original	TA472										ν					ν																
CDF	TA629										ν					ν																
Original	TA483		ν				ν	ν																		ν		ν	ν		ν	ν
CDF	TA655		ν				ν	ν	ν	ν																ν		ν	ν		ν	ν
Original	TA484						ν	ν																								ν
CDF	TA713						ν	ν																								ν
Original	TA487			ν							ν					ν	ν												ν			ν
CDF	TA796	ν		ν					ν		ν					ν	ν												ν			ν
Original	TA490																									ν			ν			ν
CDF	TA736						ν																			ν			ν			ν
Original	TA491		ν													ν	ν				ν					ν						
CDF	TA795	ν	ν								ν	ν				ν	ν				ν					ν						
Original	TA492																															
CDF	TA739						ν																									
Original	TA510							ν								ν	ν											ν			ν	
CDF	TA783						ν	ν	ν							ν	ν											ν			ν	
Original	TA517			ν				ν		ν	ν					ν	ν			ν										ν	ν	
CDF	TA691			ν				ν	ν	ν	ν					ν	ν			ν										ν	ν	
Original	TA519						ν																			ν						
CDF	TA692						ν																			ν						
Original	TA522																									ν						
CDF	TA674																									ν						
Original	TA528																															
CDF	TA784																															
Original	TA553										ν					ν					ν					ν		ν			ν	
CDF	TA683										ν					ν					ν					ν		ν			ν	
Original	TA557																								ν	ν						
CDF	TA683		ν				ν																		ν	ν			ν		ν	
Original	TA558										ν	ν				ν	ν										ν	ν		ν	ν	ν
CDF	TA684										ν	ν				ν	ν										ν	ν		ν	ν	ν
Original	TA578						ν	ν																		ν						
CDF	TA798						ν	ν																		ν						
Original	TA579																															
CDF	TA725						ν		ν	ν																						
Original	TA581																								ν	ν		ν	ν			ν
CDF	TA780																								ν	ν		ν	ν			ν
Original	TA592						ν	ν								ν	ν															
CDF	TA802						ν	ν								ν	ν															
Original	TA593																															
CDF	TA687																															
Original	TA600			ν							ν	ν				ν	ν								ν	ν						
CDF	TA770			ν																					ν	ν						

<Labels>			
a	Characteristic of population	q	Response rate (Comparators)
b	Treatment sequence	r	Time-to-progress (Comparators)
c	Choice of comparators	s	Adverse event (comparators)
d	Health state	t	Transition probability
e	Model cycle	u	Health state utility (generic measure)
f	Survival distribution (intervention)	v	Health state utility (cancer-specific)
g	Survival distribution (comparators)	w	Disutility
h	Time-to-discontinuation (intervention)	x	Resource use of health state cost
i	Time-to-discontinuation (comparators)	y	Resource use of end-of-life treatment
j	Overall survival (intervention)	z	Resource use of adverse event cost (intervention)
k	Progression-free-survival (intervention)	aa	Volume of treatment (intervention)
l	Response rate (intervention)	ab	Dose adjustment (intervention)
m	Time-to-progress (intervention)	ac	Resource use of adverse event cost (comparators)
n	Adverse event (intervention)	ad	Volume of treatment (comparators)
o	Overall survival (comparators)	ae	Dose adjustment (comparators)
p	Progression-free-survival (comparators)		

Appendix 5. Regression analysis

A.5.1 Univariate regression (Outcome variable: non-parametric, parametric use and use regardless of type)

a. Use of RWD regardless of type

*p<0.05, **p<0.01, ***p<0.001, [a] Reference group

	Use of RWD regardless of type						
Covariate	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*
Time							
	1.009 (0.999, 1.020)						
Direct treatment comparison *(AD)*							
Not available[a]							
Some available		0.213** (0.067, 0.672)					
All available		0.239*(0.076, 0.747)					
Incidence rate *(IR)*							
			1.000*(1.000, 1.000)				
Maturity of survival data *(MS)*							
Extremely immature[a]							
Immature				1.622 (0.715, 3.682)			
Mature				2.632*(1.036, 6.686)			
External Validity *(EV)*							
Low risk[a]							
Moderate					2.358*(1.107, 5.021)		
Questionable					1.816(0.554, 5.949)		
Internal Validity (*InV*)							
High quality[a]							
Low						0.975 (0.435, 2.189)	
Moderate						0.889 (0.252, 3.139)	
Questionable						3.185 (0.845, 12.008)	
Previously recommended *(PR)*							
No[a]							
Yes							2.382* (1.093, 5.192)
Constant	2.385 (0.999, 1.020)	16.75*** (6.108, 45.934)	5.348*** (3.563, 8.026)	3.474*** (2.085, 5.786)	3.167*** (1.564, 5.379)	4.5*** (2.407, 8.411)	3.778*** (2.472, 5.774)
Observations	229	229	229	229	229	229	229

b. Non-parametric use of RWD

*p<0.05, **p<0.01, ***p<0.001, [a] Reference group

	Parametric use of RWD						
Covariate	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*
Time							
	1.008 (0.999, 1.017)						
Direct treatment comparison *(AD)*							
Not available[a]							
Some available		0.567 (0.290, 1.109)					
All available		0.725 (0.384, 1.371)					
Incidence rate *(IR)*							
			1.000 (1.000, 1.000)				
Maturity of survival data *(MS)*							
Extremely immature[a]							
Immature				0.781 (0.409, 1.496)			
Mature				1.378 (0.729, 2.605)			
External Validity *(EV)*							
Low risk[a]							
Moderate					1.074 (0.601, 1.921)		
Questionable					0.882 (0.356, 2.187)		
Internal Validity (*InV*)							
High quality[a]							
Low						0.582 (0.307, 1.104)	
Moderate						0.608 (0.215, 1.717)	
Questionable						1.035 (0.487, 2.200)	
Previously recommended *(PR)*							
No[a]							
Yes							1.486 (0.873, 2.531)
Constant	0.346** (1.552, 0.769)	0.919 (0.577, 1.464)	0.637** (0.468, 0.866)	0.667 (0.432, 1.029)	0.667 (0.420, 1.058)	0.886 (0.546, 1.436)	0.573** (0.400, 0.820)
Observations	229	229	229	229	229	229	229

c. Parametric use of RWD

*p<0.05, **p<0.01, ***p<0.001, [a] Reference group

	Parametric use of RWD						
Covariate	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*	*OR (CI)*
Time							
	1.006 (0.996, 1.015)						
Direct treatment comparison *(AD)*							
Not available[a]							
Some available		0.294**(0.121, 0.716)					
All available		0.305**(0.128, 0.728)					
Incidence rate *(IR)*							
			1.000 (1.000, 1.000)				
Maturity of survival data *(MS)*							
Extremely immature[a]							
Immature				0.933 (0.455, 1.913)			
Mature				1.225 (0.577, 2.602)			
External Validity *(EV)*							
Low risk[a]							
Moderate					2.165*(1.135, 4.131)		
Questionable					3.051 (0.953, 9.766)		
Internal Validity (*InV*)							
High quality[a]							
Low						1.44 (0.726, 2.856)	
Moderate						2.00 (0.597, 6.702)	
Questionable						5.25**(1.669, 16.517)	
Previously recommended *(PR)*							
No[a]							
Yes							1.832 (0.969, 3.464)
Constant	1.959	7.875***	3.551***	3.048***	1.885**	2**	2.486***
	(0.848, 4.524)	(3.774, 16.434)	(2.495, 5.055)	(1.862, 4.989)	(1.171, 3.032)	(1.199, 3.337)	(1.698, 3.642)
Observations	229	229	229	229	229	229	229

A.5.2 Multivariate regression including InV

(Outcome variable: non-parametric, parametric use and use regardless of type)

Covariate	Model A Any use of RWD β (SE(β))	Model A Any use of RWD Odds ratio (95% CI)	Model B Non-parametric use β (SE(β))	Model B Non-parametric use Odds ratio (95% CI)	Model C Parametric use β (SE(β))	Model C Parametric use Odds ratio (95% CI)
Time						
	0.008 (0.007)	1.008 (0.995,1.021)	0.006 (0.005)	1.006 (0.996,1.016)	0.002 (0.006)	1.002 (0.991,1.013)
Direct treatment comparison *(AD)*						
Not available[a]						
Some available	-1.873* (0.874)	0.154* (0.028, 0.852)	-0.680 (0.460)	0.506 (0.206, 1.248)	-0.828 (0.596)	0.437 (0.136, 1.405)
All available	-1.509 (0.839)	0.221 (0.043, 1.144)	-0.296 (0.426)	0.744 (0.323, 1.713)	-0.905 (0.575)	0.405 (0.131, 1.248)
Incidence rate *(IR)*						
	-0.000 (0.000)	1.000 (1.000, 1.000)	0.000 (0.000)	1.000 (1.000, 1.000)	-0.000 (0.000)	1.000 (1.000, 1.000)
Maturity of survival data *(MS)*						
Extremely immature[a]						
Immature	1.099* (0.493)	3.014* (1.155, 7.867)	-0.040 (0.360)	0.961 (0.475, 1.947)	0.345 (0.419)	1.412 (0.621, 3.211)
Mature	1.652** (0.548)	5.348** (1.830, 15.632)	0.546 (0.353)	1.726 (0.864, 3.448)	0.652 (0.431)	1.919 (0.815, 4.446)
External Validity *(EV)*						
Low risk[a]						
Moderate	1.044* (0.434)	2.840* (1.212, 6.650)	0.120 (0.317)	1.127 (0.605, 2.100)	0.761* (0.358)	2.140* (1.061, 4.317)
Questionable	0.592 (0.690)	1.808 (0.467, 6.992)	-0.324 (0.503)	0.723 (0.270, 1.937)	0.990 (0.649)	2.691 (0.754, 9.601)
Internal Validity (InV)						
High quality[a]						
Low	-0.228 (0.464)	0.796 (0.321, 1.977)	-0.675 (0.345)	0.509 (0.259, 1.002)	0.336 (0.375)	1.400 (0.671, 2.919)
Moderate	-0.581 (0.746)	0.559 (0.130, 2.414)	-0.632 (0.563)	0.531 (0.176, 1.603)	0.414 (0.663)	1.513 (0.413, 5.549)
Questionable	0.112 (1.036)	1.118 (0.147, 8.512)	-0.201 (0.540)	0.818 (0.284, 2.358)	1.031 (0.778)	2.804 (0.611, 12.873)
Previously recommended *(PR)*						
No[a]						
Yes	1.021* (0.461)	2.673* (1.091, 6.548)	0.276 (0.308)	1.318 (0.721, 2.412)	0.813* (0.380)	2.255* (1.071, 4.749)
Constant	0.786	2.194	-0.599	0.549	0.813	1.368
Observations	229	229	229	229	229	229

*p<0.05, **p<0.01

[a] Reference group

A.5.3 Multivariate regression excluding $Time$, InV and IR

(Outcome variable: non-parametric, parametric use and use regardless of type)

Covariate	Model A Any use of RWD β (SE(β))	Model A Any use of RWD Odds ratio (95% CI)	Model B Non-parametric use β (SE(β))	Model B Non-parametric use Odds ratio (95% CI)	Model C Parametric use β (SE(β))	Model C Parametric use Odds ratio (95% CI)
Direct treatment comparison *(AD)*						
Not available[a]						
Some available	-2.105** (0.652)	0.122** (0.034, 0.437)	-0.706 (0.365)	0.494 (0.242, 1.009)	-1.312** (0.485)	0.269** (0.104, 0.696)
All available	-1.777** (0.613)	0.169** (0.051, 0.562)	-0.447 (0.337)	0.639 (0.330, 1.237)	-1.300** (0.461)	0.273** (0.110, 0.672)
Maturity of survival data *(MS)*						
Extremely immature[a]						
Immature	0.985* (0.474)	2.678* (1.058, 6.782)	-0.134 (0.346)	0.875 (0.444, 1.724)	0.240 (0.402)	1.272 (0.579, 2.794)
Mature	1.571** (0.533)	4.810** (1.693, 13.663)	0.478 (0.342)	1.613 (0.825, 3.151)	0.579 (0.420)	1.783 (0.784, 4.058)
External Validity *(EV)*						
Low risk[a]						
Moderate	0.934* (0.418)	2.546* (1.123, 5.771)	0.017 (0.305)	1.017 (0.560, 1.849)	0.802* (0.346)	2.230* (1.133, 4.392)
Questionable	0.609 (0.676)	1.839 (0.4889, 6.922)	-0.237 (0.485)	0.789 (0.305, 2.041)	1.062 (0.628)	2.891 (0.844, 9.904)
Previously recommended *(PR)*						
No[a]						
Yes	1.146** (0.430)	3.145** (1.353, 7.310)	0.426 (0.278)	1.532 (0.889, 2.639)	0.764* (0.343)	2.147* (1.095, 4.208)
Constant	1.496*	4.462*	-0.276	0.758	1.055*	2.873*
Observations	229	229	229	229	229	229

*p<0.05, **p<0.01

[a] Reference group

A.5.4 Multivariate binary logistic regression (Outcome variable: use of RWD in each component)

	Non-parametric use									
	Characteristic of population		Treatment sequence		Choice of comparators		Health state		Model cycle	
VARIABLES	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)
Time										
	-0.00835	0.992	-0.00382	0.996	0.00262	1.003	-0.0124	0.988	-0.00324	0.997
	(0.0122)	(0.968 - 1.016)	(0.00818)	(0.980 - 1.012)	(0.00802)	(0.987 - 1.018)	(0.0177)	(0.954 - 1.023)	(0.0399)	(0.922 - 1.078)
Direct treatment comparison *(AD)*										
Not available[a]										
Some available	-0.604	0.547	-0.770	0.463	-0.830	0.436	-	-	-	-
	(1.081)	(0.066 - 4.547)	(0.681)	(0.122 - 1.760)	(0.536)	(0.153 - 1.245)				
All available	-0.502	0.605	-0.107	0.898	-1.045*	0.352*	-0.280	0.756	-	-
	(0.925)	(0.099 - 3.706)	(0.595)	(0.280 - 2.881)	(0.538)	(0.123 - 1.010)	(1.118)	(0.085 - 6.761)		
Incidence rate *(IR)*										
	-0.000335	1.000	5.91e-05	1.000	8.38e-06	1.000	5.77e-05	1.000	-0.00720	0.993
	(0.000380)	(0.999 - 1.000)	(6.03e-05)	(1.000 - 1.000)	(5.07e-05)	(1.000 - 1.000)	(9.18e-05)	(1.000 - 1.000)	(0.0136)	(0.967 - 1.020)
Maturity of survival data *(MS)*										
Extremely immature[a]										
Immature	-0.469	0.626	0.680	1.974	0.432	1.541	-1.432	0.239	-	-
	(1.270)	(0.052 - 7.544)	(0.775)	(0.432 - 9.016)	(0.532)	(0.543 - 4.373)	(1.450)	(0.014 - 4.101)		
Mature	1.405	4.077	1.721**	5.591**	0.191	1.211	-	-	-	-
	(0.969)	(0.610 - 27.23)	(0.699)	(1.420 - 22.01)	(0.565)	(0.400 - 3.663)				
External Validity *(EV)*										
Low risk[a]										
Moderate	0.394	1.483	0.495	1.641	-0.576	0.562	-1.898	0.150	-	-
	(1.145)	(0.157 - 14.00)	(0.574)	(0.533 - 5.055)	(0.472)	(0.223 - 1.418)	(1.225)	(0.014 - 1.653)		
Questionable	-	-	-0.529	0.589	-0.127	0.881	-	-	-	-
			(1.153)	(0.062 - 5.645)	(0.668)	(0.238 - 3.265)				
Previously recommended *(PR)*										
	-0.626	0.535	-0.644	0.525	-0.107	0.899	-0.128	0.880	-	-
	(0.896)	(0.092 - 3.096)	(0.591)	(0.165 - 1.672)	(0.470)	(0.358 - 2.256)	(1.166)	(0.090 - 8.645)		
Constant	-2.172	0.114	-2.887**	0.0558**	-1.586	0.205	-0.676	0.508	5.012	150.3
	(1.575)	(0.005 - 2.495)	(1.141)	(0.006 - 0.522)	(0.926)	(0.033 - 1.257)	(2.081)	(0.009 - 30.05)	(13.00)	(1.30e-09 - 1.732e+13)
Observations	154	154	229	229	229	229	94	94	7	7

Standard errors in parentheses

*** p<0.01, ** p<0.05

(Continued Table A.5.4)

	Non-parametric use							
	Validating survival distribution (intervention)		Validating survival distribution (comparators)		Time-to-discontinuation (intervention)		Time-to-discontinuation (comparators)	
VARIABLES	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)
Time								
	0.0152**	1.015**	0.00899	1.009	0.0259	1.026	0.00787	1.008
	(0.00776)	(1.000 - 1.031)	(0.00634)	(0.997 - 1.022)	(0.0210)	(0.985 - 1.069)	(0.0147)	(0.979 - 1.037)
Direct treatment comparison *(AD)*								
Not available[a]								
Some available	0.153	1.165	-0.155	0.857	0.0774	1.081	-1.923	0.146
	(0.552)	(0.395 - 3.439)	(0.439)	(0.362 - 2.026)	(1.534)	(0.0535 - 21.84)	(1.139)	(0.0157 - 1.363)
All available	0.792	2.208	-0.0646	0.937	0.720	2.055	-1.375	0.253
	(0.504)	(0.823 - 5.925)	(0.405)	(0.424 - 2.075)	(1.384)	(0.136 - 30.97)	(0.855)	(0.0473 - 1.352)
Incidence rate *(IR)*								
	6.86e-05	1.000	-7.56e-06	1.000	0.000131	1.000	-7.65e-05	1.000
	(4.56e-05)	(1.000 - 1.000)	(4.69e-05)	(1.000 - 1.000)	(9.63e-05)	(1.000 - 1.000)	(0.000134)	(1.000 - 1.000)
Maturity of survival data *(MS)*								
Extremely immature[a]								
Immature	0.412	1.510	-0.0885	0.915	0.202	1.224	-0.571	0.565
	(0.531)	(0.534 - 4.271)	(0.422)	(0.400 - 2.092)	(1.529)	(0.0611 - 24.50)	(0.919)	(0.0932 - 3.425)
Mature	0.874*	2.396	-0.0961	0.908	1.096	2.991	-0.429	0.651
	(0.477)	(0.941 - 6.106)	(0.414)	(0.403 - 2.046)	(1.290)	(0.238 - 37.53)	(0.904)	(0.111 - 3.829)
External Validity *(EV)*								
Low risk[a]								
Moderate	0.0455	1.047	0.299	1.348	-0.192	0.826	-0.941	0.390
	(0.439)	(0.443 - 2.473)	(0.380)	(0.640 - 2.837)	(1.104)	(0.0949 - 7.179)	(0.719)	(0.0952 - 1.598)
Questionable	0.193	1.212	0.0106	1.011				
	(0.682)	(0.319 - 4.614)	(0.600)	(0.312 - 3.278)				
Previously recommended *(PR)*								
	0.837	2.310	0.674	1.961	-1.966	0.140	-0.130	0.878
	(0.440)	(0.976 - 5.467)	(0.363)	(0.964 - 3.992)	(1.458)	(0.00804 - 2.438)	(0.749)	(0.202 - 3.816)
Constant	-4.591***	0.0101***	-2.495***	0.0825***	-6.758**	0.00116**	-1.931	0.145
	(0.975)	(0.002 - 0.069)	(0.751)	(0.0189 - 0.360)	(2.867)	(4.21e-06 - 0.320)	(1.612)	(0.006 - 3.419)
Observations	229	229	229	229	202	202	202	202

Standard errors in parentheses
*** p<0.01, ** p<0.05

(continued Table A.5.4)

	Parametric use											
	Overall survival (intervention)		Progression-free survival (intervention)		Overall survival (comparators)		Progression-free survival (comparators)		Adverse events (comparators)		Transition probability	
VARIABLES	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)
Time												
	0.00616	1.006	0.0215**	1.022**	0.0106	1.011	0.0115	1.012	0.00461	1.005	0.0327**	1.033**
	(0.00569)	(0.995 - 1.017)	(0.00902)	(1.004 - 1.040)	(0.00614)	(0.999 - 1.023)	(0.00756)	(0.997 - 1.027)	(0.0434)	(0.923 - 1.094)	(0.0165)	(1.000 - 1.067)
Direct treatment comparison *(AD)*												
Not available[a]												
Some available	-1.031**	0.357**	-1.021	0.360	-1.615***	0.199***	-2.049***	0.129***	-	-	-1.840	0.159
	(0.432)	(0.153 - 0.832)	(0.655)	(0.100 - 1.300)	(0.440)	(0.084 - 0.471)	(0.557)	(0.043 - 0.384)			(1.200)	(0.015 - 1.670)
All available	-0.544	0.580	-0.636	0.529	-1.253***	0.286***	-1.775***	0.170***	-	-	0.107	1.112
	(0.364)	(0.284 - 1.183)	(0.496)	(0.200 - 1.401)	(0.381)	(0.135 - 0.602)	(0.447)	(0.071 - 0.407)			(0.675)	(0.296 - 4.176)
Incidence rate *(IR)*												
	-0.000	1.000	-0.000	1.000	-0.000	1.000	-0.000156	1.000	-0.0393	0.961	-0.000142	1.000
	(7.77e-05)	(1.000 - 1.000)	(0.000231)	(0.999 - 1.000)	(7.67e-05)	(1.000 - 1.000)	(0.000100)	(1.000 - 1.000)	(0.0259)	(0.914 - 1.012)	(0.000148)	(1.000 - 1.000)
Maturity of survival data *(MS)*												
Extremely immature[a]												
Immature	-0.900**	0.406**	-0.185	0.831	-1.050***	0.350***	-1.059**	0.347**	-1.866	0.155	-0.262	0.770
	(0.385)	(0.191 - 0.864)	(0.540)	(0.289 - 2.396)	(0.398)	(0.161 - 0.763)	(0.475)	(0.137 - 0.879)	(1.926)	(0.004 - 6.752)	(0.775)	(0.169 - 3.515)
Mature	-0.990**	0.372**	0.0151	1.015	-1.416***	0.243***	-1.220**	0.295**	-	-	-1.981	0.138
	(0.389)	(0.173 - 0.797)	(0.548)	(0.347 - 2.972)	(0.411)	(0.109 - 0.542)	(0.506)	(0.110 - 0.796)			(1.101)	(0.016 - 1.194)
External Validity *(EV)*												
Low risk[a]												
Moderate	0.162	1.176	0.380	1.462	0.689	1.993	0.182	1.200	-3.657	0.0258	-0.585	0.557
	(0.355)	(0.586 - 2.358)	(0.501)	(0.547 - 3.905)	(0.379)	(0.949 - 4.185)	(0.441)	(0.505 - 2.849)	(2.690)	(0.000 - 5.029)	(0.704)	(0.140 - 2.212)
Questionable	0.162	1.176	-0.249	0.780	0.191	1.211	-0.740	0.477	-5.581	0.00377	0.141	1.152
	(0.522)	(0.423 - 3.275)	(0.779)	(0.169 - 3.589)	(0.563)	(0.402 - 3.650)	(0.662)	(0.130 - 1.746)	(4.269)	(8.76e-07 - 16.20)	(0.981)	(0.169 - 7.871)
Previously recommended *(PR)*												
	0.0153	1.015	-0.0818	0.921	-0.485	0.616	-0.316	0.729	3.220	25.04	2.145**	8.544**
	(0.336)	(0.525 - 1.962)	(0.458)	(0.376 - 2.260)	(0.359)	(0.305 - 1.244)	(0.412)	(0.325 - 1.634)	(2.929)	(0.080 - 7,798)	(0.848)	(1.621 - 45.03)
Constant	-0.184	0.832	-3.062***	0.0468***	0.100	1.105	-0.352	0.703	3.392	29.74	-6.211***	0.00201***
	(0.667)	(0.225 - 3.077)	(1.039)	(0.006 - 0.359)	(0.710)	(0.275 - 4.442)	(0.853)	(0.132 - 3.742)	(4.540)	(0.004 - 217,738)	(1.920)	(4.66e-05 - 0.0864)
Observations	229	229	229	229	229	229	229	229	57	57	229	229

Standard errors in parentheses, *** p<0.01, ** p<0.05

(Continued Table A.5.4)

	Parametric use									
	Health utility of health state (generic)		Disutility of adverse event		Resource use for health state cost		Resource use for end-of-life care		Resource use for managing adverse event cost (intervention)	
VARIABLES	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)
Time										
	0.000777	1.001	-21.35	5.36e-10	-0.000404	1.000	-0.00135	0.999	-0.0239	0.976
	(0.0131)	(0.975 - 1.027)	(0)	(5.36e-10 - 5.36e-10)	(0.00541)	(0.989 - 1.010)	(0.00514)	(0.989 - 1.009)	(0.0189)	(0.941 - 1.013)
Direct treatment comparison *(AD)*										
Not available[a]										
Some available	-0.345	0.708			0.408	1.504	0.265	1.303	-0.0704	0.932
	(0.813)	(0.144 - 3.485)			(0.426)	(0.653 - 3.465)	(0.378)	(0.621 - 2.737)	(1.100)	(0.108 - 8.055)
All available	-	-	-	-	0.213	1.237	-0.0515	0.950	-	-
					(0.399)	(0.567 - 2.702)	(0.367)	(0.463 - 1.949)		
Incidence rate *(IR)*										
	-8.13e-05	1.000	-2.209	0.110	-0.000138*	1.000*	2.22e-05	1.000	1.58e-05	1.000
	(0.000208)	(1.000 - 1.000)	(0)	(0.110 - 0.110)	(8.25e-05)	(1.000 - 1.000)	(4.10e-05)	(1.000 - 1.000)	(0.000139)	(1.000 - 1.000)
Maturity of survival data *(MS)*										
Extremely immature[a]										
Immature	1.110	3.034			0.654	1.924	0.629*	1.875*	-0.941	0.390
	(1.223)	(0.276 - 33.36)			(0.408)	(0.865 - 4.280)	(0.374)	(0.901 - 3.903)	(1.347)	(0.028 - 5.469)
Mature	0.996	2.707			0.949**	2.583**	0.906**	2.475**	-0.651	0.522
	(1.197)	(0.259 - 28.29)			(0.403)	(1.172 - 5.695)	(0.369)	(1.201 - 5.099)	(1.368)	(0.036 - 7.622)
External Validity *(EV)*										
Low risk[a]										
Moderate	-0.175	0.839			0.540	1.715	0.495	1.641	0.483	1.621
	(0.803)	(0.174 - 4.052)			(0.361)	(0.846 - 3.481)	(0.329)	(0.862 - 3.124)	(1.268)	(0.135 - 19.46)
Questionable	-	-	-	-	0.740	2.095	0.302	1.352	-	-
					(0.543)	(0.723 - 6.071)	(0.528)	(0.480 - 3.808)		
Previously recommended *(PR)*										
	0.00405	1.004			-0.204	0.815	0.823**	2.276**	1.986	7.290
	(0.883)	(0.178 - 5.663)			(0.347)	(0.413 - 1.608)	(0.321)	(1.213 - 4.272)	(1.316)	(0.553 - 96.10)
Constant	-3.454**	0.0316**	2,383		-1.819***	0.162***	-1.805***	0.164***	-2.475	0.0841
	(1.639)	(0.001 - 0.785)	(0)	(-)	(0.687)	(0.042 - 0.623)	(0.636)	(0.047 - 0.572)	(2.240)	(0.001 - 6.791)
Observations	145	145	5	5	229	229	229	229	126	126

Standard errors in parentheses
*** p<0.01, ** p<0.05

(Continued Table A.5.4)

	Parametric use									
	Volume of treatment (intervention)		Dose adjustment (intervention)		Resource use for managing adverse event cost (comparators)		Volume of treatment (comparators)		Dose adjustment (comparators)	
VARIABLES	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)	β (SE)	OR (CI)
Time										
	0.00752	1.008	-0.0154*	0.985*	-0.0231*	0.977*	0.00186	1.002	-0.00809	0.992
	(0.00920)	(0.990 - 1.026)	(0.00872)	(0.968 - 1.002)	(0.0137)	(0.951 - 1.004)	(0.00815)	(0.986 - 1.018)	(0.00806)	(0.976 - 1.008)
Direct treatment comparison *(AD)*										
Not available[a]										
Some available	0.153	1.165	0.503	1.654	-1.402	0.246	-0.838	0.432	0.115	1.122
	(0.618)	(0.347 - 3.916)	(0.786)	(0.354 - 7.726)	(0.941)	(0.0389 - 1.557)	(0.567)	(0.142 - 1.315)	(0.612)	(0.338 - 3.721)
All available	-0.408	0.665	0.534	1.705	-2.360*	0.0944*	-1.160**	0.313**	-0.531	0.588
	(0.612)	(0.200 - 2.207)	(0.737)	(0.403 - 7.223)	(1.209)	(0.00883 - 1.010)	(0.547)	(0.107 - 0.915)	(0.631)	(0.171 - 2.025)
Incidence rate *(IR)*										
	-8.65e-05	1.000	9.24e-05	1.000	-4.17e-05	1.000	-0.000107	1.000	7.49e-05	1.000
	(0.000116)	(1.000 - 1.000)	(8.93e-05)	(1.000 - 1.000)	(0.000141)	(1.000 - 1.000)	(0.000111)	(1.000 - 1.000)	(6.81e-05)	(1.000 - 1.000)
Maturity of survival data *(MS)*										
Extremely immature[a]										
Immature	-0.230	0.794	-0.379	0.685	-0.940	0.391	-0.359	0.698	-0.00180	0.998
	(0.594)	(0.248 - 2.546)	(0.957)	(0.105 - 4.467)	(0.959)	(0.0597 - 2.557)	(0.543)	(0.241 - 2.024)	(0.755)	(0.227 - 4.380)
Mature	-0.360	0.697	1.803***	6.070***	-0.627	0.534	-0.360	0.697	1.549**	4.707**
	(0.619)	(0.207 - 2.347)	(0.694)	(1.559 - 23.64)	(0.968)	(0.0801 - 3.561)	(0.566)	(0.230 - 2.113)	(0.633)	(1.362 - 16.27)
External Validity *(EV)*										
Low risk[a]										
Moderate	1.273*	3.570*	2.402**	11.05**	-0.958	0.384	0.576	1.779	1.398**	4.048**
	(0.658)	(0.982 - 12.98)	(1.062)	(1.377 - 88.62)	(0.789)	(0.082 - 1.802)	(0.513)	(0.650 - 4.863)	(0.675)	(1.077 - 15.21)
Questionable	-0.0971	0.907	2.107	8.225	-	-	-1.210	0.298	0.972	2.644
	(1.199)	(0.087 - 9.510)	(1.301)	(0.642 - 105.4)			(1.131)	(0.033 - 2.736)	(0.986)	(0.383 - 18.26)
Previously recommended *(PR)*										
	0.589	1.802	-0.0555	0.946	0.0537	1.055	0.399	1.490	-0.452	0.636
	(0.531)	(0.637 - 5.099)	(0.617)	(0.282 - 3.169)	(0.813)	(0.215 - 5.191)	(0.471)	(0.592 - 3.752)	(0.568)	(0.209 - 1.936)
Constant	-3.759***	0.0233***	-4.473***	0.0114***	0.681	1.977	-1.728*	0.178*	-3.128***	0.0438***
	(1.136)	(0.003 - 0.216)	(1.471)	(0.001- 0.204)	(1.499)	(0.105 - 37.29)	(0.948)	(0.028 - 1.139)	(1.133)	(0.005 - 0.404)
Observations	229	229	229	229	202	202	229	229	229	229

Standard errors in parentheses, *** p<0.01, ** p<0.05

A.5.5 Results of Brant test

	χ^2	P > χ^2	df
All	8.10	0.524	9
Time			
	0.06	0.808	1
Incidence rate (IR)			
	0.02	0.883	1
Availability of direct treatment comparison (AD)			
Some available	0.59	0.444	1
All available	0.99	0.320	1
External validity (EV)			
Moderate	1.02	0.312	1
Questionable	0.20	0.652	1
Previous recommendation (PR)			
	3.90	0.048*	1
Maturity of survival data (MS)			
Immature	0.63	0.429	1
Mature	0.05	0.829	1

* A significant test statistic provides evidence that the parallel regression assumption has been violated.

A.5.6 Multivariate ordinal regression including InV
(Outcome variable: intensity of use of real-world data)

Covariate	Univariate model		Full model	
	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Model 1 (Y > 1 vs Y≤ 1)				
Time				
	0.008 (0.004)	1.008 (1.000, 1.016)	0.005 (0.005)	1.005 (1.000, 1.016)
Availability of head-to-head comparison				
Not available[a]				
Some available			-2.075*** (0.463)	0.126*** (0.051, 0.311)
All available			-1.775*** (0.429)	0.169*** (0.073, 0.393)
Incidence rate				
			-0.000 (0.000)	1.000 (1.000, 1.000)
Previous recommendation states				
			-0.329 (0.324)	0.720 (0.382,1.358)
Maturity of survival data				
Extremely immature[a]				
Immature			-0.679* (0.346)	0.507* (0.258, 0.999)
Mature			-1.007** (0.353)	0.365** (0.183, 0.729)
External validity				
Low risk[a]				
Moderate risk			0.382 (0.313)	1.465 (0.793, 2.707)
High risk			0.079 (0.484)	1.082 (0.419, 2.794)
Internal validity				
High quality[a]				
Low risk			-0.350 (0.346)	0.704 (0.357, 1.388)
Moderate risk			-0.163 (0.541)	0.849 (0.294, 2.453)
Questionable risk			-0.812 (0.540)	0.444 (0.154, 1.281)
Model 2 (Y > 2 vs Y≤ 2)				
Time				
	0.008 (0.004)	1.008 (1.000, 1.016)	0.005 (0.005)	1.005 (0.995, 1.015)
Availability of head-to-head comparison				
Not available[a]				
Some available			-2.075*** (0.463)	0.126*** (0.051, 0.311)
All available			-1.775***	0.169***

			(0.429)	(0.073, 0.393)
Incidence rate				
			-0.000	1.000
			(0.000)	(1.000, 1.000)
Previous recommendation states				
			0.680	1.973
			(0.496)	(0.747, 5.213)
Maturity of survival data				
Extremely immature[a]				
Immature			-0.679*	0.507*
			(0.346)	(0.258, 0.999)
Mature			-1.007**	0.365**
			(0.353)	(0.183, 0.729)
External validity				
Low risk[a]				
Moderate			0.382	1.465
			(0.313)	(0.793, 2.707)
Questionable			0.079	1.082
			(0.484)	(0.419, 2.794)
Internal validity				
High quality				
Low risk			-0.350	0.704
			(0.346)	(0.357, 1.388)
Moderate risk			-0.163	0.849
			(0.541)	(0.294, 2.453)
Questionable risk			-0.812	0.444
			(0.540)	(0.154, 1.281)
α_1	-0.736	0.479	1.761	5.815
	(0.383)	(0.226,1.015)	(0.724)	(1.406, 24.051)
α_2	-2.919***	0.054***	-1.338	0.262
	(0.436)	(0.023, 0.127)	(0.776)	(0.057, 1.201)
Observations	229		229	
LR R^2	0.008		0.133	
Log likelihood	-211.865		-183.949	
LR χ^2	3.47		59.30	

**p<0.01, ***p<0.001

[a] Reference group

Note: Standard errors are shown in parentheses.

A.5.7 Ordinal logistic regression using different time unit
(Outcome variable: maturity of survival data)

Covariate	Model A (Time unit: Monthly)		Model B (Time unit: Quarterly)		Model C (Time unit: Annually)	
	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)	β (SE(β))	Odds ratio (95% CI)
Time						
	-0.008 (0.006)	0.992 (0.980, 1.004)	-0.023 (0.019)	0.978 (0.942, 1.014)	-0.131 (0.077)	0.877 (0.758, 1.015)
Post 2016 CDF						
	-0.254 (0.453)	0.776 (0.319, 1.886)	-0.287 (0.453)	0.750 (0.309, 1.822)	-0.091 (0.453)	0.913 (0.376, 2.217)

CI: confidence interval.

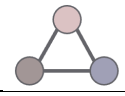
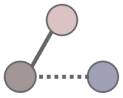
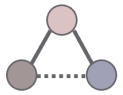
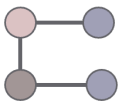
A.5.8 Results of regression using a secondary criterion of maturity
(Outcome variable: use of RWD in estimating overall survival)

Covariate	Regression A: Use of RWD in estimating OS for intervention β (SE(β))	Regression A: Use of RWD in estimating OS for intervention Odds ratio (95% CI)	Regression B: Use of RWD in estimating OS for comparators β (SE(β))	Regression B: Use of RWD in estimating OS for comparators Odds ratio (95% CI)
Maturity of survival data (MS)				
Extremely immature[a]	-	-	-	-
Immature	-0.987***	0.373***	-1.164***	0.312***
	(0.378)	(0.178, 0.781)	(0.389)	(0.146, 0.669)
Relatively mature	-1.016**	0.362**	-1.593***	0.203***
	(0.450)	(0.150, 0.875)	(0.481)	(0.079, 0.522)
Mature	-1.351**	0.259**	-1.522**	0.218**
	(0.612)	(0.078, 0.860)	(0.618)	(0.065, 0.732]
Incidence rate (IR)				
	-0.000	1.000	-0.000	1.000
	(0.000)	(1.000, 1.000)	(0.000)	(1.000, 1.000)
Direct treatment comparison (AD)				
Not available[a]	-	-	-	-
Some available	-1.052**	0.349**	-1.674***	0.187***
	(0.430)	(0.150, 0.811)	(0.438)	(0.079, 0.442)
All available	-0.534	0.587	-1.301***	0.272***
	(0.368)	(0.285, 1.208)	(0.387)	(0.128, 0.581)
External Validity (EV)				
Low risk[a]	-	-	-	-
Moderate risk	0.132	1.141	0.615	1.849
	(0.354)	(0.570, 2.284)	(0.373)	(0.890, 3.843)
High risk	0.156	1.168	0.186	1.204
	(0.522)	(0.420, 3.250)	(0.561)	(0.401, 3.613)
Previously recommended (PR)				
No[a]	-	-	-	-
Yes	0.157	1.169	-0.246	0.782
	(0.318)	(0.627, 2.180)	(0.333)	(0.407, 1.502)
Constant	0.357	1.429	1.043	2.837
Observations	229		229	
LR R^2	0.1205		0.2097	

**p<0.05, ***p<0.01.
[a] Reference group, CI: confidence interval.

Appendix 6. Classification used in the thesis

6.1 Possible combination of treatment comparison in appraisals of oncological medicine

Types of treatment comparison	Illustration	Direct treatment comparison	Indirect treatment comparison	Availability of RCT	Anchored comparison	Population adjusted comparison
Head-to-head comparison		Yes, all available	Not used	Yes, available	Yes	No
Mixed treatment comparison		Yes, all available	Yes, used	Yes, available	Yes	No
Population-adjusted indirect treatment comparison		No, not available	Yes, used	No, not available	No	No (Naïve) MAIC STC Other methods
		No, not available	Yes, used	Yes, available	No	No (Naïve) MAIC STC Other methods
Network meta-analysis					Yes	No (Naïve) MAIC STC Other methods
		Only available for some comparators	Yes, used	Yes, available	Yes	No (Naïve) MAIC STC Other methods

MAIC: Matching adjusted indirect treatment comparisons, STC: Simulated treatment comparisons

Technology of interest Comparator(s) Other treatments

Direct evidence Indirect evidence