ECONOMIC EVALUATION METHODS IN ONCOLOGY

By GRAEME BALL, M.A., B.A., B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the

Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2023) Hamilton, Ontario (Health Research Methodology)

TITLE: Economic Evaluation Methods in Oncology AUTHOR: Graeme Ball, M.A., B.A., B.A. (McMaster University) SUPERVISOR: Professor J.-E. Tarride NUMBER OF PAGES: ii, 103

Table of Contents

Chapter 5:

Chapter 1: Introduction

Since at least the early 2000s, expenditures on pharmaceutical innovations have continued to rise worldwide, and drug budgets are under increasing pressure[1,2]. While the reasons for these increases may be complex and multifactorial, several key factors have been identified, including an aging population[3] and the extremely high costs of innovation[4]. Against this backdrop of ever-increasing drug expenditures, the advent of HTA has provided a practical means for payers, governments, and other stakeholders in a growing number of countries to pursue cost-containment of innovative pharmaceuticals in a systematic, value-based way. Previously, HTA had been defined as "the systematic evaluation of the properties and effects of a health technology, addressing the direct and intended effects of this technology, as well as its indirect and unintended consequences, and aimed mainly at informing decision making regarding health technologies"[5]. However, while the lifecycle of a new pharmaceutical may last years or even decades, HTA is typically conducted only once at the time of technological adoption based on data that is available at that point in time and is rarely, if ever, updated with new data. As a result, the cost-effectiveness of pharmaceuticals across their lifecycle remains largely unresearched and unknown.

More recently, a shift in HTA towards encompassing a life-cycle approach has resulted in the emergence of a new paradigm of health technology management (HTM). Accordingly, HTA itself has been redefined as "a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle"[5]. This new definition of health technology decision making explicitly considers the value of medications at multiple points across their full lifespan, accounts for maturity of clinical data over time, and thus may be better suited to the comprehensive economic evaluation of pharmaceuticals. Several HTA agencies have expressed intent to pursue some form of HTM in order to not only evaluate the economic value of drugs at the point of adoption, but also later in their lifecycle. For example, the Canadian Agency for Drugs and Technologies in Health (CADTH) has signaled intent to

re-assess already reimbursed drugs to ensure that the clinical and economic benefits described in initial economic evaluations continues to be realized over time[6]. In the United Kingdom (UK), the National Institute for Health and Care Excellence (NICE) has recently launched a 5-year strategy in which a dynamic approach to HTM is being considered[7]. It seems likely that HTM could be implemented in some form by HTA agencies in the not-to-distant future, however, at present reimbursement decisions are rarely re-assessed, even in cases where additional data have become available.

Economic evaluations to support HTA and HTM

To support HTA in their respective countries, reimbursement submission guidelines have been issued by HTA agencies which provide prescriptive guidance for the conduct of economic evaluations. These guidelines specify the content and structure required to be submitted to facilitate review by the HTA agency. In Canada, CADTH has published several iterations of guidelines for reimbursement submissions of new drugs[8]. At NICE in the UK, extensive technology appraisal guidance documents[9], as well as a comprehensive suite of methodology guidelines, have been published. In Australia, the Pharmaceutical Benefits Advisory Committee (PBAC) also provides guidelines for appraisals of new drugs submitted for reimbursement[10]. Similarly, a multitude of guidelines are available from a growing number of countries which have implemented HTA processes, including but not limited to Japan, Korea, China, Taiwan, Sweden, Scotland, and the Netherlands[11-14]. In addition to HTA agency guidance documents, best practice guidelines have been published by several academic groups to assist researchers in the conduct of economic evaluation outside of HTA agency reimbursement submissions[11,15,16].

Theoretically, HTAs can be conducted based on either clinical trials[11] or predictive models[17]. However, trial-based economic evaluations are associated with a number of downsides, despite being a potentially superior source of data, which can include the length of time needed to conduct the analysis, the cost required to run the trial, and the limitations imposed by the analysis scope[17]. In practice, likely at least in

part due to the limitations noted above, greater than 80% of published economic evaluations have been observed to be model-based[18]. Methods development in cost-effectiveness research has grown substantially over the past several decades. The application of cost-effectiveness analysis to healthcare has occurred since at least 1982, and a formal willingness-to-pay (WTP) threshold per quality-adjusted life-year (QALY) gained has been in place in several countries since 1996[19]. An increasing number of countries are formalizing health technology assessment (HTA) as part of their drug coverage decision making processes, and value-for-money calculations acquire ever-greater importance as list prices for drugs continue to increase over time.

The evolution of economic modeling methods has expanded to include state transition modeling using Markov processes[20], and partitioned survival modeling (PSM) using the area-under-the-curve framework[21,22]. In oncology, the PSM modeling framework has become commonly utilized[18,23] for economic evaluations due to its relative ease of conducting as it directly incorporates clinical trial endpoints and can be developed without having access to individual patient-level data[24]. The analysis of survival curves from clinical studies has also evolved to include multi-state modeling (MSM) for the analysis of time-to-event data[25]. More recently, the application of cubic splines[26] in extrapolation of survival curves has allowed analysts and researchers to more accurately model survival outcomes over time, and the application of the mixture-cure[27] modeling framework to estimate the clinical and economic benefits of potentially curative therapies has also been developed for use in economic evaluation.

Thesis rationale and overview

Despite a proliferation of economic evaluation methods development over the past several decades, gaps remain in the published literature regarding which methods are being used, by whom, and what the impact of using different methods is on the results of economic evaluation. It is currently unknown

the extent to which different methods, which may include survival analysis, parametric model selection, testing the proportional hazards assumption, guideline compliance, and attributes of clinical data used (short-term vs long-term), are used in economic evaluations in oncology.

To fill a gap in the literature and to better inform decision making in oncology, this doctoral thesis investigates the role and impact of analytical methods in the economic evaluation of oncology medications through three main chapters which have been recently published.

Chapter 2 presents a systematic literature survey of published economic evaluations in oncology over a 10-year period in order to identify, examine, and describe analytical methods that have been utilized (published in Pharmacoeconomics Open in 2021). This chapter demonstrated that greater detail in reporting of extrapolation methods, statistical techniques, and validation procedures is needed in order to conform with best practices outlined in existing economic evaluation guidelines[18].

Chapter 3 complements the work of chapter 2 but takes a different perspective through an examination of the methods reported in economic evaluations published by HTA agencies in Canada, the UK, and Australia (published in Current Oncology in 2022). This chapter revealed significant reporting discrepancies across the agencies and concluded that common standards for reporting the results of HTAs should be implemented[23].

Building on chapters 2 and 3, chapter 4 provides a model-based health technology re-assessment of an oncology drug approved on the basis of interim trial data using recently published long-term follow up data (published in Current Oncology in 2023). The findings from this chapter highlight the importance of transparency in the reporting of methods, the impact of using a life-cycle approach to HTA, and demonstrate the existence of a tradeoff between clinical/economic uncertainty and the value of the incremental cost-effectiveness ratio (ICER). The final chapter provides the overall conclusions of the research and presents avenues for future research.

Importance of topic

The importance of investigating the use of methods in economic evaluation is highlighted by the fact that different methods may have differential influence on the results of analyses (e.g. ICER value, HTA recommendation), and differing results may have implications for decision making. Very few previous studies have investigated economic evaluation methods in the published oncology literature and those that have were either focused on a single disease area[28], a single methodological technique[29], a single geographic region[30,31], or examined data over a very limited timeframe of analysis[28]. Chapter 2 provides a view of the evolution of economic evaluation methods over time across all disease areas in oncology, and systematically captures a cross-section of methods usage over multiple geographic regions. This chapter presents the first comprehensive documentation of economic evaluation methods use in the published oncology literature.

Consistency in the application and reporting of economic evaluation methods is also important as manufacturers are required to conform to economic guidelines published by HTA agencies, yet HTA agencies themselves are not required to follow any sort of guideline in the reporting of their appraisals. Non-standardized reporting and appraisal may result in access disparities between countries. Divergence in outcomes from HTA agency appraisals can be most commonly observed in terms of differing numerical ICER values, and in some cases, in terms of HTA recommendations. Health technology assessment agencies are also financed with public tax dollars, and non-standardized reporting and appraisal may restrict the public's access to reimbursement information. Chapter 3 assessed the consistency of reporting and appraisal of reimbursement submissions from manufacturers across HTA agencies in Canada, the UK, and Australia, and represents a first publication looking at methods in oncology.

Innovative oncology drugs are increasingly appraised, approved, and reimbursed based on interim clinical data from phase 2 non-comparative studies in which median survival outcomes have not been reached[32], and immature data, even if initially promising, may or may not reflect the survival outcomes observed in later follow-up data[33]. Consequently, once long-term follow-up data becomes available (often several years after these medications have been reimbursed) economic evaluation based on longer-term data may yield different results, yet rarely are such reassessments undertaken. The approach taken in chapter 4 reflects a model-based application of economic evaluation methods to investigate a novel decision problem: the impact of longer-term follow-up data on the results of cost-effectiveness analysis. Prior to the publication of chapter 4, the extent to which economic evaluation results differ based on long-term clinical follow-up compared with interim data was unknown. This is the first published evidence of the direct impact of taking a life-cycle approach to HTA.

Collectively, these chapters represent the first comprehensive effort to identify, describe, and assess how economic evaluation methods in oncology are utilized in actual practice, both in the published literature and among selected HTA agencies, as well as how economic evaluation methods can impact the results of cost-effectiveness analysis using a health technology reassessment framework.

Chapter 1 references

1. Tichy EM, Hoffman JM, Suda KJ, et al. National trends in prescription drug expenditures and projections for 2022. Am J Health Syst Pharm 2022;79(14):1158-1172. (In eng). DOI: 10.1093/ajhp/zxac102.
2. Conti RM, Turner A, Hughes-Cromwick P. Projections of US Prescription Drug Spending and Key Policy Implications. JAMA Health Forum 2021;2(1):e201613-e201613. DOI: 10.1001/jamahealthforum.2020.1613.
3. Statistics Canada. A portrait of Canada's growing population aged 85 and older from the 2021 Census. (https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-X/2021004/98-200-x2021004-eng.cfm).
4. Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. Jama 2020;323(9):844-853. (In eng). DOI: 10.1001/jama.2020.1166.
5. O'Rourke B, Oortwijn W, Schuller T. The new definition of health technology assessment: A milestone in international collaboration. Int J Technol Assess Health Care 2020;36(3):187-190. (In eng). DOI: 10.1017/s0266462320000215.
6. Canadian Agency for Drugs and Technologies in Health. Health Technology Reassessment: An Overview of Canadian and International Processes. Environmental scan 86 2019 (Available online: https://www.cadth.ca/sites/default/files/es/es0334-health-technology-reassessment-es.pdf).
7. National Institutute for Health and Care Excellence. NICE Strategy 2021 to 2026. 2023 (Available online: https://static.nice.org.uk/NICE%20strategy%202021%20to%202026%20-%20Dynamic,%20Collaborative,%20Excellent.pdf (accessed on 6 January 2023).).
8. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada (4th edition). 2017 (In Available online: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf).
9. National Institutute for Health and Care Excellence. NICE health technology evaluations: the manual. Process and methods 2022 (In Available online: https://www.nice.org.uk/process/pmg36).
10. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (version 5.0). 2016 (Available online: https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf).
11. Ramsey SD, Willke RJ, Glick H, et al. Cost-Effectiveness Analysis Alongside Clinical Trials II—An ISPOR Good Research Practices Task Force Report. Value in Health 2015;18(2):161-172. DOI: https://doi.org/10.1016/j.jval.2015.02.001.
12. Research group on economic evaluation for Japanese public medical benefits. Guideline for Economic Evaluation of Healthcare Technologies in Japan. 2013.
13. China Center for Health Economic Research. China guidelines for pharmacoeconomic evaluations. 2017.
14. Bae S, Lee S, Bae EY, Jang S. Korean guidelines for pharmacoeconomic evaluation (second and updated version) : consensus and compromise. PharmacoEconomics 2013;31(4):257-67. (In eng). DOI: 10.1007/s40273-012-0021-6.

15. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. Jama 2016;316(10):1093-103. (In eng). DOI: 10.1001/jama.2016.12195.

16. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ 346: f1049. 2013.

17. Briggs A CK, Sculpher M. Decision modelling for health economic evaluation: Oxford University Press, 2006.

18. Ball G, Levine M, Thabane L, Tarride J-E. Onwards and Upwards: A Systematic Survey of Economic Evaluation Methods in Oncology. PharmacoEconomics - Open 2021;5(3):397-410. DOI: 10.1007/s41669-021-00263-w.

19. Laupacis A, Keown P, Pus N, et al. A study of the quality of life and cost-utility of renal transplantation. Kidney Int 1996;50(1):235-42. (In eng). DOI: 10.1038/ki.1996.307.

20. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998;13(4):397-409. (In eng). DOI: 10.2165/00019053-199813040-00003.

21. Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. J Clin Oncol 1989;7(1):36-44. (In eng). DOI: 10.1200/jco.1989.7.1.36.

22. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. Stat Med 1990;9(11):1259-76. (In eng). DOI: 10.1002/sim.4780091106.

23. Ball G, Levine MAH, Thabane L, Tarride JE. Appraisals by Health Technology Assessment Agencies of Economic Evaluations Submitted as Part of Reimbursement Dossiers for Oncology Treatments: Evidence from Canada, the UK, and Australia. Curr Oncol 2022;29(10):7624-7636. (In eng). DOI: 10.3390/curroncol29100602.

24. Woods BS, Sideris E, Palmer S, Latimer N, Soares M. Partitioned Survival and State Transition Models for Healthcare Decision Making in Oncology: Where Are We Now? Value in Health 2020;23(12):1613-1621. DOI: https://doi.org/10.1016/j.jval.2020.08.2094.

25. Meira-Machado L, de Uña-Alvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 2009;18(2):195-222. (In eng). DOI: 10.1177/0962280208092301.

26. Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. Journal of Statistical Computation and Simulation 2015;85(4):777-793. DOI: 10.1080/00949655.2013.845890.

27. Lambert PC. Modeling of the Cure Fraction in Survival Studies. The Stata Journal 2007;7(3):351-375. DOI: 10.1177/1536867x0700700304.

28. Bell Gorrod H, Kearns B, Stevens J, et al. A Review of Survival Analysis Methods Used in NICE Technology Appraisals of Cancer Treatments: Consistency, Limitations, and Areas for Improvement. Medical decision making : an international journal of the Society for Medical Decision Making 2019;39(8):899-909. (In eng). DOI: 10.1177/0272989x19881967.

29. Grigore B, Ciani O, Dams F, et al. Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. PharmacoEconomics 2020 (In eng). DOI: 10.1007/s40273-020-00935-1.

30. Bullement A, Cranmer HL, Shields GE. A Review of Recent Decision-Analytic Models Used to Evaluate the Economic Value of Cancer Treatments. Applied Health Economics and Health Policy 2019;17(6):771-780. DOI: 10.1007/s40258-019-00513-3.

31. Benedict A MN. Survival Modelling in UK Oncology Technology Appraisals Since the Publication of Good Practice Guidelines. 2018.

32.   Patel D, Grimson F, Mihaylova E, et al. Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. Value Health 2021;24(8):1118-1125. (In eng). DOI: 10.1016/j.jval.2021.01.015.

33.   Ball G, Levine MAH, Thabane L, Tarride J-E. Health Technology Reassessment: Addressing Uncertainty in Economic Evaluations of Oncology Drugs at Time of Reimbursement Using Long-Term Clinical Trial Data. Current Oncology. (https://mdpi-res.com/d_attachment/curroncol/curroncol-30-00484/article_deploy/curroncol-30-00484.pdf?version=1688971491).

Chapter 2: Onwards and Upwards: A Systematic Survey of Economic Evaluation Methods in Oncology

1.   Introduction

Economic evaluation in healthcare estimates the value for money of health technologies through assessment of the comparative costs and clinical impacts, the results of which can inform the value of a specified allocation of healthcare resources[1]. To increase transparency and reporting of economic methods of health technologies, several guidelines have been published. For example, the National Institute for Clinical and Care Excellence (NICE) in the United Kingdom (UK) publishes a series of practice guidelines to aid submissions from manufacturers seeking reimbursement of their health technologies. Similar guidelines have been published by the Canadian Association for Drugs and Technologies in Health (CADTH) in Canada, the 2nd Panel on Cost-Effectiveness in Health and Medicine in the US, the Research Group on Economic Evaluation for Japanese Public Medical Benefits in Japan, and the Health Insurance Review and Assessment Service (HIRA) in South Korea, among many others[2-7]. Despite widespread publication and availability of guidelines, documentation of the utilization of specific methods in economic evaluation in oncology remains limited. This is important as the last decade has seen many important methodological advances when conducting economic evaluations of oncologic treatments.

In addition to extrapolation and other survival analysis techniques, other quantitative methods (e.g. statistical testing, crossover adjustment techniques, alternative model structures) have been developed to overcome the limitations of previous methods. For example, clinical studies are finite in length, and in order for the results of clinical studies to be amenable to economic evaluation it is often necessary to extrapolate clinical outcomes beyond the study duration through survival analysis techniques[5, 8, 2, 9, 6]. In addition, the analytical methods selected for economic evaluation in oncology have been shown to

influence survival results[10], and therefore it is important for appropriate methods to be used when evaluating oncology products.

However, only two previous studies have detailed the use of survival analysis in economic evaluations in oncology[11, 12]. The first of these studies examined survival modeling and extrapolation techniques used in oncology submissions to NICE in the UK before and after publication of the NICE Decision Support Unit's Technical Support Document (TSD) on survival analysis[13]. The authors extracted data from 20 technology appraisals and reviewed information on model structure, data sources, extrapolation methods, and validation. The authors found that extrapolation techniques in practice have improved since the publication of the guidelines: some form of parametric extrapolation was used in almost all of the NICE submissions except one. However, nearly 30% of the submissions did not identify the source of overall survival data, and although the authors reported which parametric distributions were tested in the submissions, the specific distribution(s) chosen for extrapolation were not identified. Statistical testing methods used in the submissions were also documented but not disaggregated by type.

A second study from 2019 reviewed 58 NICE technology appraisals and examined the extent to which recommendations made in the NICE Decision Support Unit's Technical Support Document (TSD) on survival analysis have been followed since its publication[11]. The authors found that while there were increases in validation of results using data and/or clinical opinion following publication of the TSD, the proportion of submissions that adhered to the TSD recommendations did not change substantively over time. The authors concluded that despite the publication of the guidelines, survival analysis conducted as part of NICE technology appraisals remains suboptimal[11]. The study was limited to assessment of survival analysis and did not examine other characteristics of the NICE submissions. While these two UK studies are informative, their generalizability may be limited outside of the UK, and the results are not necessarily reflective of economic models found in the published oncology literature.

2.   Objectives

To fill a gap in the literature, the present study aims to identify, examine, and describe the analytical methods used in economic evaluations, including study characteristics and model structure, though a systematic survey of the published oncology cost-effectiveness literature over a 10-year period between 2010-2019 using commonly cited English language databases for economic evaluations in oncology. This approach allows a wider range of analytical techniques to be catalogued over a longer period of time than has been presented in previous studies. Secondary objectives of the study include examining the use of identified methods across different geographic regions.

3.   Methods

A systematic search of the published oncology literature was conducted to identify economic evaluations of advanced or metastatic cancers published between 2010-2019 using PUBMED, Ovid MEDLINE, and EMBASE databases. The PICOS method was followed for determining literature search criteria: identified studies were limited to English-language economic evaluations of advanced or metastatic cancer among adult populations, both treatment(s) and comparator(s) had to be explicitly reported, and outcomes of interest included incremental cost-effectiveness ratios (ICERs) and/or cost-utility effectiveness ratios (ICURs). The literature search was limited to economic evaluations, and duplicates and published abstracts were excluded. All eligibility criteria were defined a priori.

The following keywords were used in the database search queries: "cost-effectiveness analysis" "cost-benefit analysis", "cost-utility analysis", "quality-adjusted life-years", "metastasis", "advanced cancer", "advanced neoplasm", "metastatic neoplasm", "economics", "cost", "health economics", "budget", "costing", "price", "pharmacoeconomic", "expenditure", "expenses", "statistical model", "economic model", "probability", "Markov", "monte carlo method", "Decision Theory", and "Decision Tree". See Appendix for specific search strategies for each database.

Initial screening of titles and abstracts for relevance to the study objectives was conducted according to the stated eligibility criteria by the primary author. Full-text articles meeting all of the inclusion criteria and none of the exclusion criteria were reviewed by the primary author, and for each included study, data were extracted (see Supplementary Material for complete list of extracted studies) to describe study characteristics (i.e. disease area, patient population, type of cancer, source of clinical data, type of economic evaluation, study perspective, overall conclusions, funding source, validation methods, software used), key assumptions and modeling techniques (model structure, number of modeled health states, time horizon, intervention, comparator, treatment line, discount rates, outcomes of interest, types of analyses, key study results, total costs, base case ICER, sensitivity analyses, WTP threshold, data sources) and extrapolation methods (i.e. statistical techniques used for fitting curves, type of distribution, crossover adjustment, digitization method use).

Analytical methods identified in each included study were extracted and documented in the extraction sheet, and similarities and differences were descriptively assessed. Economic evaluations sponsored by industry have been observed to utilize longer time horizons than those conducted by HTA agencies[14]. Previous observations have also suggested that industry-sponsored studies are more likely than academic-sponsored studies to report favourable conclusions of cost-effectiveness[15]. In addition, since novel approaches have been suggested to model immunotherapy (IO)[16,17], a comparison of the model structure between IO and non-IO drugs was conducted. Chi-squared tests were used to probe relationships between categorical variables in order to substantiate these previous observations. Identified studies were also grouped according to geography in order to capture potential variation across regions. Statistical testing was performed in Microsoft Excel 2019.

A large number of studies was anticipated to be identified through the search strategies. Based on the number of studies reported in previous publications, it was determined that a sample size approximately equal to that reported in Benedict 2018[12], (n=58) and three times larger than the sample size of n=20

studies reported in Gorrod et al. 2019[11], would be an appropriate and representative snapshot of the large number of studies captured in the literature search. It was initially assumed that a 20% random sample of all eligible studies over the 2010-2019 timeframe, conducted in Microsoft Excel using a combination of RAND, INDEX, and MATCH functions, would yield at least 60 studies. Prior to knowing the exact number of studies that would meet the inclusion criteria, the random sample size of 20% was to be increased to reach a total of 60 studies if necessary. This desired sample of 60 studies was a based on previous studies on methods (Benedict 2018 [N=58], and Gorrod et al. 2019 [N=20][11,12]). To validate the representativeness of our sample, we took another random sample of 20% of the studies and compared the two samples in terms of type of economic evaluation and model structure.

4.   Results

4.1 Results of systematic survey

A total of 8481 abstracts were identified through the literature search and 1671 duplicates were removed (Figure 1). 5907 studies were excluded in level 1 screening (according to the PICOS criteria). The remaining 916 full-text studies were assessed using a pre-defined eligibility form. Of these studies, 538 were removed through level 2 screening (according to the PICOS criteria) and a total of 378 met the eligibility criteria (Figure 1). The 20% random sampling of the 378 studies meeting the inclusion criteria resulted in 76 studies to be included in the abstraction set (see Supplementary Material for complete list of included studies).

Figure 1. PRISMA diagram

```
┌─────────────────────────────┐
│ Records identified through  │
│    database searching       │
│       (n = 8481)            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Records after duplicates    │───────▶│   Duplicates removed        │
│        removed              │        │       (n = 1671)            │
│       (n = 6810)            │        └─────────────────────────────┘
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Title and abstract screening│───────▶│ Records excluded (n = 5907) │
│       (n = 6810)            │        │                             │
│                             │        │ Studies not meeting         │
│                             │        │ eligibility criteria        │
└─────────────────────────────┘        └─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌──────────────────────────────────────┐
│ Full-text articles assessed │───────▶│ Full-text articles excluded (n = 538)│
│      for eligibility        │        │                                      │
│       (n = 903)             │        │ Excluded due to:                     │
│                             │        │                                      │
│                             │        │ Inappropriate disease area  (n=0)    │
│                             │        │ Inappropriate population (n=1)       │
│                             │        │ Inappropriate intervention (n=0)     │
│                             │        │ Inappropriate comparator (n=0)       │
│                             │        │ Inappropriate outcome measure(s)(n=34)│
│                             │        │ Inappropriate study design (n=96)    │
│                             │        │ Not English language (n=10)          │
│                             │        │ Duplicates (n=5)                     │
│                             │        │ Abstract only (n=392)                │
└─────────────────────────────┘        └──────────────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Studies included for        │───────▶│ 20% random sample (80%      │
│      extraction             │        │ articles removed)           │
│       (n = 378)             │        └─────────────────────────────┘
│                             │
│ 20% random sample           │
│       (n = 76)              │
└─────────────────────────────┘
```

4.2 Study characteristics

A detailed description of the study characteristics is presented in Table 1. Briefly, close to half of the included studies originated from North America (38%), and the most commonly assessed cancer types were lung (18%), colorectal (16%) and breast cancers (15%). Approximately half of the studies were published within the last four years. A majority (82%) of identified papers were based on clinical data from phase III randomized controlled trials, while another 16% utilized observational data from real-world evidence studies. The remaining 3% of studies were based on data from random effects network meta-analyses, including one study employing a network of 5 trials[18] and the other comprised of a network of 16 studies[19]. The real-world studies were overwhelmingly retrospective in nature (92%), comprised of predominantly database analyses (67%) or based on registry data (25%). Several of the real-world studies (17%) conducted propensity score matching to balance prognostic factors between treatment arms, and 33% utilized Cox models.

Table 1. Characteristics of studies included in random sample (N=76)

Characteristic	Proportion of studies n (%)
Regions/Country	
North America	29 (38%)
East Asia (China, Japan, Korea, Singapore, Taiwan)	16 (21%)
Europe	14 (18%)
United Kingdom	12 (16%)
South America	2 (3%)
Australia	2 (3%)

Africa	1 (1%)
Type of cancer studied	
Lung cancer	14 (18%)
Colorectal cancer	12 (16%)
Breast cancer	10 (13%)
Diagnostics	8 (11%)
Prostate cancer	7 (9%)
Pancreatic cancer	5 (7%)
Other cancer types (n<5)	20 (26%)
Main source of clinical data	
Phase III RCT	62 (82%)
Real-world evidence	12 (16%)
Network meta-analysis	2 (3%)
Funding source	
Industry	25 (33%)
Public grant	23 (30%)
No funding	21 (28%)
Not reported	6 (8%)
Mix of public and private funding	1 (1%)
Journal type	
Open access	13 (17%)
Standard	63 (83%)

Abbreviations: n number, RCT randomized controlled trial

Funding sources of the identified studies were relatively well-balanced between industry sponsorship (33%), public grants (32%), and no declaration of funding (28%). Cross-referencing the time horizon selected by study authors with the source of funding revealed that industry-sponsored studies were more likely to use longer time horizons. Of the 32 included studies that used a time horizon of 10+ years (including lifetime), 47% were industry-sponsored whereas 16% were funded through public grants. Conversely, for the studies using shorter time horizons of 5 years or less, funding sources were more evenly distributed with 39% sponsored by industry and 39% funded through public grants. However, no statistically significant relationship was identified between funding sources and time horizon (Chi-square test: p=0.2939).

4.3 Key assumptions and modeling techniques from identified studies

Over three-quarters of all included studies (82%) were cost-utility analyses and 83% were conducted from a public healthcare system perspective. The most common model structure was the Markov model (47%) followed by partitioned survival model (17%). More than half (57%) of all included studies concluded that the intervention under investigation was cost-effective and this proportion was higher for industry-sponsored studies (76%) (Chi-square test: p=0.0054). Details are presented in Table 2.

Table 2. Key assumptions and modeling techniques used in the studies included in the random sample (N=76)

Characteristic	Proportion of studies n (%)
Analytical technique	
CUA	62 (82%)
CEA	13 (17%)

Other	1 (1%)
Study perspective	
Public healthcare system	63 (83%)
Societal	5 (7%)
Hospital	5 (7%)
Not reported	3 (4%)
Model structure	
Markov	37 (49%)
Partitioned survival	13 (17%)
Not reported	7 (9%)
Decision tree	5 (7%)
Combination (decision tree + Markov)	5 (7%)
Other	4 (5%)
Microsimulation	3 (4%)
Discrete event simulation	2 (3%)
Time horizon	
≤1 year	1 (1%)
1-5 years	19 (25%)
6-10 years	18 (24%)
11+ years	9 (12%)
Lifetime	25 (33%)
Not reported	4 (5%)
Crossover adjustment (N=76)	

Efficacy results adjusted for crossover	7 (9%)
Reporting of results	
Deterministic sensitivity analysis	60 (79%)
Probabilistic sensitivity analysis	64 (84%)
Cost-effectiveness acceptability curves	47 (62%)
Scenario analysis	24 (32%)
Authors' primary conclusion	
Intervention cost-effective	43 (57%)
Intervention not cost-effective	29 (38%)
Not reported	4 (5%)
Modeling software used	
TreeAge Pro	24 (32%)
Not reported	20 (26%)
Microsoft Excel	24 (32%)
SAS	4 (5%)
R	2 (3%)
Stata	1 (1%)
C++	1 (1%)

Some percentages may add up to more than 100% due to rounding

Abbreviations: CEA cost-effectiveness analysis, CUA cost-utility analysis

In approximately 9% of studies, methods for cross-over adjustments were reported to have been used. Of these 7 studies, some of which reported more than one crossover adjustment method, the most commonly cited methods were Cox regression with crossover as a time-dependent covariate (29%), rank

preserving structural failure time (29%), and inverse probability of censoring weights (29%). However, 3 of the 7 studies (43%) did not report the specific crossover adjustment method used.

4.4 Extrapolation methods

Forty-nine percent of studies reported extrapolation of survival endpoints and 19% of these studies reconstructed Kaplan-Meier curves using digitization techniques (Table 3). When reported, the average number of months of extrapolation beyond the clinical study duration was approximately 90 months. Among the 49% of studies that extrapolated results, 89% reported extrapolation using fitted parametric curves. Hybrid models combining both Kaplan-Meier trial data and extrapolated data was relatively rare (5%), and only 2 studies (5%) used solely the hazard ratio method to extrapolate over time (both based on patient-level observational data). Other non-common methods used for extrapolation included use of a simple average monthly transition probability applied across years (3%), and transition probabilities calibrated to minimize mean squared differences between trial survival endpoints and model-generated curves (3%). Less than 20% of the extrapolated studies reported testing the proportional hazards assumption to justify their extrapolation. Half of these used the log-cumulative hazards plot to assess proportional hazards, while the other half did not report the method of assessment used.

Table 3. Extrapolation techniques and methods reported in the articles included in the random sample

Characteristic	Number of studies n (%)
Extrapolation (N=37)	
Kaplan-Meier curves included	2 (5%)
Fitted curves only	35 (95%)
Extrapolation method (N=37)	

Fitted parametric curves	33 (89%)
Hazard ratio method	2 (5%)
Other methods (n=1)	2 (5%)
Statistical fit (N=37)	
Akaike Information Criterion (AIC)	12 (33%)
Bayesian Information Criterion (BIC)	7 (19%)
Digitization of survival curves (N=37)	
Yes	7 (19%)
No	19 (53%)
Not reported	10 (28%)
Proportional hazards assumption (N=37)	
Tested	6 (16%)
Validation (N=76)	
Study results validated (e.g. using RWE, etc.)	16 (21%)
Distributions selected for extrapolation (N=28)	

Treatment	**PFS (N=21)**		**OS (N=27)**	
Weibull	11	39%	18	64%
Log logistic	2	7%	4	14%
Log normal	5	18%	1	4%
Exponential	1	4%	1	4%
Generalized gamma	2	7%	3	11%
Comparator	**PFS (N=21)**		**OS (N=27)**	
Weibull	10	36%	17	61%

Log logistic	1	4%	4	14%
Log normal	6	21%	1	4%
Exponential	2	7%	2	7%
Generalized gamma	2	7%	3	11%

Abbreviations: N number, OS overall survival, PFS progression-free survival, RWE real-world evidence

Overall, 28 studies (37%) reported which distributions were used for extrapolation, and among these 28 studies the Weibull distribution was the most commonly used parametric distribution for both treatment and comparator arms in modeling PFS (39%, 36%) and OS (64%, 61%), followed by the log normal distribution for PFS (18%, 21%) and the log-logistic for OS (14%, 14%). Distribution selection for PFS was reported in 21 studies, 27 studies reported distribution selection for OS, and the Akaike Information Criterion (AIC) was the most commonly reported method for identifying best statistical fit (32%). The majority of identified studies did not validate the results of their analyses and extrapolations; only 21% of the identified studies performed a validation procedure, and validation was more commonly performed in studies published in later years compared with earlier years. The most common validation techniques reported were clinical experts (44%), comparison with previous studies (31%), comparison with RWE (13%), and creation of a separate validation model (13%).

4.5 Additional analyses

The greatest number of identified studies were from North American countries (38%), followed by countries in East Asia (21%), continental Europe (18%), and the UK (16%). A majority of the North American studies (59%), European studies (64%), and UK studies (58%) found the treatment under investigation to be cost-effective at the willingness-to-pay threshold cited by the authors, while this proportion was 50% for studies from East Asia. Among studies which reported conducting validation exercises, some small geographic variation was noted: 11% studies from North America reported

validation, compared with 6% of studies from East Asia, 5% of UK studies, 4% of studies from continental

Europe. In addition, a much higher proportion of studies from the UK (100%) and North America (93%)

reported use of probabilistic sensitivity analysis than studies from other geographic regions (continental

Europe: 71%, East Asia: 69%). Comparable trends were observed for utilization of deterministic

sensitivity analysis, scenario analyses, and use of cost-effectiveness acceptability curves. Sensitivity

analyses were also reported in some studies from South America, Australia, and Africa, but the number

of studies included for each of these regions was too low to draw clear inferences or conclusions.

Some regional variation was also observed when analyzing survival extrapolation methods across

jurisdictions. The use of fitted parametric curves for extrapolation was frequently reported in studies

from the UK (83%), about twice as often as reported in studies from other geographic regions (East Asia:

44%, North America: 38%, continental Europe: 29%). Additional variation between regions included the

number of studies reporting adjustment for crossover, which was higher in the UK (25%) than in other

regions (0-7%), and testing of the proportional hazards assumption which was consistently rare across

most regions (between 7%-17%) but was not reported at all in studies from East Asia. Other methods

were broadly similar across regions. Results of the analysis comparing IO (n=19) and non-IO (n=57)

economic evaluations indicated that while approximately half of both IO (58%) and non-IO studies (46%)

used Markov models, a greater proportion of IO-focused studies used partitioned survival models (32%)

compared with non-IO studies (12%).

Finally, to investigate the representativeness of the 20% random sample, a second 20% random sample

was taken. Comparing types of models in the 2nd random sample with the original random sample, the

proportion of studies reporting the primary analysis as cost-utility, cost-effectiveness, or "other" were

84%, 13%, and 2% in the second random sample, and 82%, 17%, and 1% in the original random sample.

A high degree of concordance between the two random samples was also observed for the proportions

of studies reporting common model structures: Markov models (58% vs. 49%), partitioned survival (17%

vs. 17%), decision tree (8% vs. 7%), "other" (7% vs. 5%), and not reported (9% vs. 9%). A small degree of variation was seen for less commonly used model structures: combination (0% vs. 7%), discrete event simulation (0% vs. 3%), and microsimulation (1% vs. 4%).

5. Discussion

5.1 Summary of main results

This 20% random sample of published economic evaluations over the past decade has shown that many advances in economic evaluation methods have diffused into common usage. These methods included deterministic and probabilistic sensitivity analysis, extrapolation of outcomes beyond the duration of clinical trials, utilization of cost-effectiveness acceptability curves, both Markov and partitioned survival model structures, and the cost-utility analytical framework. Less frequently or inconsistently utilized methods included testing of the proportional hazards assumption (for those studies in which it would have been appropriate to do so), assessing statistical fit of survival extrapolations, and validating study results. Looking at the study sample across geographic regions, heterogeneity was observed in the use and reporting of procedures for validating results, statistical curve fitting techniques, testing of proportional hazards assumption, and adjustment for crossover. While new methods may be developed over time, uniform uptake across regions is not guaranteed, even when supported by the publication of economic evaluation guidelines.

5.2 Explanation of findings and comparison with other studies

Previously cited reviews of economic evaluations in oncology have examined data over a very limited period or have been focused on a specific jurisdiction[12, 11]. An additional study from late 2019 examined modeling approaches in 100 NICE technology appraisals and 124 published studies, finding that the state transition model (41.0%, 82.3%) and partitioned-survival model (54.0%, 12.1%) were the most commonly utilized model structures in NICE submissions and published oncology literature,

respectively[20]. However, this study was limited to a 5-year period and investigated model structure

exclusively and did not review for example the model assumptions regarding PFS or survival

extrapolations. To the authors' current knowledge, the present analysis provides the first examination of

published English-language economic evaluations in oncology across a 10-year period, focused on

multiple modeling methods across multiple jurisdictions, and cataloguing trends in methods uptake

across geographies.

Comparing the study characteristics observed in the present study based on published economic

evaluations in oncology between 2010-2019 to the results of previously conducted studies reveals a

number of similarities. First, Markov models and partitioned survival models appear to be the most

commonly utilized model structures across most geographic regions, which suggests that these methods

have been broadly accepted and integrated into economic evaluation processes. Comparisons between

UK studies included in the present analysis and previous studies (also UK studies) demonstrates further

similarities in terms of average model duration (time horizon), use of fitted parametric curves and

extrapolation techniques, the use of procedures to validate results, and testing of the proportional

hazards assumption. For example, use of fitted parametric survival curves were found to be similar

between previous publications and the UK subset from the present study (76%[12], 91%[11], and 83%,

respectively).

Insights beyond those presented in previous studies include results presented across more than a single

geography. For example, while the use of partitioned survival models has been extensively observed by

Benedict and colleagues (61%)[12], Bullement and colleagues (54%)[20], and UK studies included in the

present analysis (42%), this model structure was seldomly observed in studies from North America

(17%), continental Europe (7%) and East Asia (6%). Studies from jurisdictions outside the UK tended to

favour the use of Markov models (24%, 43% and 56%, respectively). In contrast to the results from

previous UK studies, fitted parametric survival curves were also much less frequently reported in studies

from East Asian countries (44%), North America (38%), and continental Europe (29%). These observations suggest that there may be important differences in uptake of economic evaluation methods across geographic regions, and these could potentially lead to differences in decision making.

5.3 Limitations

The present study is associated with a number of limitations. Search parameters were limited to articles published in English, exhibiting a bias towards studies from countries that have English as a first language. Second, not all economic evaluations in oncology require or report extrapolation of survival endpoints, and thus the total number of studies from which inferences may be drawn around survival outcomes may be limited. In addition, while the 20% random sample taken was assumed to be representative of the entire 378 identified studies that met the inclusion criteria, there is no guarantee that the studies not included in the random sample would provide similar or corroborative results, though this may be an intuitively plausible conjecture. To address this limitation, a second 20% random sample was taken in which the analysis type and model structure characteristics were found to be similar to the proportions observed in the original 20% random sample, providing some reassurance regarding its representativeness. It was also assumed that if study authors did not mention extrapolation, it was assumed that study results were not extrapolated. In some cases, absence of this evidence may not be evidence of its absence, though the number of studies to which this applies might be expected to be small. Comparisons between regions were also limited by the relatively small number of studies per group.

This study was focused on published literature exclusively, and thus there is a potential for publication bias since oncology models built for reimbursement submissions were not included in the study. The study results also do not directly capture the impact of the evolution of methods on oncology models submitted to HTA agencies or the subsequent reimbursement recommendations made based on those

models. Conference abstracts, too, were excluded, implying the potential for not having captured some of the most up-to-date economic evaluation methods in use, and this may in turn affect the external validity of this study. However, since the content of conference abstracts is necessarily limited and may differ from the content included in full publications, excluding them in the present study may be justifiable for the purposes of comparison with published economic evaluations.

5.4 Future directions

As economic evaluation becomes increasingly embedded in decision making, a subsequent increase in the aggregate number of published studies can be expected. This increase will provide opportunities to re-evaluate the uptake of methods in light of development of new guidelines, survival analysis techniques, and other methods. While our study found that Markov and partitioned survival models were the most common structures used among our sample of studies published between 2010 and 2019, future research could focus on the use of novel modeling techniques such as discrete event simulation[20-24], multi-state modeling[20, 25, 26], and mixture cure models[27, 28] which are more frequently used to overcome specific limitations inherent in more rudimentary analytical approaches. Discrete event simulation, for example, is typically used when the implementation of a defined model structure is not manageable as a cohort- based state transition model, or when baseline heterogeneity, continuous disease markers, time varying event rates, and the influence of prior events on subsequent event rates are of relevance to decision making[22]. Recent advances in the development of anti-cancer therapies have led to the advent of therapeutics that may be curative for certain patients, leading to recommendations of using mixture cure modeling [29]. Since we might expect to see more frequent use of newer modeling methods[20], and given that these novel methods have not (yet) been incorporated into current guidelines, it could be helpful and informative for future research efforts to track and document the diffusion of these newer methods into use over time, both in the published oncology literature and in technology appraisals from health technology assessment bodies.

6. Conclusion

This review of published economic evaluations in oncology has shown that over the past decade a majority of the identified papers reported basic characteristics of study type, data source used, modeling techniques, and utilization of survival analysis methods. However, greater detail in reporting extrapolation methods, statistical analyses, and validation of results could be potential improvements. Regional variation observed in the use of these methods warrants further examination in order to support greater consistency in decision making. Future research efforts could be dedicated towards documenting the diffusion of novel modeling techniques into economic evaluation.

Chapter 2 references

1. Richardson J, Schlander M. Health technology assessment (HTA) and economic evaluation: efficiency or fairness first. Journal of market access & health policy. 2019;7(1):1557981. doi:10.1080/20016689.2018.1557981.

2. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. Jama. 2016;316(10):1093-103. doi:10.1001/jama.2016.12195.

3. Bae S, Lee S, Bae EY, Jang S. Korean guidelines for pharmacoeconomic evaluation (second and updated version): consensus and compromise. PharmacoEconomics. 2013;31(4):257-67. doi:10.1007/s40273-012-0021-6.

4. Research group on economic evaluation for Japanese public medical benefits. Guideline for Economic Evaluation of Healthcare Technologies in Japan [in Japanese]. 2013. Available at: https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=201402257860290440 [Accessed June 16, 2019]

5. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013 [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013 Apr 4. Process and Methods Guides No. 9. PMID: 27905712. Available at: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [Accessed February 9, 2020]

6. Drummond MF, Schwartz JS, Jonsson B, Luce BR, Neumann PJ, Siebert U et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. Int J Technol Assess Health Care. 2008;24(3):244-58; discussion 362-8. doi:10.1017/s0266462308080343.

7. China Center for Health Economic Research. China guidelines for pharmacoeconomic evaluations [in Chinese]. 2011. Available at: https://tools.ispor.org/PEguidelines/source/China-Guidelines-for-Pharmacoeconomic-Evaluations_2011_Chinese.pdf

8. Briggs A CK, Sculpher M. Decision Modelling for Health Economic Evaluation. 2006(New York):Oxford University Press.

9. CADTH. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. 2018. Available at: https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada [Accessed June 29, 2019]

10. Wan X, Peng L, Li Y. A review and comparison of methods for recreating individual patient data from published Kaplan-Meier survival curves for economic evaluations: a simulation study. PLoS One. 2015;10(3):e0121353. doi:10.1371/journal.pone.0121353.

11. Bell Gorrod H, Kearns B, Stevens J, Thokala P, Labeit A, Latimer N et al. A Review of Survival Analysis Methods Used in NICE Technology Appraisals of Cancer Treatments: Consistency, Limitations, and Areas for Improvement. Medical decision making: an international journal of the Society for Medical Decision Making. 2019;39(8):899-909. doi:10.1177/0272989x19881967.

12. Benedict A MN. Survival Modelling in UK Oncology Technology Appraisals Since the Publication of Good Practice Guidelines. 2018.

13. Latimer N. NICE DSU Technical Support Document 14: undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf [Accessed September 20, 2019]

14. Masucci L, Beca J, Sabharwal M, Hoch JS. Methodological Issues in Economic Evaluations Submitted to the Pan-Canadian Oncology Drug Review (pCODR). PharmacoEconomics - open. 2017;1(4):255-63. doi:10.1007/s41669-017-0018-3.

15. Valachis A, Polyzos NP, Nearchou A, Lind P, Mauri D. Financial relationships in economic analyses of targeted therapies in oncology. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2012;30(12):1316-20. doi:10.1200/jco.2011.38.6078.

16. Gibson EJ, Begum N, Koblbauer I, Dranitsaris G, Liew D, McEwan P, Tahami Monfared AA, Yuan Y, Juarez-Garcia A, Tyas D, Lees M. Modeling the economic outcomes of immuno-oncology drugs: alternative model frameworks to capture clinical outcomes. Clinicoecon Outcomes Res. 2018 Mar 8;10:139-154.

17. Gibson EJ, Begum N, Koblbauer I, Dranitsaris G, Liew D, McEwan P, Yuan Y, Juarez-Garcia A, Tyas D, Pritchard C. Cohort versus patient level simulation for the economic evaluation of single versus combination immuno-oncology therapies in metastatic melanoma. J Med Econ. 2019 Jun;22(6):531-544.

18. Lu S, Ye M, Ding L, Tan F, Fu J, Wu B. Cost-effectiveness of gefitinib, icotinib, and pemetrexed-based chemotherapy as first-line treatments for advanced non-small cell lung cancer in China. Oncotarget. 2017;8(6):9996-10006. doi:10.18632/oncotarget.14310.

19. Coyle D, Ko YJ, Coyle K, Saluja R, Shah K, Lien K et al. Cost-Effectiveness Analysis of Systemic Therapies in Advanced Pancreatic Cancer in the Canadian Health Care System. Value Health. 2017;20(4):586-92. doi:10.1016/j.jval.2016.11.002.

20. Bullement A, Cranmer HL, Shields GE. A Review of Recent Decision-Analytic Models Used to Evaluate the Economic Value of Cancer Treatments. Applied Health Economics and Health Policy. 2019;17(6):771-80. doi:10.1007/s40258-019-00513-3.

21. Pan F, Reifsnider O, Zheng Y, Proskorovsky I, Li T, He J et al. Modeling Clinical Outcomes in Prostate Cancer: Application and Validation of the Discrete Event Simulation Approach. Value Health. 2018;21(4):416-22. doi:10.1016/j.jval.2017.09.022.

22. Karnon J, Haji Ali Afzali H. When to Use Discrete Event Simulation (DES) for the Economic Evaluation of Health Technologies? A Review and Critique of the Costs and Benefits of DES. PharmacoEconomics. 2014;32. doi:10.1007/s40273-014-0147-9.

23. Wang HI, Roman E, Crouch S, Aas E, Burton C, Patmore R et al. A Generic Model for Follicular Lymphoma: Predicting Cost, Life Expectancy, and Quality-Adjusted-Life-Year Using UK Population-Based Observational Data. Value Health. 2018;21(10):1176-85. doi:10.1016/j.jval.2018.03.007.

24. Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. Value Health. 2012;15(8):1127-36. doi:10.1016/j.jval.2012.07.008.

25. Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with

Partitioned Survival and Markov Decision-Analytic Modeling. Medical Decision Making. 2016;37(4):427-39. doi:10.1177/0272989X16670617.

26. Williams C, Lewsey J, Briggs A, Mackay D. Cost-Effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. Medical Decision Making. 2016;37. doi:10.1177/0272989X16651869.

27. Othus M, Bansal A, Koepl L, Wagner S, Ramsey S. Accounting for Cured Patients in Cost-Effectiveness Analysis. Value Health. 2017;20(4):705-9. doi:10.1016/j.jval.2016.04.011.

28. Ishak J, Villalobos CF, Proskorovsky I. Projecting survival with cure mixture models: when are the data mature enough for reliable analysis? Value in Health. 2018;21:S53. doi:10.1016/j.jval.2018.04.443.

Chapter 2 supplementary materials

Table A1. List of included studies

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Asukai, et al.	2010	Spain	Markov	trial	pemetrexed	N
Sher, et al.	2010	USA	Markov	trial	positron emission test/computed tomography	N
Retel, et al.	2010	Netherlands	Markov	trial	70-gene signature test	N
Marino, et al.	2010	France	NR	trial	FEC-D (fluorouracil, epirubicin, cyclophosphamide, docetaxel)	N
Cunio et al.	2011	Brazil	Markov	trial	clodronate	N
Iannazzo, et al.	2011	Italy	Microsimulation	RWE	leuprorelin 22.5mg	N
Xie, et al.	2011	USA	Markov	trial	denosumab	Y
De et al.	2012	Netherlands	Other	trial	adjuvant systemic therapy	N
Casciano, et al.	2012	USA	Markov	trial	everolimus	N
Hannouf, et al.	2012	Canada	Markov	trial	cetuximab+platinum-based chemotherapy	Y
Dranitsaris, et al.	2012	Spain	decision tree	trial	FOLFOX+new drug	N
Delea, et al.	2012	UK	partitioned survival	trial	lapatinib+capecitabine	Y
Hornberger, et al.	2013	USA	NR	trial	gene-expression profiling on tissue of origin	N

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Mihajlovic, et al.	2013	Serbia	Markov	trial	everolimus	N
Brown, et al.	2013	UK	Markov	trial	venorelbine	N
Vuong, et al.	2013	Germany	decision tree	RWE	stereotactic radiosurgery	N
Alba, et al.	2013	Spain	Markov	trial	nanoparticle albumin-bound paclitaxel	N
Zeng, et al.	2013	China	Markov	trial	continuation maintenance pemetrexed	N
Lawrence, et al.	2013	Canada	Markov	trial	bevacizumab+chemotherapy	Y
Hoyle, et al.	2013	UK	partitioned survival	trial	cetuximab	Y
Kilonzo, et al.	2013	UK	partitioned survival	trial	pazopanib	N
Amdahl, et al.	2014	UK	combination	trial	pazopanib	N
Gong et al.	2014	USA	Markov	trial	abiraterone	N
Bentley, et al.	2014	USA	decision tree	RWE	92-gene assay	N
Hannouf, et al.	2014	Canada	combination	RWE	21-gene recurrence score assay	N
Zheng, et al.	2014	USA	Other	RWE	1 line of treatment	N

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Lee, et al.	2014	UK	Markov	trial	degarelix	N
Hong, et al.	2015	Canada	NR	trial	positron emission test/computed tomography	N
Roberts, et al.	2015	UK	Markov	RWE	operative cohort	N
Wen, et al.	2015	China	Markov	trial	RAS testing+cetuximab+FOLFIRI RAS testing+bevacizumab+FOLFIRI	Y
Goldstein, et al.	2015	USA	Markov	trial	bevacizumab + chemotherapy (FOLFOX)	Y
Li, et al.	2015	USA	partitioned survival	trial	single-site mutation test	N
Kumar, et al.	2015	USA	Other	trial	multiple	Y
Pennington et al.	2015	UK	partitioned survival	trial	radiation	N
Lin, et al.	2016	Taiwan	Other	RWE	chemoradiotherapy	N
Zhou, et al.	2016	China	Markov	trial	FOLFOX+cetuximab FOLFOX+bevacizumab	Y
Balcik, et al.	2016	Turkey	Markov	trial	pemetrexed/cisplatin	N
Zhou, et al.	2016	China	Markov	trial	FOLFIRINOX	N
Nishie, et al.	2017	Japan	Markov	trial	EOB magnetic resonance imaging	N
Raphael et al.	2017	Canada	discrete event simulation	trial	palbociclib+letrozole	N

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Lu, et al.	2017	China	combination	NMA	icotinib	N
Shiroiwa, et al.	2017	Japan	NR	trial	S-1	N
Bongers, et al.	2017	Netherlands	microsimulation	RWE	sequential chemo-radiation	N
Parikh, et al.	2017	USA	discrete event simulation	RWE	oxaliplatin/irinotecan, then oxaliplatin/irotecan+bevacizumab	Y
Doble, et al.	2017	Australia	combination	trial	mulitplex targeted sequencing	N
Zheng, et al.	2017	China	Markov	trial	docetaxel+androgen deprivation therapy	N
Huang, et al.	2017	USA	partitioned survival	trial	pembrolizumab	Y

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Coyle, et al.	2017	Canada	Markov	NMA	FOLFIRINOX	N
Gharaibeh, et al.	2017	USA	partitioned survival	trial	FOLFIRINOX	N
Keller, et al.	2017	Australia	Markov	trial	screening	N
Miguel, et al.	2017	Portugal	partitioned survival	trial	Pembrolizumab	Y
Lertjanyakun, et al.	2018	Japan	Markov	trial	exemestane	N
Wu, et al.	2018	China	combination	trial	osimertinib	N
Lotan, et al.	2018	USA	decision tree	RWE	biomarker-based approaches DNA-repair genes ERCC2 RNA subtyping	N
Nixon, et al.	2018	Canada	Markov	trial	PRO monitoring (web-based self-monitoring of symptoms tool)	N
Gharaibeh, et al.	2018	UK	partitioned survival	trial	FOLFIRINOX	N

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Kimura, et al.	2018	Japan	NR	trial	afatinib	N
Ball, et al.	2018	Canada	partitioned survival	trial	bevacizumab	Y
James, et al.	2018	UK	microsimulation	trial	docetaxel+standard of care	N
Uyl-de et al.	2018	Netherlands	Markov	RWE	cetuximab	Y
Li, et al.	2018	USA	Markov	trial	diagnostic staging laparoscopy	N
Leung, et al.	2018	Taiwan	Markov	trial	pertuzumab+trastuzumab+docetaxel	Y
Kimura, et al.	2018	Japan	NR	trial	ramucirumab+paclitaxel	Y
Mujica-Mota, et al.	2018	UK	Markov	trial	everolimus (lutetium-177 dotatate - identified, but study removed during screening) sunitinib	N
Harty, et al.	2018	UK	Markov	trial	cetuximab+FOLFIRI	Y
Elsisi et al.	2018	Egypt	Markov	trial	sorafenib	N
Bolagnos-Diaz et al.	2018	Peru	Markov	trial	cetuximab+chemotherapy	Y
Li, et al.	2019	China	Markov	trial	bevacizumab+chemotherapy	Y
Garrison, et al.	2019	USA	Markov	trial	pertuzumab+trastuzumab+chemotherapy	Y
Redig et al.	2019	Sweden	NR	RWE	targeted therapies (early)	Y

Author	Year	Country	Type of model	Clinical data source	Intervention(s)	I/O therapy? (Y/N)
Hamilton, et al.	2019	USA	Markov	trial	cytoreductive surgery+hyperthermic intraperitoneal chemotherapy	N
Pruis, et al.	2019	Singapore	partitioned survival	trial	sunitinib	N
Ondhia, et al.	2019	Canada	partitioned survival	trial	atezolizumab	Y
Insinga, et al.	2019	USA	partitioned survival	trial	pembrolizumab+chemo	Y
Quinn, et al.	2019	USA	decision tree	trial	sentinel lymph node biopsy	N
Raldow et al.	2019	USA	Markov	trial	short-course radiation	N

Abbreviations: I/O, immuno-oncology; N, no; NMA, network meta-analysis; RWE, real-world evidence, UK, United Kingdom; US, United States; Y, yes

Table caption: list of all articles included in this study

Table A2. List of included studies CONT'D

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Asukai, et al.	docetaxel	CUA	Y	Y	Y	N	fitted parametric curves
Sher, et al.	computed tomography no imaging	CUA	Y	Y	Y	Y	NR
Retel, et al.	Sankt Gallen guidelines Adjuvant Online Software	CEA	N	N	Y	Y	NR
Marino, et al.	FEC100 (fluorouracil, epirubicin, cyclophosphamide)	CUA	N	N	Y	Y	No extrapolation
Cunio et al.	zoledronate	CUA	Y	Y	Y	N	No extrapolation
Iannazzo, et al.	leuprorelin 11.25mg goserelin triptorelin buserelin	CUA	N	N	Y	N	HR method
Xie, et al.	zoledronic acid	CEA	Y	Y	Y	Y	No extrapolation
De et al.	No adjuvant systemic therapy	CEA	Y	Y	Y	Y	No extrapolation
Casciano, et al.	sunitinib	CUA	Y	Y	Y	N	fitted parametric curves
Hannouf, et al.	Platinum-based chemotherapy	CUA	Y	Y	Y	N	Other
Dranitsaris, et al.	FOLFOX	Other	N	N	N	N	NR
Delea, et al.	capecitabine trastuzumab+capecitabine	CUA	Y	Y	Y	Y	fitted parametric curves
Hornberger, et al.	"usual care"	CUA	Y	Y	Y	N	No extrapolation
Mihajlovic, et al.	best supportive care	CUA	Y	Y	Y	N	fitted parametric curves

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Brown, et al.	chemotherapy	CUA	Y	Y	Y		fitted parametric curves
Vuong, et al.	surgical resection	CEA	N	N	Y	N	No extrapolation
Alba, et al.	paclitaxel	CUA	Y	Y	Y	N	fitted parametric curves
Zeng, et al.	placebo	CUA	Y	Y	Y	N	fitted parametric curves
Lawrence, et al.	chemotherapy panitumumab+chemotherapy cetuximab+chemotherapy	CUA	Y	Y	Y	N	Other
Hoyle, et al.	cetuximab+irotecan panitumumab best supportive care	CUA	Y	Y	Y	N	fitted parametric curves
Kilonzo, et al.	sunitinib interferon-α best supportive care	CUA	Y	Y	Y	N	fitted parametric curves
Amdahl, et al.	**Direct comparison**: best supportive care **Indirect comparison**: trabectedin, ifosfamide, gemcitabine, docetaxel, lenograstim, Mesna	CUA	Y	Y	Y	N	fitted parametric curves
Gong et al.	prednisone Sipuleucel-T	CUA	Y	Y	Y	N	NR
Bentley, et al.	standard of care	CUA	Y	Y	Y	N	NR
Hannouf, et al.	Canadian clinical practice	CUA	N	N	Y	Y	NR
Zheng, et al.	2 lines of treatment	CEA	N	N	N	Y	NR
Lee, et al.	leuprorelin	CUA	Y	Y	Y	Y	fitted parametric curves

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Hong, et al.	chemotherapy physical exam/radiography	CEA	Y	Y	Y	N	No extrapolation
Roberts, et al.	non-operative cohort	CUA	Y	Y	Y	Y	No extrapolation
Wen, et al.	KRAS testing+cetuximab+FOLFIRI KRAS testing+bevacizumab+FOLFIRI	CUA	Y	Y	Y	N	NR
Goldstein, et al.	chemotherapy (FOLFIRI)	CUA	Y	Y	Y	Y	fitted parametric curves
Li, et al.	Next generatino gene-sequencing panel	CEA	Y	Y	Y	N	NR
Kumar, et al.	multiple	CEA	Y	Y	Y	N	fitted parametric curves
Pennington et al.	best supportive care	CUA	Y	Y	Y	Y	fitted parametric curves
Lin, et al.	esophagectomy	CEA	Y	Y	N	N	NR
Zhou, et al.	FOLFIRI+cetuximab FOLFIRI+bevacizumab	CUA	Y	Y	N	N	NR
Balcik, et al.	gemcitabine/cisplatin	CUA	Y	Y	N	N	NR
Zhou, et al.	gemcitabine+nab-paclitaxel	CUA	Y	Y	N	N	NR
Nishie, et al.	ECCM magnetic resonance imaging CE computed tomography	CUA	Y	Y	Y	Y	No extrapolation
Raphael et al.	letrozole	CUA	N	N	Y	Y	NR
Lu, et al.	pemetrexed+cisplatin pemetrexed gefitinib gefitinib+patient assistance program icotinib+patient assistance program	CUA	Y	Y	Y	N	fitted parametric curves
Shiroiwa, et al.	taxane	CUA	N	N	Y	N	No extrapolation

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Bongers, et al.	concurrent chemotherapy-radiation standard sequential radiation concurrent chemotherapy-radiation	CUA	N	N	N	N	HR method
Parikh, et al.	oxaliplatin/irinotecan+bevacizumab, then oxaliplatin/irotecan+bevacizumab oxaliplatin/irinotecan, then oxaliplatin/irotecan+bevacizumab, then targeted biologic oxaliplatin/irinotecan+bevacizumab, then oxaliplatin/irotecan+bevacizumab, then targeted biologic	CUA	N	N	Y	N	fitted parametric curves
Doble, et al.	chemotherapy, best supportive care	CUA	Y	Y	N	N	NR
Zheng, et al.	androgen deprivation therapy	CUA	Y	Y	Y	N	NR
Huang, et al.	pemextred+carboplatin pemextred+cisplatin gemcitabine+cisplatin gemcitabine+carboplatin paclitaxel+carboplatin	CUA	Y	Y	Y	Y	fitted parametric curves
Coyle, et al.	gemcitabine, gemcitabine+5-fluorouracil, gemcitabine+capecitabine), gemcitabine+cisplatin, gemcitabine+oxaliplatin, gemcitabine+erlotinib, gemcitabine+nab-paclitaxel	CUA	Y	Y	Y	N	fitted parametric curves
Gharaibeh, et al.	nab-paclitaxel+gemcitabine gemcitabine	CUA	Y	Y	Y	Y	fitted parametric curves

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Keller, et al.	no screening	CUA	Y	Y	Y	N	NR
Miguel, et al.	ipilimumab	CUA	Y	Y	Y	Y	fitted parametric curves
Lertjanyakun, et al.	fulvestrant 250mg fulvestrant 500mg toremifene exemestane+everolimus	CUA	Y	Y	Y	Y	fitted parametric curves
Wu, et al.	standard chemotherapy	CUA	Y	Y	Y	N	fitted parametric curves
Lotan, et al.	traditional approaches radical cystectomy radical cystectomy+neoadjuvant chemotherapy	CEA	Y	Y	N	N	NR
Nixon, et al.	standard of care monitoring	CUA	Y	Y	Y	N	NR
Gharaibeh, et al.	gemcitabine, cisplatin+gemcitabine, oxaliplatin+gemcitabine, capecitabine+gemcitabine, nab-paclitaxel+gemcitabine	CUA	Y	Y	Y	N	fitted parametric curves
Kimura, et al.	gefitinib erlotinib	CEA	N	N	N	N	NR
Ball, et al.	chemotherapy	CUA	Y	Y	Y	N	fitted parametric curves
James, et al.	standard of care	CUA	N	N	Y	Y	fitted parametric curves
Uyl-de et al.	best supportive care	CUA	N	N	Y	N	No extrapolation
Li, et al.	no diagnostic staging laparoscopy	CUA	Y	Y	Y	N	NR

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Leung, et al.	trastuzumab+docetaxel	CUA	Y	Y	Y	N	fitted parametric curves
Kimura, et al.	paclitaxel irinotecan	CEA	N	N	N	N	NR
Mujica-Mota, et al.	placebo (octreotide - identified as potential comparator, but study removed during screening) placebo	CUA	Y	Y	Y	Y	fitted parametric curves
Harty, et al.	FOLFIRI	CUA	Y	Y	Y	N	fitted parametric curves
Elsisi et al.	best supportive care	CUA	Y	Y	Y	N	NR
Bolagnos-Diaz et al.	chemotherapy	CUA	Y	Y	Y	N	No extrapolation
Li, et al.	chemotherapy	CUA	Y	Y	Y	N	fitted parametric curves
Garrison, et al.	trastuzumab+chemotherapy	CUA	Y	Y	Y	N	fitted parametric curves
Redig et al.	targeted therapies (medium) targeted therapies (late)	CEA	N	N	N	N	NR
Hamilton, et al.	systemic chemotherapy	CUA	Y	Y	Y	N	NR
Pruis, et al.	interferon-α	CUA	Y	Y	Y	Y	fitted parametric curves
Ondhia, et al.	docetaxel nivolumab	CUA	Y	Y	Y	Y	fitted parametric curves

Author	Comparator(s)	Type of model	Sensitivity analyses	DSA?	PSA?	Scenario analyses	Survival extrapolation method
Insinga, et al.	chemotherapy pembrolizumab monotherapy	CUA	Y	Y	Y	Y	fitted parametric curves
Quinn, et al.	no biopsy	CUA	Y	Y	Y	N	NR
Raldow et al.	long-course chemoradiation	CUA	Y	Y	Y	Y	NR

Abbreviations: CEA, cost-effectiveness analysis; CUA, cost-utility analysis; DSA, deterministic sensitivity analysis; HR, hazard ratio; N, no; NR, not reported; PSA, probabilistic sensitivity analysis; Y, yes

Table caption: list of all articles included in this study

Table A3. List of included studies CONT'D

Author	Test of proportional hazards	AIC	BIC	CEACs	Crossover adjustment	Funding source	Modeling software
Asukai, et al.	N	NR	NR	Y	N	Industry	Excel
Sher, et al.	N	NR	NR	Y	N	NR	TreeAge
Retel, et al.	N	NR	NR	Y	N	Public grant	Excel
Marino, et al.	N	NR	NR	Y	N	Public grant	NR
Cunio et al.	N	NA	NA	N	N	Industry	NR
Iannazzo, et al.	Y	NA	NA	Y	N	Industry	TreeAge
Xie, et al.	N	NR	NR	Y	N	Industry	Excel
De et al.	N	NA	NA	N	N	Public grant	NR
Casciano, et al.	N	NR	NR	N	N	Industry	Excel
Hannouf, et al.	N	NR	NR	Y	N	Public grant	TreeAge
Dranitsaris, et al.	N	NR	NR	N	N	No funding	TreeAge
Delea, et al.	N	NR	NR	Y	Y	Industry	Excel
Hornberger, et al.	N	NA	NA	Y	N	Industry	NR
Mihajlovic, et al.	N	Y	NR	Y	Y	NR	R
Brown, et al.	N	NR	NR	Y	N	Public grant	Excel
Vuong, et al.	N	NR	NR	N	N	NR	SAS
Alba, et al.	N	NR	NR	N	N	Industry	NR
Zeng, et al.	N	NR	NR	Y	N	Public grant	TreeAge
Lawrence, et al.	N	NR	NR	Y	N	Industry	Excel
Hoyle, et al.	N	NR	NR	Y	Y	Public grant	Excel

Author	Test of proportional hazards	AIC	BIC	CEACs	Crossover adjustment	Funding source	Modeling software
Kilonzo, et al.	N	NR	NR	N	Y	Public grant	NR
Amdahl, et al.	N	NR	NR	Y	N	Industry	Excel
Gong et al.	N	NR	NR	Y	N	No funding	NR
Bentley, et al.	N	NA	NA	N	N	Industry	TreeAge
Hannouf, et al.	N	NR	NR	Y	N	Public grant	TreeAge
Zheng, et al.	N	NR	NR	N	N	Industry	NR
Lee, et al.	N	Y	N	N	N	Industry	NR
Hong, et al.	N	NA	NA	Y	N	NR	TreeAge
Roberts, et al.	N	NR	NR	Y	N	NR	TreeAge
Wen, et al.	N	NR	NR	Y	N	No funding	TreeAge
Goldstein, et al.	N	Y	Y	Y	N	No funding	C++
Li, et al.	N	NR	NR	N	N	Industry	Excel
Kumar, et al.	N	Y	N	N	N	Industry	SAS
Pennington et al.	N	Y	NR	Y	N	Industry	NR
Lin, et al.	N	N	N	N	N	Public grant	SAS
Zhou, et al.	N	Nr	NR	N	N	No funding	TreeAge
Balcik, et al.	N	NR	NR	N	N	Public grant	NR
Zhou, et al.	N	NR	NR	N	N	No funding	TreeAge
Nishie, et al.	N	NR	NR	Y	N	Industry	Excel
Raphael et al.	N	NR	NR	Y	N	No funding	TreeAge

Author	Test of proportional hazards	AIC	BIC	CEACs	Crossover adjustment	Funding source	Modeling software
Lu, et al.	N	NR	NR	Y	N	Public grant	R
Shiroiwa, et al.	N	NR	NR	Y	N	Public grant	SAS
Bongers, et al.	N	NR	NR	N	N	Public grant	Excel
Parikh, et al.	N	NR	NR	Y	N	Public grant	TreeAge
Doble, et al.	N	NR	NR	N	N	Public grant	Excel
Zheng, et al.	N	NR	NR	Y	N	No funding	TreeAge
Huang, et al.	N	NR	NR	Y	Y	Industry	Excel
Coyle, et al.	N	NR	NR	Y	N	No funding	NR
Gharaibeh, et al.	N	NR	NR	Y	N	No funding	Excel
Keller, et al.	N	NA	NA	Y	N	Public grant	TreeAge
Miguel, et al.	N	Y	Y	Y	N	Industry	Excel

Author	Test of proportional hazards	AIC	BIC	CEACs	Crossover adjustment	Funding source	Modeling software
Lertjanyakun, et al.	N	Y	Y	Y	N	Public grant	Excel
Wu, et al.	N	N	N	Y	N	Public grant	NR
Lotan, et al.	N	NR	NR	N	N	Public grant	TreeAge
Nixon, et al.	N	NR	NR	Y	N	NR	NR
Gharaibeh, et al.	Y	NR	NR	Y	N	No funding	Excel
Kimura, et al.	N	NR	NR	N	N	No funding	NR
Ball, et al.	N	Y	Y	Y	N	No funding	Excel
James, et al.	N	NR	NR	N	N	Mix	NR
Uyl-de et al.	N	NR	NR	Y	N	Industry	NR
Li, et al.	N	NR	NR	Y	N	Public grant	TreeAge
Leung, et al.	N	NR	NR	Y	N	No funding	TreeAge
Kimura, et al.	N	NR	NR	N	N	No funding	NR
Mujica-Mota, et al.	N	NR	NR	N	Y	Public grant	Excel
Harty, et al.	Y	N	N	Y	N	Industry	Excel
Elsisi et al.	N	NR	NR	N	N	No funding	Excel
Bolagnos-Diaz et al.	N	NR	NR	Y	N	Industry	TreeAge
Li, et al.	N	N	N	Y	N	No funding	TreeAge

Author	Test of proportional hazards	AIC	BIC	CEACs	Crossover adjustment	Funding source	Modeling software
Garrison, et al.	Y	Y	Y	Y	N	Industry	NR
Redig et al.	N	NR	NR	N	N	No funding	Stata
Hamilton, et al.	N	NR	NR	N	N	No funding	TreeAge
Pruis, et al.	N	Y	N	N	N	No funding	Excel
Ondhia, et al.	Y	Y	Y	Y	Y	Industry	Excel
Insinga, et al.	Y	Y	Y	Y	N	Industry	NR
Quinn, et al.	N	NR	NR	N	N	No funding	TreeAge
Raldow et al.	N	NR	NR	N	N	Public grant	TreeAge

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CEACs, cost-effectiveness acceptability curves; N, no; NR, not reported; Y, yes

Table caption: list of all articles included in this study

Chapter 3: Appraisals by Health Technology Assessment Agencies of Economic Evaluations Submitted as Part of Reimbursement Dossiers for Oncology Treatments: Evidence from Canada, the UK, and Australia

1. Introduction

Publicly funded healthcare systems in Canada, the United Kingdom (UK), and Australia use a health technology assessment (HTA) framework to inform drug reimbursement decision-making. In order for a new medication to be publicly reimbursed in these countries, pharmaceutical companies are required to submit a reimbursement dossier which includes, at minimum, the clinical data used for regulatory approval, as well as a model-based economic evaluation to demonstrate the value for money of this new therapy in a given therapeutic area. Following a critical review of the manufacturers' clinical and economic evidence by the HTA agencies, the HTA appraisals and funding recommendations associated with these products are publicly posted on the HTA agency websites.

It has been shown that drug funding recommendations by HTA agencies may differ, due to differences in political priorities [1], agency mandates [1], processes and procedures [2,3], or healthcare systems [2]. However, the level of reporting and appraisal by HTA agencies of economic models submitted by manufacturers for reimbursement appears to not have been previously investigated. This is important for several reasons. First, while the economic guidelines that manufacturers are required to follow for drug submissions to each HTA agency are relatively detailed [4–6], there are no explicit guidelines that HTA bodies are required to follow in the reporting of their economic appraisals of manufacturers' reimbursement submissions. Secondly, physicians, patients, or patient associations, as well as the general public, rely on the public information provided by these HTA agencies to understand the rationale behind the funding recommendations made by the HTA agencies. Finally, in Canada and to a lesser extent, the UK, the appraisal of the economic evidence serves as a basis for price negotiations

between the manufacturers and public plans. To fill a gap in the literature, we sought to answer the question of whether HTA agencies in Canada, the UK, and Australia are consistent in their reporting and appraisal of the economic evaluations submitted by drug manufacturers for the reimbursement of oncology medications. Building on our previous work regarding economic evaluations in oncology in the published literature [7], we hypothesized that consistency would be observed for oncology medications evaluated by the three HTA agencies, due to the same product being assessed based on the same or similar clinical data.

2. Materials and Methods

2.1. Study Data

Publicly posted funding recommendations and appraisal documents for all oncology drug indications issued by Canadian Agency for Drugs and Technologies in Health (CADTH) in 2019 and 2020 were identified. Second, the websites of the National Institute for Health and Care Excellence (NICE) in the UK and Pharmaceutical Benefits Advisory Committee (PBAC) in Australia were searched to identify publicly posted recommendations matching the same drug and indication as those identified from CADTH. Any documents published by NICE and PBAC before the end of 2021 were considered for inclusion. The final study sample comprised oncology drug submissions for which all three HTA agencies had issued a public reimbursement recommendation.

2.2. Data Abstraction

To facilitate comparison between the reporting and critical appraisal of the manufacturers' economic models by the three HTA agencies, a set of commonly required attributes for economic evaluations submitted by drug manufacturers for reimbursement purposes was compiled from CADTH's Guidelines for the Economic Evaluation of Health Technologies [4], the NICE Guide to the Methods of Technology Appraisals [5], and the Guidelines for Preparing a Submission to the PBAC [6]. Due to the focus of our

study on oncology products, recommendations from NICE Decision Support Unit technical support

document 14 (NICE DSU 14) [8] for the conduct and reporting of survival analysis for economic

evaluation were also reviewed, as the NICE DSU 14 is explicitly referenced in CADTH and PBAC

guidelines for economic evaluations of oncology indications. Recommendations from these appraisal

guidelines were cross-referenced with the Consolidated Health Economic Evaluation Reporting

Standards (CHEERS) statement [9] and economic evaluation guidelines from the International Society for

Pharmacoeconomics and Outcomes Research (ISPOR) [10,11], in order to determine a minimum set of

common reporting measures expected to be included in economic models submitted by manufacturers

to CADTH, NICE, and PBAC. Based on this review, 21 common data elements expected to be described in

manufacturer submissions to CADTH, NICE, and PBAC were identified, as shown in Table S2 in the

Supplementary Online Material. Based on these common elements expected to be included in the

economic models and reports submitted for reimbursement by manufacturers, an abstraction sheet was

developed to capture to what extent CADTH, NICE, and PBAC report on characteristics of the

manufacturer economic submissions, in terms of the type of analysis (e.g., cost-utility), utility value

method for cost-utility analyses, model structure, time horizon, indirect comparison, equity issues,

treatment of uncertainty, and validation of results. For interested readers, a glossary of technical terms

is provided in Table S6 in the Supplementary Materials. For survival analyses and extrapolations, HTA

reports were reviewed to document whether the following information was reported: whether a

parametric approach was used, parametric distributions used for extrapolations, goodness-of-fit testing,

testing of the proportional hazards assumption, curve fitting assessment, validation of extrapolations,

treatment effect scenario analyses, justification for any use of external data, whether distributions were

fitted to the tail of Kaplan–Meier curves or entirety of the curves, and whether or not alternative curve-

fitting approaches were examined. A methodological element of interest was considered to be reported

as long as it was mentioned in the HTA reports, irrespective of the quantity of information reported. If

one HTA agency published multiple paragraphs describing a given category, the equity considerations, for example, while another HTA agency published a single sentence, they would both be categorized as having reported on equity. The data was abstracted by one reviewer, and 20% of the abstracted data was checked by a second reviewer.

HTA agencies' methodological criticisms of the economic dossiers submitted by manufacturers were grouped into a set of seven categories: (1) time horizon; (2) treatment benefit; (3) utility values; (4) comparator; (5) subgroups; (6) progression-free survival estimates; (7) overall survival estimates; (8) costs; and (9) extrapolation of survival data. These common thematic categories were adapted from previous studies detailing nine methodological issues described in CADTH economic guidance reports [12] and ten common issues identified by CADTH's economic guidance panel [13]. Incremental costs and QALYs reported in manufacturer submissions and re-calculated by the three HTA agencies were also documented. To facilitate appropriate comparison of incremental cost-effectiveness ratios (ICERs) between the three agencies, which use different currencies, ICERs reported by CADTH, NICE, and PBAC were converted to USD using 2021 purchasing power parity (PPP)-adjusted exchange rates published by the Organisation for Economic Co-operation and Development (OECD) [14]. Finally, the funding recommendation (list, do not list) was also abstracted for each oncology product evaluated by the three agencies in 2020 and 2021. Our initial expectation was that each of the elements described in each of the HTA agency submission guidelines would be summarized and reported in the published assessment reports, since they are required to be submitted by the manufacturer.

2.3. Data Analyses

In order to explore whether reporting of methodological approaches and sources of clinical evidence considered by each agency were similar, potential relationships between variables were assessed where appropriate. Dichotomous differences in the reporting of methods and recommendation outcomes were

assessed using Chi-squared tests. Where appropriate, potential relationships between categorical

variables, which include several modeling characteristics and survival curve extrapolation techniques,

were substantiated through Chi-squared tests. Dichotomous differences in methodological criticisms

observed between the three agencies were assessed using Cochran's Q tests. For statistically significant

Cochran's Q test results, post hoc pairwise McNemar tests were conducted to identify pairwise

relationships. To support the generalizability of our results, we conducted similar analyses for HTAs,

which had recommendations available from only two of the three HTA agencies.

3. Results

3.1. Number of HTA Submissions Reviewed by CADTH between 2019–2020 Matched with

Corresponding HTAs from NICE and PBAC

A total of 83 indications in oncology were identified from the CADTH website between 2019–2020.

Matching these 83 indications with their corresponding public appraisal documents from NICE and the

PBAC, 36 indications were found to have been reviewed by all three agencies, and these 36 indications

(108 individual HTA appraisals) comprised our comparative study sample. Out of the 108 appraisals by

NICE, CADTH, and PBAC, 14 recommendations were published by PBAC and NICE before the CADTH

recommendations (i.e., before 2019), with 7 in 2021.

Of note, we excluded 17 indications (51 individual submission appraisals) that were evaluated by two of

the three agencies, 19 indications (57 individual submission appraisals) that were evaluated by only one

of the agencies, and 11 indications that were listed on the CADTH website, but for which none of the

three agencies (including CADTH) had published a recommendation. Table S1 in the Supplementary

Materials presents the list of indications/products reviewed by the three agencies, two agencies, and

one agency only, as well as those indications for which no recommendations were issued.

3.2. Manufacturer Economic Submissions' Characteristics

Table 1 presents the characteristics of the 108 economic evaluations, which were submitted to CADTH, NICE, and PBAC, as reported by these three HTA agencies in their public documents providing the rationale for the funding decision. Two thirds (67%) of the manufacturer economic submissions utilized a single phase 3 study as the main source of clinical data, and the most frequent therapeutic areas were lung cancer (25%) and leukemia (14%). Approximately two-thirds (64%) of the submissions were related to treatments for late-stage disease (stage IV or metastatic disease).

Table 1. Characteristics of included studies.

Characteristic	n	%
HTA agency (n = 108)		
pCODR	36	33%
NICE	36	33%
PBAC	36	33%
Data source type (n = 108)		
Ph3	79	67%
Ph2 (single arm)	16	15%
Mix of Ph3 and Ph2	4	3%
RWE	0	0%
Mix of Ph2 and RWE	5	6%
Mix of Ph3 and RWE	4	7%
Ph4	0	1%

Characteristic	n	%
Type of cancer studied (n = 108)		
Leukemia	15	14%
Breast	12	11%
Lung	27	25%
Genitourinary	9	8%
Gastrointestinal	12	11%
Lymphoma	6	6%
Skin and melanoma	12	11%
Other	3	3%
Myeloma	3	3%
Gynecology	6	6%
Head and neck	3	3%
Neurological	0	0%
Cancer stage (n = 108)		
Early/stage I	12	11%
Stage II/III	27	25%
Stage IV/metastatic	69	64%

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; RWE, real-world evidence.

3.3. HTA Agency Reporting on Economic Model Characteristics Submitted by Manufacturers

As shown in Table 2, all three HTA agencies were consistent in their reporting of the basic characteristics of the economic models submitted by the manufacturers, in terms of the type of economic analyses, model structure, time horizon, treatment of uncertainty, and use of indirect treatment comparison used by the manufacturers, as these items were reported almost all the time by the three HTA agencies. However, some differences were observed between HTA agencies in the model characteristics submitted by the manufacturer or HTA reporting on some elements. Briefly, all submissions to CADTH and NICE were based on cost-utility techniques, while 17% of PBAC submissions were based on cost-minimization techniques (p = 0.013). Differences were observed between the HTA agencies in terms of reporting the instrument used to derive the utility values for CADTH (44%), NICE (92%), and PBAC (61%) (p = 0.001), reporting on equity issues (p < 0.001), and which types of analyses were conducted to deal with uncertainty (p < 0.001) with NICE reporting this information more frequently than the CADTH and PBAC. Partitioned survival models were used in approximately 70% of the models, and most models used three health states. Indirect comparisons were used in more than half of the submissions.

Table 2. Common economic evaluation attributes reported by HTA agencies (N = 108).

Reported Characteristic	Number of Studies n (%)			p-Value (χ2)
	CADTH	NICE	PBAC	
Type of analysis				
CUA	36 (100%)	36 (100%)	30 (83%)	0.013
CEA	0 (0%)	0 (0%)	0 (0%)	
Other (e.g., CMA)	0 (0%)	0 (0%)	6 (17%)	

Reported Characteristic	Number of Studies n (%)			p-Value (χ2)
	CADTH	NICE	PBAC	
QALYs reported (Y/N)				
Yes	34 (94%)	36 (100%)	30 (83%)	0.023
No	2 (6%)	0 (0%)	6 (17%)	
Utility value method				
EQ5D	15 (42%)	33 (92%)	18 (50%)	0.001
SF36	0 (0%)	0 (0%)	0 (0%)	
HUI	0 (0%)	0 (0%)	0 (0%)	
Other	1 (3%)	1 (3%)	4 (11%)	
Not reported	20 (56%)	2 (6%)	14 (39%)	
Model structure				
Partitioned survival	25 (69%)	25 (69%)	24 (67%)	0.112
Markov	11 (31%)	10 (28%)	6 (17%)	
Not reported	0 (0%)	0 (0%)	6 (17%)	
Decision tree	0 (0%)	0 (0%)	0 (0%)	
Combination (decision tree + Markov)	0 (0%)	1 (3%)	0 (0%)	
Other	0 (0%)	0 (0%)	0 (0%)	
Number of modeled health states				
Three	24 (67%)	29 (81%)	21 (58%)	0.516

Reported Characteristic	Number of Studies n (%)			p-Value (χ2)
	CADTH	NICE	PBAC	
Four	2 (6%)	3 (8%)	4 (11%)	
Five	4 (11%)	2 (6%)	1 (3%)	
Six	1 (3%)	0 (0%)	3 (8%)	
Seven or more	0 (0%)	2 (6%)	0 (0%)	
Not reported	5 (14%)	0 (0%)	7 (19%)	
Time horizon (submitted by manufacturer)				
1–5 years	4 (11%)	0 (0%)	3 (8%)	<0.001
6–10 years	14 (39%)	4 (11%)	16 (44%)	
11–20 years	7 (19%)	10 (28%)	3 (8%)	
21–30 years	3 (8%)	7 (19%)	3 (8%)	
31–40 years	2 (6%)	6 (17%)	3 (8%)	
40+ years	6 (17%)	8 (22%)	1 (3%)	
Not reported	0 (0%)	1 (3%)	7 (19%)	
Indirect treatment comparison (Y/N)				
Yes	20 (56%)	24 (67%)	20 (56%)	0.541
No	16 (44%)	12 (33%)	16 (44%)	
Equity issues reported				
Yes	0 (0%)	15 (42%)	0 (0%)	<0.001

Reported Characteristic	Number of Studies n (%)			*p*-Value (χ2)
	CADTH	NICE	PBAC	
No	36 (100%)	21 (58%)	36 (100%)	
Handling of uncertainty				
Deterministic sensitivity analysis	12 (33%)	33 (92%)	9 (25%)	<0.001
Probabilistic sensitivity analysis	11 (31%)	36 (100%)	4 (11%)	<0.001
Scenario analysis	13 (36%)	36 (100%)	27 (75%)	<0.001
Validation (Y/N)				
Yes	2 (6%)	35 (97%)	0 (0%)	<0.001
No	34 (94%)	1 (3%)	36 (100%)	
Reimbursement recommendation				
Reimburse	28 (78%)	34 (94%)	19 (53%)	<0.001

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; CEA, cost-effectiveness analysis; CMA, cost-minimization analysis; CUA, cost-utility analysis; EQ5D, European Quality of Life 5 dimensions; HUI, health utilities index; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; QALYs, quality-adjusted life-years; SF36, Short Form 36.

3.4. HTA Agency Reporting on Methods Used to Extrapolate Survival Data in Manufacturers' Cost-Effectiveness Models

Important numerical and statistical differences between the three HTA agencies were seen in the reporting of the methods used by manufacturers when analyzing and extrapolating survival data for cost-effectiveness modeling, with NICE reporting more often on the characteristics of the survival extrapolation methods used by manufacturers than CADTH and PBAC. For example, NICE consistently

reported on whether parametric distributions were used (100% of the time), which statistical tests (e.g., AIC, BIC) were used to select the best fitting curves (94% of the time), whether the PH assumption was tested (89%), whether survival curves were fitted jointly or separately (86%), or if the extrapolations were validated (97%). In comparison, CADTH and PBAC discussed whether parametric distributions were used 56% and 78% of the time and rarely reported on the PH assumption (CADTH: 10% and PBAC: 32%). Table 3 presents the details, while Table 4 presents the parametric distributions used for modeling PFS and OS, as reported by the HTA agencies. Compared to NICE, who provided information on which statistical distributions were used, CADTH rarely reported which statistical distribution was used. While the Weibull, exponential, log-logistic, log-normal, and generalized gamma were used by manufacturers to model PFS or OS, no single distribution was reported more than 25% of the time (Table 4).

Table 3. Survival analysis attributes reported by HTA agencies.

Reported Characteristic	Number of Studies n (%)			*p*-Value (χ2)
	CADTH	NICE	PBAC	
Parametric approach				
Yes	20 (56%)	36 (100%)	28 (78%)	<0.001
No	16 (44%)	0 (0%)	8 (22%)	
Standard parametric distributions tested	**N = 20**	**N = 36**	**N = 28**	
Yes	17 (85%)	36 (100%)	21 (75%)	0.008
No	3 (15%)	0 (0%)	7 (25%)	
Curve fitting assessment	**N = 20**	**N = 36**	**N = 28**	
AIC	1 (5%)	2 (6%)	1 (4%)	<0.001

Reported Characteristic	Number of Studies n (%)			*p*-Value (χ2)
	CADTH	NICE	PBAC	
BIC	1 (5%)	0 (0%)	0 (0%)	
Both AIC and BIC	6 (30%)	30 (83%)	8 (29%)	
Other	0 (0%)	2 (6%)	1 (4%)	
Not reported	28 (60%)	2 (6%)	26 (64%)	
PH assumption tested (if appropriate)	**N = 20**	**N = 36**	**N = 28**	
Yes	2 (10%)	32 (89%)	9 (32%)	<0.001
No	18 (90%)	4 (11%)	19 (68%)	
Fitted parametric curves	**N = 20**	**N = 36**	**N = 28**	
Jointly fitted models	1 (5%)	20 (56%)	10 (36%)	
Separately fitted models	0 (0%)	11 (31%)	4 (14%)	<0.001
Not reported	19 (95%)	5 (14%)	14 (50%)	
Validation of extrapolations				
Yes	1 (3%)	35 (97%)	6 (17%)	<0.001
No	35 (97%)	1 (3%)	30 (83%)	
Scenario analyses of treatment effect				
Yes	12 (33%)	19 (53%)	11 (31%)	0.109
No	24 (67%)	17 (47%)	25 (69%)	
Use/source of external data justified				

Reported Characteristic	Number of Studies n (%)			p-Value (χ2)
	CADTH	NICE	PBAC	
Yes	1 (3%)	20 (56%)	4 (11%)	<0.001
No	35 (97%)	16 (44%)	32 (89%)	
Curves fitted to tail of Kaplan–Meier curves only				
Yes	1 (3%)	4 (11%)	3 (8%)	0.389
No	35 (97%)	32 (89%)	33 (92%)	
Alternative curve-fitting approaches examined				
Yes	3 (8%)	9 (25%)	1 (3%)	0.011
No	33 (92%)	27 (75%)	35 (97%)	

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; OS, overall survival; PBAC, Pharmaceutical Benefits Advisory. Committee; PFS, progression-free survival.

Table 4. Parametric distributions selected for survival curve extrapolations.

Selected Parametric Curve Reported	Treatment						Comparator					
	CADTH (n = 20)		NICE (n = 36)		PBAC (n = 28)		CADTH (n = 20)		NICE (n = 36)		PBAC (n = 28)	
	PFS	OS	PFS	OS	PFS	OS	PFS	OS	PFS	OS	PFS	OS
Weibull	5%	10%	22%	17%	11%	14%	0%	5%	22%	14%	11%	14%
Exponential	0%	5%	8%	25%	25%	32%	0%	5%	8%	25%	25%	25%
Log-logistic	0%	5%	17%	19%	7%	11%	0%	0%	17%	19%	11%	14%

Selected Parametric Curve Reported	Treatment						Comparator					
	CADTH (n = 20)		NICE (n = 36)		PBAC (n = 28)		CADTH (n = 20)		NICE (n = 36)		PBAC (n = 28)	
	PFS	OS	PFS	OS	PFS	OS	PFS	OS	PFS	OS	PFS	OS
Log-normal	15%	5%	19%	17%	32%	14%	10%	5%	17%	17%	25%	18%
Gamma	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%
Generalized gamma	0%	0%	14%	6%	11%	4%	0%	5%	11%	3%	11%	4%
Gompertz	5%	0%	8%	6%	0%	7%	0%	0%	8%	6%	0%	7%
Other	0%	0%	0%	3%	0%	0%	0%	0%	0%	3%	0%	0%
Not reported	75%	75%	11%	8%	14%	18%	90%	80%	14%	14%	18%	18%

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; OS, overall survival; PBAC, Pharmaceutical Benefits Advisory. Committee; PFS, progression-free survival.

3.5. HTA Agency Reporting on Methodological Criticisms of Manufacturer Economic Submissions

In general, the three HTA agencies tended to focus on broadly similar areas of criticism, regarding the cost-effectiveness models for a given drug/indication, most often relating to the extrapolation of treatment benefit beyond the trial duration (CADTH: 36%, NICE: 47%, and PBAC: 39%), estimation of PFS (CADTH: 36%, NICE: 61%, and PBAC: 47%), and estimation of OS (CADTH: 53%, NICE: 61%, and PBAC: 44%). Notable differences between HTA agencies include NICE rarely criticizing manufacturers' submitted time horizon (8%), while almost always criticizing extrapolations (69%), and CADTH usually criticizing both the manufacturers' submitted time horizon (44%) and cost assumptions (64%) (Figure 1).

Figure 1. Methodological criticisms. Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; PBAC, Pharmaceutical Benefits Advisory Committee; OS, overall survival; PFS, progression-free survival; NICE. National Institute for Health and Care Excellence.
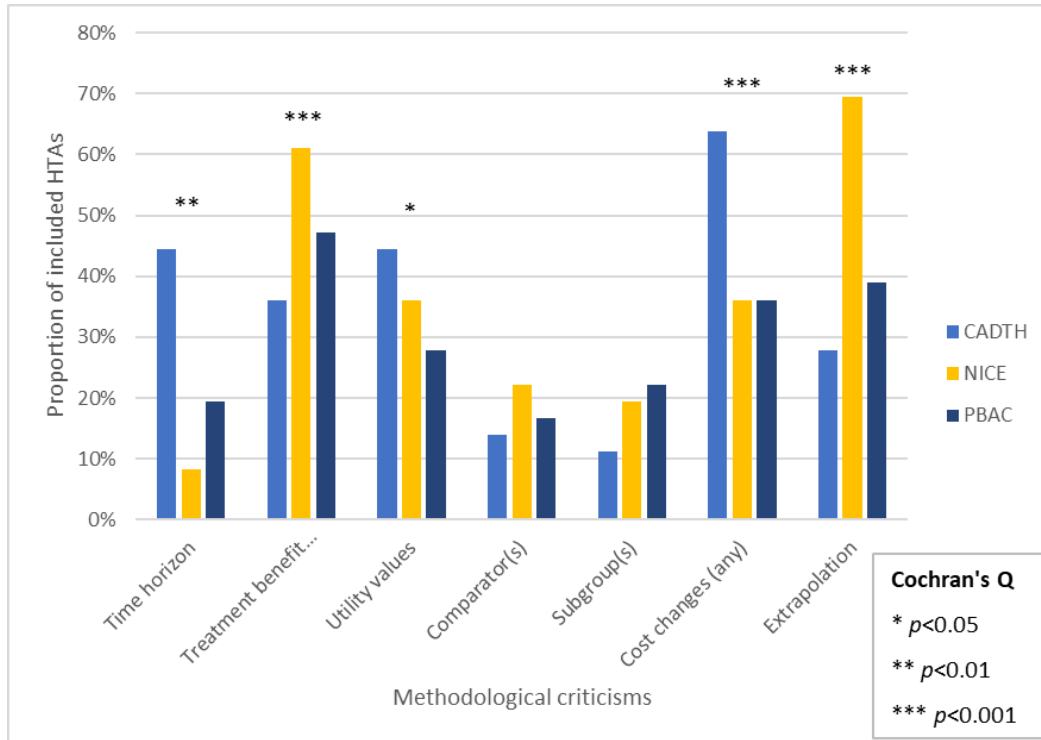
Abbreviations: CADTH, Canadian Agency for Drugs and Technologies in Health; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee

3.6. HTA Agency Reporting on Economic Results, HTA Economic Re-Analyses and Funding Recommendations

Table 5 presents the incremental QALY and incremental cost per QALY gained submitted by the manufacturers and following the re-analyses conducted by the HTA agencies. While all three HTA agencies reported the economic results submitted by the manufacturers, PBAC (50% of the time) and NICE (42%) redacted the QALYs results more often that CADTH (22%). Among those HTAs that reported unredacted QALYs, average incremental QALYs were of broadly similar magnitude across the three

agencies, both for the manufacturer-submitted QALYs (CADTH: 1.30, NICE: 1.17, and PBAC: 1.52) and in

the CADTH and NICE reanalyses of the model results (CADTH: 0.78 and NICE: 0.68) (Table 5). As the

PBAC did not report the reanalysis of QALYs, no values were available for comparison. The average

difference between the manufacturer-submitted and agency reanalyzed QALYs were also similar across

agencies (CADTH: −60.3% and NICE: −58.5%). In terms of incremental cost-effectiveness ratios (ICERs),

the ICERs expressed in the PPP were found to vary across the individual agencies, both in the

manufacturer's submitted estimates (CADTH: USD$110K/QALY, NICE: USD$66K/QALY, and PBAC:

USD$49K/QALY) and agency reanalyses (CADTH: USD$201K/QALY and NICE: USD$113K/QALY). NICE and

CADTH re-analyses almost doubled the ICER submitted by the manufacturer. In terms of

recommendations, 94% of NICE recommendations were positive, 78% were positive for CADTH, and

PBAC issued positive recommendations for 53% of the submissions. Statistical differences in

recommendation status were seen between PBAC and NICE ($p < 0.001$) and PBAC and CADTH ($p = 0.029$). The three agencies issued the same recommendation (either positive or negative) in 39% of the

included HTAs.

Table 5. Comparison of manufacturer and agency-reanalyzed incremental quality-adjusted life-years and

incremental cost-effectiveness ratios (ICERs).

	Incremental QALYs					
HTA Agency	Manufacturer: Base Case	Range		Agency Re-Analysis: Base Case	Range	Average Change
CADTH (n = 32)	1.30	0.13 to 4.34	CADTH (n = 28)	0.78	0.08 to 2.25	−60.3%

Incremental QALYs						
HTA Agency	Manufacturer: Base Case	Range		Agency Re-Analysis: Base Case	Range	Average Change
NICE (n = 21)	1.17	0.07 to 3.44	NICE (n = 15)	0.68	0.07 to 2.75	−58.5%
PBAC (n = 18)	1.52	0.13 to 6.84	N/A	N/A	N/A	N/A

	ICER					
HTA Agency	Manufacturer: Base Case	Range		Agency Re-Analysis	Range	Average Change
CADTH (n = 32)	$109,581	$12,242 to $388,172	CADTH (n = 32)	$200,923	$41,414 to $983,977	183.4%
NICE (n = 27)	$65,778	$6631 to $137,200	NICE (n = 26)	$112,891	$23,744 to $229,381	171.6%
PBAC (n = 23)	$48,665	$18,910 to $129,217	N/A	N/A	N/A	N/A

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; HTA, health technology assessment; ICER, incremental cost-effectiveness ratio; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; PPP, purchasing power parity; QALY, quality-adjusted life-year.

As a partial validation of the representativeness of our results, our supplementary analyses of economic appraisals, conducted by 2 of the 3 agencies (19 indications and 57 individual HTAs), confirmed the results of the main comparative study sample, as the observed frequencies of reporting among this alternative dataset (Tables S3, S4, and S5 in the Supplementary Materials) were broadly similar to those included in the main study.

4. Discussion

We hypothesized that consistency would be observed for the oncology medications evaluated by the three HTA agencies, due to the same product being assessed for the same indication, based on the same or similar clinical data.

4.1. Summary of Findings

We undertook a review of 36 oncology-based economic evaluations submitted by drug manufacturers for reimbursement purposes in Canada over the 2-year period, 2019–2020, which matched with corresponding submissions to the UK and Australia, for which an appraisal and funding recommendation report was publicly available from each HTA agency. Although we hypothesized that consistency of reporting would be observed due to the same product being assessed for the same indication, based on the same or similar clinical data, we found important differences in reporting. While the three HTA agencies consistently reported the baseline characteristics of these economic evaluations, NICE provided more information than CADTH or PBAC when describing the methods used for the extrapolations of survival data, despite the similar requirements for drug manufacturers to follow the same DSU guidelines [8]. Differences were also observed in the HTA agency criticisms of manufacturers' submitted models and extent of the reanalysis undertaken. The level of detail provided by each HTA agency, as a rationale for their appraisal and funding recommendations, was also found to vary substantially. In general, NICE provided extensive documents that comprehensively detailed the clinical, economic, and technical aspects of manufacturer submissions, as well as in-depth assessment notes from the evidence review group (ERG). The appraisals by PBAC and CADTH, while providing relatively extensive review documents, nevertheless did not provide the same level of detail and transparency as NICE. Both NICE and the PBAC, in contrast to CADTH, often redacted key outcomes in their HTAs (e.g., QALYs and ICERs). This discrepancy seems notable, given that all three agencies are publicly funded and

assess the same drug products and indications using the same or very similar economic model. While each agency may approach their respective HTA process with a similar degree of rigor, it seems that the agencies have pursued different approaches in the quantity of reporting that they make available to the public.

This situation could be explained by different levels of resources assigned to the review of the economic evidence submitted by the manufacturers. For example, CADTH assigns a panel of external reviewers to critically appraise the information submitted by the manufacturer, while NICE utilizes a number of academic centers of excellence, the individual members of which may differ for each reimbursement submission. The PBAC is comprised of an independent statutory body of clinical and economic experts appointed by the Australian government. While criticisms of model assumptions varied across HTA agencies, the re-analyses conducted by CADTH and NICE to address model limitations nearly doubled the ICERs on average. It was difficult to assess ICERs re-analyzed by the PBAC, as these ICERs were presented as ranges with no point estimate, and often the range was quite wide for both manufacturer-submitted and PBAC re-analyzed ICERs. However, the percentage of positive recommendations were lower for PBAC than CADTH and NICE, which might be at least partially explained through different approaches to reimbursement (drug reimbursement in Australia does not include price negotiation, and the PBAC is instead a yes/no decision-making body).

4.2. Previous Studies

It is difficult to compare our study with the previous literature for several reasons. First, previous studies have examined jurisdictional differences across the published HTAs, focusing on the factors that influence HTA reimbursement recommendations from HTA agencies in Australia, Canada, England, and Scotland [1], differences in rates of positive and negative recommendations between Canada and the UK [15], and the impact of differing clinical evidence bases on the HTA recommendations from Australia,

Canada, and the UK [16]. Each of these previous studies has been limited in scope, focusing predominantly on recommendation status across jurisdictions utilizing data that are now considerably dated (2014 or older [1]). Other studies have sought to identify relationships between HTA recommendations across jurisdictions, though most have been published more than 5 years ago [15,17–23], and are focused exclusively on one specific component of HTA submissions (e.g., surrogate endpoints) or one single disease area (e.g., schizophrenia) [18,24,25]. While a large majority of the previous studies have been focused on areas outside of oncology, we previously examined the published oncology literature regarding economic evaluation methods [7]. We showed that greater detail in reporting of survival analysis methods, including extrapolation, statistical analyses, and validation of results, is needed, in order to support greater consistency in decision making. To the authors' knowledge, at the time of writing, no previous studies [1,17,19,21,26–28] have specifically examined how HTA agencies evaluate and report on economic evaluations submitted by manufacturers for reimbursement.

Our current study offers insights into the reporting by three HTA agencies across a broad spectrum of economic evaluation methods, including study characteristics, common economic evaluation attributes, survival analysis, recommendation status, and methodological criticisms. The differences in recommendation status we observed across 36 oncology indications, assessed by CADTH between 2019–2020, matched with corresponding HTAs from NICE and the PBAC, might be at least partially explained through different approaches to reimbursement. Nonetheless, our study does corroborate recent work [12,13], which showed that the time horizon and cost estimates were the most frequently criticized elements of manufacturer submissions to CADTH in the periods 2011–2014 and 2012–2018, respectively. However, our results also suggest that these criticisms may be unique to CADTH, as both NICE and PBAC were found to rarely criticize manufacturer-submitted time horizons (NICE: 8%, PBAC: 19%) or cost estimates (NICE: 36%, PBAC: 36%).

4.3. Limitations

This analysis provides useful insights into methods reporting in HTA appraisal documents, but there are a number of important limitations that should be recognized. First, we conducted our study over a limited time period of 2 years (2019–2020); thus, publication bias may affect our results and conclusions. A different level of detail may have been reported in HTAs before 2019, and recent guideline updates or changes in the HTA review process at CADTH, NICE, and PBAC may impact what and how the HTA results are reported. For example, NICE announced, in 2021, an overhaul of methods to optimize evidence generation and global HTA strategy [29], while the Australian government has recently announced a new strategic agreement and the first independent review of Australia's HTA system [30]. In addition, the Canadian study data included in our analyses was produced through the pCODR assessment pathway, which was specifically designed for review of cancer medications. In late 2020, CADTH announced a new review pathway, in which all submitted drugs, oncology or otherwise, would be reviewed under a single CADTH review procedure that would commence in 2021 [31]. Second, as we focused exclusively on oncology HTAs, caution should be exercised in generalizing our results to other therapeutic areas. Third, as long as an element of interest was mentioned in the HTA reports from CADTH, NICE, and PBAC, irrespective of the quantity of information reported, we considered it as reported. While not a specific objective of our study, the differences we observed in the quantity of information reported from agencies highlights the need for greater consistency in reporting for HTA bodies. We also converted the ICERs reported in the published HTAs across the three agencies, using purchasing power parity (PPP); however, it is difficult to directly compare the ICERs between regions, due to the differences in treatment costs or other relative prices. In addition, the PBAC reports only ranges of ICERs, rather than point estimates, which further inhibits the ability to compare PBAC ICERs with those from other HTA agencies. Finally, we assumed that the reimbursement submissions sent by manufacturers to CADTH, NICE, and the PBAC were similar, which may or may not be true. However, as

demonstrated in our results, the main characteristics of the submissions to these three HTA agencies were observed to be similar.

It should be noted that our sample set included 36 indications in total (108 individual HTAs) that were available from all three HTA agencies out of an overall set of 83 indications (249 individual HTAs). In order to ensure that our comparative sample was representative, we analyzed public recommendations made by two agencies (e.g., N = 19 indications), and the results were consistent with the main analysis (N = 36 indications). From our research, it does appear that CADTH received a slightly higher number of HTA submissions than either NICE or PBAC. One speculative explanation could be that NICE and PBAC are known to be more restrictive in their assessments of submitted dossiers, and this may or may not have prompted a number of pharmaceutical manufacturers to not submit a reimbursement dossier to NICE and/or PBAC for some indications, due to a comparatively lower probability of success. Unfortunately, we cannot demonstrate or substantiate this point using our current research and dataset.

4.4. Future Directions

This study focused on recent recommendations published by three HTA agencies over a 2-year period and is therefore limited in both time horizon and scope. While 2 years was judged to be adequate for assessing the reporting of methods, and previous studies have used similar time scales and/or smaller sample sizes [12,13,32], future studies could be expanded to encompass HTA recommendations from additional years, in order to account for recent changes. Efforts could also be put into expanding comparisons beyond CADTH, NICE, and PBAC, in order to include other countries that have adopted HTA processes, such as South Korea, Taiwan, and more recently, Japan. Our study data from 2019–2020 may provide a useful dataset for future comparisons with oncology drug submissions assessed under CADTH's new procedures.

5. Conclusions

Based on our 2-year sample of oncology HTAs published by CADTH, NICE, and PBAC, the variations in the reporting we observed, especially for technical aspects such as survival analysis, suggest that in addition to the guidelines for HTA submissions, the community of HTA agencies should also have common standards for reporting the results of their assessments, though the information and opinions reported may differ.

Chapter 3 references

1. Allen, N.; Walker, S.R.; Liberti, L.; Salek, S. Health Technology Assessment (HTA) Case Studies: Factors Influencing Divergent HTA Reimbursement Recommendations in Australia, Canada, England, and Scotland. Value Health 2017, 20, 320–328.
2. Vreman, R.A.; Mantel-Teeuwisse, A.K.; Hövels, A.M.; Leufkens, H.G.M.; Goettsch, W.G. Differences in Health Technology Assessment Recommendations Among European Jurisdictions: The Role of Practice Variations. Value Health 2020, 23, 10–16.
3. European Observatory Health Policy Series. In Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies; Busse, R., Klazinga, N., Panteli, D., Quentin, W., Eds.; World Health Organization and OECD (2019); Copenhagen, Denmark, 2019.
4. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada, 4th ed; Canadian Agency for Drugs and Technologies in Health (CADTH): Ottawa, ON, Canada, 2021. Available online: https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada (accessed on 29 October 2021).
5. National Institute for Health and Care Excellence (NICE): Guide to the Methods of Technology Appraisal 2013; National Institute for Health and Care Excellence (NICE): London, UK, 2013. Available online: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 (accessed on 14 November 2021).
6. Department of Health AG. Guidelines for Preparing a Submission to the Pharmaceutical benefits Advisory Committee (PBAC), Version 5.0. September 2016. Available online: https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf (accessed on 22 November 2021).
7. Ball, G.; Levine, M.; Thabane, L.; Tarride, J.-E. Onwards and Upwards: A Systematic Survey of Economic Evaluation Methods in Oncology. PharmacoEconomics-Open 2021, 5, 397–410.
8. Latimer, N.R. Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. Med. Decis. Making 2013, 33, 743–754.
9. Husereau, D.; Drummond, M.; Petrou, S.; Carswell, C.; Moher, D.; Greenberg, D.; Augustovski, F.; Briggs, A.H.; Mauskopf, J.; Loder, E. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ 2013, 346, f1049.
10. Ramsey, S.D.; Willke, R.J.; Glick, H.; Reed, S.D.; Augustovski, F.; Jonsson, B.; Briggs, A.H.; Sullivan, S. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health 2015, 18, 161–172.
11. Caro, J.J.; Briggs, A.H.; Siebert, U.; Kuntz, K.M. Modeling good research practices–overview: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Value 2012, 15, 796–803.
12. Masucci, L.; Beca, J.; Sabharwal, M.; Hoch, J.S. Methodological Issues in Economic Evaluations Submitted to the Pan-Canadian Oncology Drug Review (pCODR). PharmacoEconomic-Open 2017, 1, 255–263.
13. Saluja, R.; Jiao, T.; Koshy, L.; Cheung, M.; Chan, K.K.W. Comparing Manufacturer Submitted and Pan-Canadian Oncology Drug Review Reanalysed Incremental Cost-Effectiveness Ratios for Novel Oncology Drugs. Curr. Oncol. 2021, 28, 606–618.
14. Organization for Economic Cooperation and Development (OECD). Purchasing Power Parities (PPP) Indicator. https://doi.org/10.1787/1290ee5a-en (accessed on 22 February 2022).

15. Chabot, I.; Rocchi, A. Oncology drug health technology assessment recommendations: Canadian versus UK experiences. ClinicoEconomics Outcomes Res. 2014, 6, 357–367.
16. Spinner, D.S.; Birt, J.; Walter, J.W.; Bowman, L.; Mauskopf, J.; Drummond, M.F.; Copley-Merriman, C. Do different clinical evidence bases lead to discordant health-technology assessment decisions? An in-depth case series across three jurisdictions. ClinicoEconomics Outcomes Res. 2013, 5, 69–85.
17. Lexchin, J.; Mintzes, B. Medicine reimbursement recommendations in Canada, Australia, and Scotland. Am. J. Manag. Care 2008, 14, 581–588.
18. Clement, F.M.; Harris, A.; Li, J.J.; Yong, K.; Lee, K.M.; Manns, B.J. Using effectiveness and cost-effectiveness to make drug coverage decisions: A comparison of Britain, Australia, and Canada. JAMA 2009, 302, 1437–1443.
19. Levy, A.R.; Mitton, C.; Johnston, K.M.; Harrigan, B.; Briggs, A.H. International comparison of comparative effectiveness research in five jurisdictions: Insights for the US. PharmacoEconomics 2010, 28, 813–830.
20. Rocchi, A.; Miller, E.; Hopkins, R.B.; Goeree, R. Common Drug Review recommendations: An evidence base for expectations? PharmacoEconomics 2012, 30, 229–246.
21. Samjoo, I.A.; GDT. Comparison of Cancer Therapy Reimbursement Decisions Made in Canada to Australia, Sweden and the United Kingdom. Value Health 2013, 16, A153.
22. Bae, G.; Bae, E.Y.; Bae, S. Same drugs, valued aring comparators and methods used in reimbursement recommendations in Australia, Canada, and Korea. Health Policy 2015, 119, 577–587.
23. Odom, D.M.L.; Sherif, B.; Nelson, L.; McSorley, D. Methods to assess the association of patient-reported outcomes and clinical endpoints, In Proceedings of the 20th Annual International Meeting of the International Society for Pharmacoeconomics and Outcomes Research, Philadelphia 18 May 2015.
24. Grigore, B.; Ciani, O.; Dams, F.; Federici, C.; de Groot, S.; Mollenkamp, M.; Rabbe, S.; Shatrov, K.; Zemplenyi, A.; Taylor, R.S. Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. PharmacoEconomics 2020, 38, 1055–1070.
25. Jin, H.; Tappenden, P.; Robinson, S.; Achilla, E.; Aceituno, D.; Byford, S. Systematic review of the methods of health economic models assessing antipsychotic medication for schizophrenia. PLoS ONE 2020, 15, e0234996.
26. Lung, T.W.; Muhunthan, J.; Laba, T.L.; Shiell, A.; Milat, A.; Jan, S. Making guidelines for economic evaluations relevant to public health in Australia. Aust. N. Z. J. Public Health 2017, 41, 115–117.
27. Smith, M.K.; Khan, O.K.; Yip, S.; Tang, P.A. Comparison between Canadian and British oncology drug review recommendations and their impact on patient access. J. Clin. Oncol. 2017, 35, 6537.
28. Gibson, S.A.M.; Beale, R.; Wickstead, M.; Griffiths, M. Parallel scientific advice from NICE and CADTH: Would one submission fit all? Value Health 2019, 22, S501.
29. National Institute for Health and Care Excellence (NICE). NICE Launches Ambitious Strategy to Provide Quicker Access to New Treatments and Innovations. 2021. Available online: https://www.nice.org.uk/news/article/nice-launches-ambitious-strategy-to-provide-quicker-access-to-new-treatments-and-innovations (accessed on 6 December 2021).
30. Australian Department of Health and Aged Care. Landmark New Medicines Agreements to Bring Significant Benefits for Australian Patients. Media Release 2021; 7 September 2021. Available online: https://www.health.gov.au/ministers/the-hon-greg-hunt-mp/media/landmark-new-medicines-agreements-to-bring-significant-benefits-for-australian-patients (accessed on 17 December 2021).

31. Canadian Agency for Drugs and Technologies in Health (CADTH). CADTH Pharmaceutical Reviews Update—Issue 17. Reimbursement Reviews 2020. Available online: https://cadth.ca/cadth-pharmaceutical-reviews-update-issue-17 (accessed on 28 October 2021).
32. Mathes, T.; Jacobs, E.; Morfeld, J.-C.; Pieper, D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Serv. Res. 2013, 13, 371.

Chapter 3 supplementary materials

Table S1. 83 publicly available therapeutic indications and availability of corresponding HTA documents from CADTH, NICE, and the PBAC

Drug (trade name)	Drug (generic name)	Product / indication assessed by:				Disease area
		0 agencies	1 agency	2 agencies	3 agencies	
		(n=11)	(n=19)	(n=17)	(n=36)	
Adcetris	brentuximab vedotin				***	Peripheral t-cell lymphoma
Kisqali	ribociclib		*			Advanced or metastatic breast cancer
Tecentriq & Avastin	atezolizumab & bevacizumab				***	Hepatocellular Carcinoma
Zejula	niraparib				***	Ovarian Cancer
Keytruda	pembrolizumab				***	Head and Neck Squamous cell Carcinoma
Tecentriq & Avastin	atezolizumab & bevacizumab				***	Non-Squamous Non-Small Cell Lung Cancer
Xospata	gilteritinib			**		Acute Myeloid Leukemia
Calquence	acalabrutinib			**		Chronic Lymphocytic Leukemia
Odomzo	sonidegib			**		Basal Cell Carcinoma
Venclexta	venetoclax				***	Chronic Lymphocytic Leukemia
Cabometyx	cabozantinib		*			Hepatocellular Carcinoma
Erleada	apalutamide		*			Metastatic Castration-Sensitive Prostate Cancer

Brand	Generic					Indication
Kisqali	ribociclib with fulvestrant				***	advanced or metastatic breast cancer
Nubeqa	darolutamide				***	Non-Metastatic Castration Resistant Prostate Cancer
Daurismo	glasdegib		*			Acute Myeloid Leukemia
Blincyto	blinatumomab				***	Minimal Residual Disease-Positive B-Cell Precursor Acute Lymphoblastic Leukemia
Calquence	acalabrutinib				***	Chronic Lymphocytic Leukemia (previously untreated)
Xtandi	enzalutamide			**		Metastatic Castration-Sensitive Prostate Cancer
Adcetris	brentuximab vedotin		*			Stage IV Hodgkin lymphoma
Keytruda	pembrolizumab			**		Renal Cell Carcinoma
Mylotarg	gemtuzumab ozogamicin				***	Acute Myeloid Leukemia
Rydapt	midostaurin		*			Systemic Mastocytosis
Adcetris	brentuximab vedotin				***	Primary Cutaneous Anaplastic Large Cell Lymphoma or CD30-Expressing Mycosis Fungoides
Rozlytrek	entrectinib				***	ROS1-positive Non-Small Cell Lung Cancer
Lonsurf	trifluridine-tipiracil				***	Gastric Cancer
Darzalex	daratumumab		*			Myeloma
Lynparza	olaparib	(red)				BRCA-mutated HER2-negative metastatic breast cancer

Brand	Generic					Indication
Keytruda	pembrolizumab	RED				Metastatic microsatellite instability high or mismatch repair deficient endometrial cancer
Keytruda	pembrolizumab	RED				Metastatic microsatellite instability high or mismatch repair deficient colorectal cancer
Lorbrena	lorlatinib				***	Non-Small Cell Lung Cancer
Inrebic	fedratinib		*			Myelofibrosis
Tecentriq	atezolizumab				***	Small Cell Lung Cancer
TBD	entrectinib		*			Neurotrophic Tyrosine Receptor Kinase Fusion-Positive Solid Tumours
Tecentriq	atezolizumab			**		Advanced or Metastatic Triple-Negative Breast Cancer
Kadcyla	trastuzumab emtansine				***	Early Breast Cancer
Libtayo	cemiplimab				***	Cutaneous Squamous Cell Carcinoma
Keytruda	pembrolizumab			**		Squamous Non-Small Cell Lung Cancer
Lynparza	olaparib				***	Ovarian Cancer
Nerlynx	neratinib				***	Hormone Receptor-Positive Breast Cancer
Lonsurf	trifluridine-tipiracil				***	Metastatic Colorectal Cancer
Idhifa	enasidenib		*			Acute Myeloid Leukemia
Vitrakvi	larotrectinib				***	Neurotrophic Tyrosine Receptor Kinase Locally Advanced or Metastatic Solid Tumours
Zytiga	abiraterone	RED				Prostate Cancer
Atriance	nelarabine	RED				Acute Lymphoblastic Leukemia

Brand	Generic	Red	*	**	***	Indication
Keytruda	pembrolizumab		*			Metastatic Urothelial Carcinoma
Pomalyst	pomalidomide			**		Multiple Myelom
Darzalex	daratumumab		*			Multiple Myeloma
Alunbrig	brigatinib				***	Non-Small Cell Lung Cance
Keytruda	pembrolizumab				***	Melanoma Adjuvant Treatment
Lutathera	lutetium Lu 177 dotatate			**		Gastroenteropancreatic neuroendocrine tumors
Lenvima	lenvatinib				***	Hepatocellular Carcinoma
Bosulif	bosutinib		*			Chronic Myeloid Leukemia
Ninlaro	ixazomib			**		Multiple Myeloma
Verzenio	abemaciclib				***	Advanced or metastatic breast cancer
Imbruvica	ibrutinib		*			Chronic Lymphocytic Leukemia/Small Lymphocytic Leukemia
Imbruvica	ibrutinib	(red)				Waldenstrom's Macroglobulinemia
Revlimid	lenalidomide				***	Multiple Myeloma
Keytruda	pembrolizumab				***	Non-Squamous Non-Small Cell Lung Cancer
Oncaspar	pegaspargase	(red)				Adult Acute Lymphocytic Leukemia
Zirabev	bevacizumab (biosimilar)		*			Metastatic Colorectal Cancer; Non-Small Cell Lung Cancer
Venclexta	venetoclax				***	Chronic Lymphocytic Leukemia
Vizimpro	dacomitinib			**		Non-Small Cell Lung Cancer
Zevalin	ibritumomab	(red)				Non-Hodgkin's Lymphoma

Brand	Generic					Indication
Trazimera	trastuzumab (biosimilar)		*			Breast and Gastric Cancer Biosimilar
Truxima	rituximab (biosimilar)		*			Non-Hodgkin's Lymphoma and Chronic Lymphocytic Leukemia
Xalkori	crizotinib				***	ROS1-positive Non-Small Cell Lung Cancer
Ogivri	trastuzumab (biosimilar)		*			Early Breast Cancer / Metastatic Breast Cancer / Metastatic Gastric Cancer
Demylocan	decitabine	RED				Myelodysplastic Syndromes
Ibrance	palbociclib			**		Advanced or Metastatic Breast Cancer
Imfinzi	durvalumab				***	Non-Small Cell Lung Cancer
Tafinlar & Mekinist	dabrafenib & trametinib				***	Melanoma Adjuvant Treatment
Zytiga	abiraterone	RED				Prostate Cancer
Not reported	rituximab (biosimilar)	RED				Non-Hodgkin's Lymphoma and Chronic Lymphocytic Leukemia
Lenvima	lenvatinib				***	Renal Cell Carcinoma
Blincyto	blinatumomab			**		Philadelphia chromosome positive B-cell precursor acute lymphoblastic leukemia
Folotyn	pralatrexate			**		Peripheral t-cell lymphoma
Unituxin	dinutuximab			**		Neuroblastoma
Xtandi	enzalutamide			**		Non-metastatic castration-resistant prostate cancer
Adcetris	brentuximab Vedotin			**		Hodgkin Lymphoma

					***	Melanoma Adjuvant Therapy
Opdivo	nivolumab				***	Melanoma Adjuvant Therapy
Cabometyx	cabozantinib				***	Renal Cell Carcinoma
Tagrisso	osimertinib				***	Non-Small Cell Lung Cancer
Mvasi	bevacizumab (biosimilar)		*			Metastatic Colorectal Cancer / Non-Small Cell Lung Cancer

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee

Table S2: Common economic evaluation attributes from existing guidelines

Extraction element	Synthesis of common categories	CADTH Guidelines	NICE Guidelines	PBAC Guidelines	AMCP Guidelines	CHEERS Checklist	ISPOR Guidelines I	ISPOR Guidelines II
Perspective	**Perspective**							
Indication	**Indication**							
Target population	**Target population**							
Subgroups	**Subgroup analysis**							
Comparator(s)	**Choice of comparator**							
Time horizon	**Time horizon**							
Type of analysis	**Preferred analytical technique**							
Types of costs	**Costs to be included**							

Model structure	**Modeling**	green	green	green	green	green	green	green
SLR (Y/N)	**Systematic review of evidences**	green	green	green	green	orange	orange	orange
QALYs (Y/N)	**Preference for effectiveness over efficacy**	green	green	green	green	green	orange	orange
	Preferred outcome measure stated	green	green	green	green	green	green	green
Utility value method	**Preferred method for deriving utility values**	green	green	green	green	green	orange	orange
Equity	**Equity issues stated**	green	green	green	green	orange	green	orange
Discount rate	**Discounting costs**	green	green	green	green	green	green	orange
Discount rate	**Discounting outcomes**	green	green	green	green	green	green	orange
PSA, DSA, scenarios	**Sensitivity analysis-methods**	green	green	green	green	green	orange	green
Incremental (Y/N)	**Incremental analysis**	green	green	green	green	green	orange	orange
ICERs/ICURs	**Total costs vs effectiveness**	green	green	green	green	green	orange	orange
Validation (Y/N)	**Portability of results (Generalizability)**	green	green	green	green	green	orange	green
BIA (Y/N)	**Financial impact analysis**	green	green	green	green	orange	orange	orange

Survival analysis methods	Survival analysis							

	Included
	Not included in published academic guidelines

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; DSA, deterministic sensitivity analysis; ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; N, No; N/A, Not applicable; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; PSA, probabilistic sensitivity analysis; QALY, quality-adjusted life-year; SLR, systematic literature review; Y, Yes.

Table S3. Selected results from alternative dataset: main source of clinical data

Characteristic	N	%	Number of studies n (%)
Data source type			
Phase 3 trial	29	55%	29 (55%)
Phase 2 trial (single arm)	8	15%	8 (15%)
Mix (Phase 2 and 3 trials)	1	4%	1 (4%)
RWE	0	15%	0 (15%)
Mix (Phase 2 trial and RWE)	5	9%	5 (9%)
Mix (Phase 3 trial and RWE)	8	2%	8 (2%)
Phase 4	2	0%	2 (0%)

Abbreviations: RWE, real-world evidence

Table S4. Selected results from alternative dataset: economic evaluation attributes

Reported characteristic	Number of studies n (%)		
	CADTH	NICE	PBAC
Type of analysis			
CUA	19 (66%)	11 (100%)	8 (62%)
CEA	8 (28%)	0 (0%)	0 (0%)
Other (e.g. CMA)	2 (3%)	0 (0%)	5 (31%)
Model structure			
Partitioned survival	21 (75%)	9 (82%)	3 (23%)
Markov	4 (14%)	2 (18%)	1 (8%)
Decision tree	0 (0%)	0 (0%)	0 (0%)
Combination (decision tree + Markov)	2 (7%)	0 (0%)	1 (8%)
Other	1 (4%)	0 (0%)	0 (0%)
Not reported	1 (4%)	0 (0%)	8 (62%)
Reimbursement recommendation			
Reimburse	22 (76%)	9 (82%)	12 (0%)
Do not reimburse	7 (24%)	2 (18%)	1 (100%)

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; CEA, cost-effectiveness analysis; CMA, cost minimization analysis; CUA, cost-utility analysis; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee

Table S5. Selected results from alternative dataset: survival analysis

Reported characteristic	Number of studies n (%)		
	CADTH (N=29)	NICE N=11	PBAC (N=13)
Parametric approach			
Yes	16 (55%)	11 (100%)	4 (31%)
No	13 (45%)	0 (0%)	9 (69%)
Curve fitting assessment	**N=16**	**N=11**	**N=4**
AIC	1 (3%)	0 (0%)	0 (0%)
BIC	1 (3%)	0 (0%)	0 (0%)
Both AIC and BIC	6 (21%)	10 (91%)	2 (50%)
Other	0 (0%)	1 (9%)	0 (0%)
Not reported	8 (72%)	0 (0%)	2 (50%)

Abbreviations: CADTH, Canada Agency for Drugs and Technologies in Health; NICE, National Institute for Health and Care Excellence; PBAC, Pharmaceutical Benefits Advisory Committee

Table S6. Glossary of technical terms

Term	Abbreviation	Description
cost-effectiveness acceptability curve	CEAC	A graphical representation of the uncertainty associated with the results of an economic evaluation. It plots for a range of cost effectiveness thresholds against the probability that the new technology /intervention will be cost effective at that threshold. This helps decision-makers understand the uncertainty surrounding the optimal treatment strategy.
cost-effectiveness analysis	CEA	A form of economic evaluation that is best suited to addressing questions of technical efficiency. Comparisons are limited to services or treatment options that produce the same type of benefit, which is valued strictly in one-dimensional, natural units.
cost-minimization analysis	CMA	A special type of cost-effectiveness analysis, which is possible only if it has been determined (or more often assumed) that there are no differences in benefits between the alternate interventions compared and thus the evaluation is based on only the costs of the interventions.
cost-utility analysis	CUA	A variant of cost-effectiveness analysis where the health outcome measure of interest is usually expressed as a quality adjusted life year, a single index that combines length of life and a quality adjustment for less than perfect health (i.e. the utility score).
health technology assessment	HTA	A multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision making in order to promote an equitable, efficient, and high-quality health system.

incremental cost-effectiveness ratio	ICER	The ratio of the difference in costs between an intervention and a specified comparator to the difference in effectiveness between that intervention and the specified comparator. From the results of a cost-effectiveness analysis, an incremental cost-effectiveness ratio can be calculated that depicts the extra cost per unit of outcome obtained, in comparing one treatment option to another.
network meta-analysis	NMA	A technique used in systematic reviews to compare the relative effectiveness of three or more interventions simultaneously that have not been compared in a single randomised trial or a single analysis by combining both direct and indirect effectiveness across a network of studies.
probabilistic sensitivity analysis	PSA	Probabilistic sensitivity analysis represents parameters (inputs) as distributions of possible mean values instead of single point estimates.
quality-adjusted life-year	QALY	A measure of health outcome, which captures both length of life and the quality of life. QALYs are calculated by multiplying the total time (years) in a specific health state (or the number of life years remaining) by the "utility" of those years (measured from zero, representing the worst imaginable health (values less than zero represents health states worse than death), to one, representing perfect health).

Chapter 4: Health Technology Reassessment: Addressing Uncertainty in Economic Evaluations of Oncology Drugs at Time of Reimbursement Using Long-Term Clinical Trial Data

1. Introduction

Drug reimbursement decision making often employs health technology assessments (HTAs) to detail the comparative value for money of one treatment versus another. HTA appraisals and funding recommendations are typically made on the basis of evidence from a single point in time when treatments enter the healthcare system. To enhance early access to novel health technologies, reimbursement decisions are increasingly made when the evidence base to support these decisions is lacking or far from mature [1], and HTA recommendations based on immature data or extrapolated short-term data often include the suggestion to collect additional data [2,3]. In recent years, HTA agencies have been advocating for a lifecycle management approach to health technology adoption and reimbursement decisions. In the United Kingdom, the National Institute for Health and Care Excellence (NICE) recently launched a 5 year strategy to adapt to a rapidly changing health and care landscape, which involves a more dynamic approach to health technology management [4]. The Canadian Agency for Drugs and Technologies in Health (CADTH) has also been messaging the purported benefits of health technology management in which longer-term trial data and real-world evidence (RWE) could be used to re-assess already reimbursed drugs to ensure continued clinical and/or economic benefits are continuing to be realized by patients and in the marketplace. However, reimbursement decisions are rarely reconsidered, even once additional data have been collected. Publication of long-term follow-up data from clinical trials provides the opportunity to reassess decision making under substantially reduced clinical and economic uncertainty.

We present an economic evaluation of pembrolizumab for treatment of patients with advanced melanoma, which was studied in the KEYNOTE-006 phase 3 randomized controlled trial [5–7].

Pembrolizumab was approved by the Food and Drug Administration (FDA) [8] and recommended for reimbursement by HTA agencies including NICE [9] and CADTH [10] using data from the first interim analysis (median duration of follow-up: 7.9 months) [5]. The CADTH recommendation [10] noted uncertainty in the modeling of long-term survival, stating "the original ipilimumab data demonstrated a sustained separation in the tail of the survival curve, a benefit that is yet to be confirmed in the pembrolizumab study". Similarly, the appraisal from NICE [9] noted that "the long-term benefits of pembrolizumab are highly uncertain". The evidence base for the approvals and reimbursement recommendations was based on short-term follow-up data with noted uncertainty, yet attempts to address this uncertainty do not seem to have been undertaken once the initial reimbursement decision was made. Despite the fact that an increasing number of HTAs published by national agencies are based on evidence that is assessed to be "uncertain", there is a paucity of available evidence for addressing this uncertainty in the context of reimbursement decision making.

Following publication of the FDA approval and several HTA agency recommendations [9,10], 5 year results from KEYNOTE-006 (median duration of follow-up: 57.7 months) [7] were published in a post hoc analysis of long-term follow-up data. The availability of this long-term follow-up data provided the opportunity to investigate the degree to which the results from a cost-effectiveness analysis based on interim data, where uncertainty is high, can accurately predict cost-effectiveness results based on longer-term data, where uncertainty is substantially reduced. To address this gap in the existing literature, we sought to determine the impact of using trial data of different maturity (long term versus short term) on survival curve extrapolations, and the impact of these different data on the results of a cost-effectiveness analysis, using the example of a pembrolizumab, which was reimbursed in several jurisdictions based on interim data.

2. Materials and Methods

A partitioned survival model was used to assess the cost-effectiveness of pembrolizumab versus ipilimumab for treatment of advanced melanoma from a US payer perspective over a 20 year time horizon. The results of 20 year survival extrapolations and cost-effectiveness based on interim data (median 7.9 months follow-up) were compared with the cost-effectiveness model based on the long-term follow-up data (median 57.7 months follow-up). Model inputs are presented in Table S1 in the Supplementary Materials.

2.1. Modeling Approach

We developed a three-health-state partitioned survival model (Figure 1) in Microsoft Excel® populated with two sets of data based on the published Kaplan–Meier (KM) curves for progression-free survival (PFS) and overall survival (OS) from the KEYNOTE-006 trial [5,7].

Figure 1. Model structure and health states.
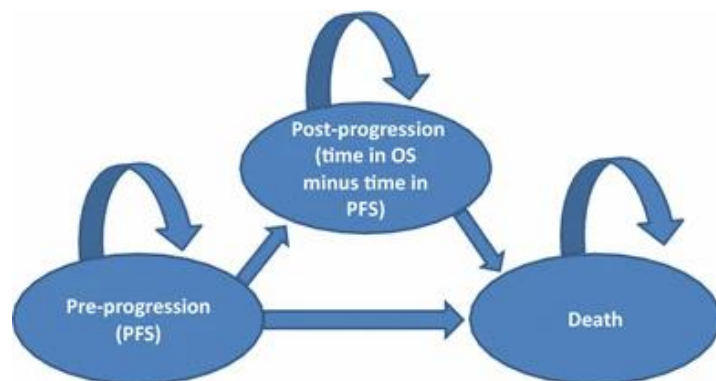
Abbreviations: PFS, progression-free survival; OS, overall survival.

Progression-free survival (Figure 2) and OS (Figure 3) were extrapolated beyond the follow-up period of the trial using standard parametric curve fitting methods over a 20 year time horizon (a description of the extrapolation procedure is provided in Section 2.7).

(A)
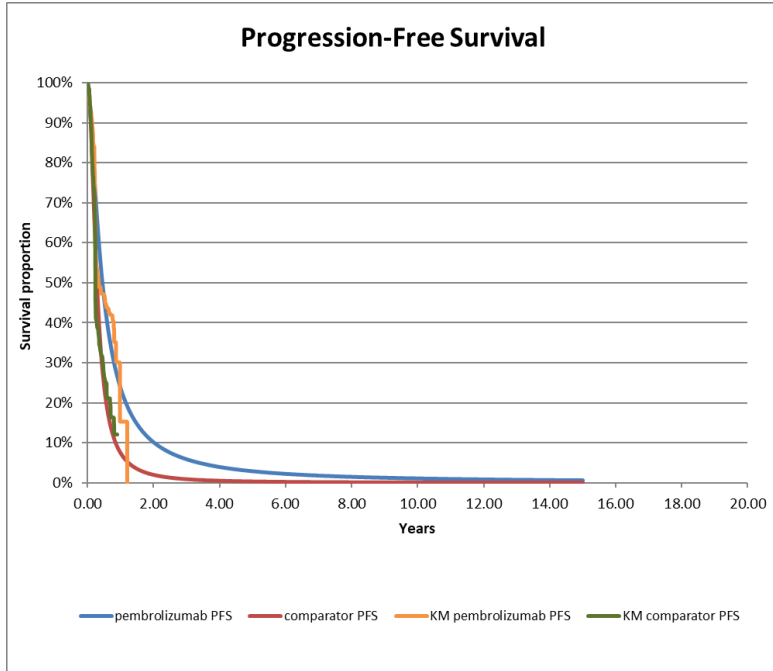
Abbreviation: KM, Kaplan–Meier; PFS, progression-free survival

(B)
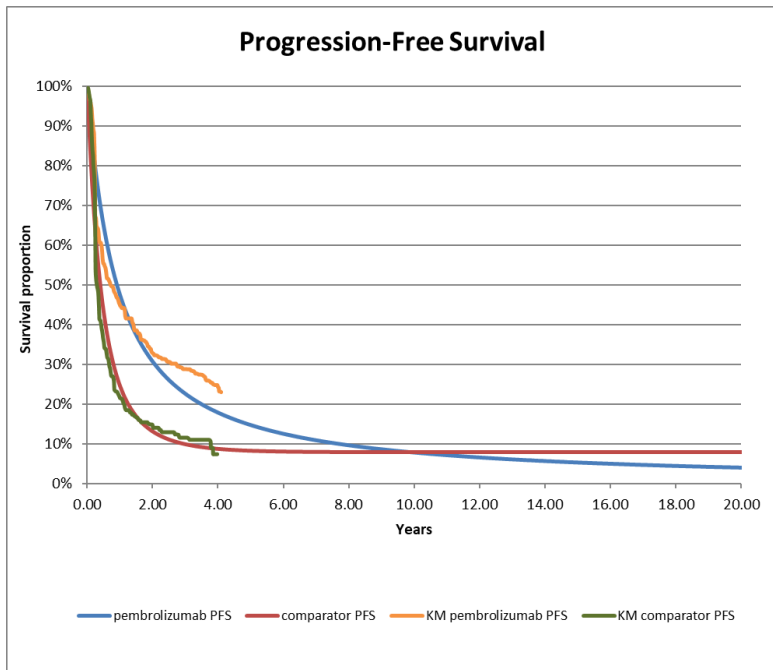
Abbreviation: KM, Kaplan–Meier; PFS, progression-free survival

Figure 2. Survival curve extrapolations based on interim data from KEYNOTE-006. (A)—Progression-free survival based on interim data from KEYNOTE-006; (B)—Progression-free survival based on long-term follow-up data from KEYNOTE-006.
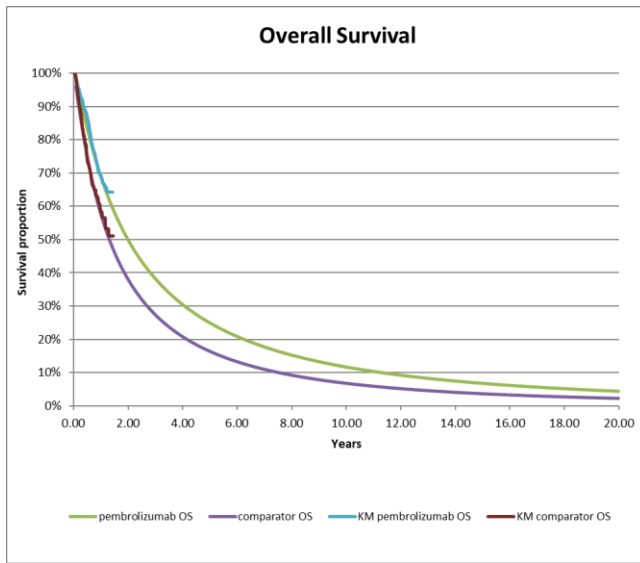
(A)

Abbreviations: KM, Kaplan–Meier; OS, overall survival; PFS, progression-free survival

(B)
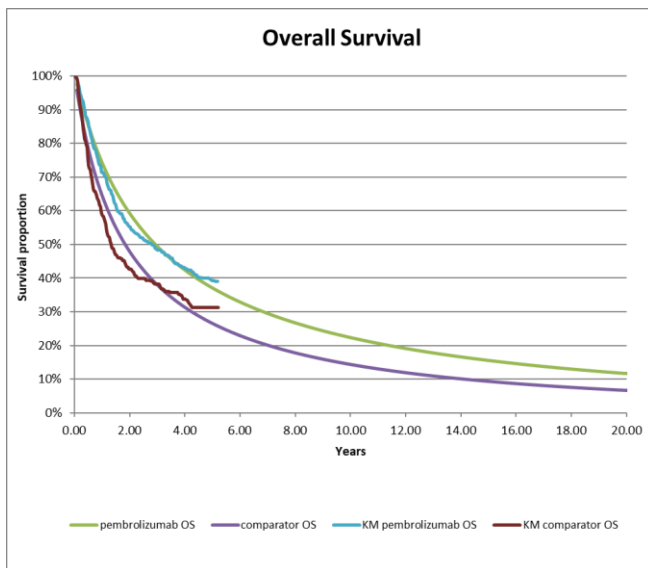
Abbreviations: KM, Kaplan–Meier; OS, overall survival; PFS, progression-free survival

Figure 3. Survival curve extrapolations based on long-term follow-up data from KEYNOTE-006. (A)—Overall survival based on interim data from KEYNOTE-006; (B)—Overall survival based on long-term follow-up data from KEYNOTE-006.

The starting age of the cohort was 63 years [5], and the cycle length was set as monthly. All patients entered the model through the pre-progression health state and could stay in this health state or transition either to the post-progression health state or to death according to transition probabilities calculated from reconstructed KM curves from KEYNOTE-006 [5]. Survival data for PFS and OS were used to determine the distribution of patients in the 'pre-progression' health state over time and the proportion of patients that transition to the 'death' health state for each treatment arm, respectively. The difference between the OS curve and the PFS curve yielded the proportion of patients experiencing progressive disease. The external validity of the modeling approach and survival analysis results were assessed through comparisons with pooled long-term ipilimumab data from patients with advanced melanoma reported by Schadendorf and colleagues [11] and through comparisons with a previously published cost-effectiveness analysis by Wang and colleagues [12] based on the interim data.

2.2. Clinical Inputs

The population modeled in our analyses included adult patients with advanced melanoma who were treated with either pembrolizumab or ipilimumab as depicted in the open-label, multicenter, randomized, controlled phase 3 KEYNOTE-006 trial. KEYNOTE-006 enrolled 834 patients, 556 of which were randomized to pembrolizumab and 278 to ipilimumab. The efficacy was analyzed in the trial according to the intention-to-treat population with OS and PFS as co-primary endpoints.

Adverse event rates of grade 3 or higher were modeled and sourced from the published clinical trial results based on interim data [5] and long-term follow-up data [7] from each of the KEYNOTE-006 treatment arms.

2.3. Regimen and Dosing

Dosing for ipilimumab was 3 mg/kg every 3 weeks up to a maximum of 4 doses, as per the FDA label. For pembrolizumab, the FDA-approved dosing of 2 mg/kg every 3 weeks was implemented in the model for a maximum of 2 years. In accordance with the KEYNOTE-006 trial protocol, a second course of pembrolizumab of up to 12 months was modeled for the proportion of patients who had not progressed by the end of 24 months of pembrolizumab treatment. The dose intensity was assumed to be 100%, and vial sharing was allowed in the base case (medication wastage was not explicitly accounted for).

2.4. Utility Values

Utility values were applied to each health state based on the EuroQoL five-dimension (EQ-5D) preference instrument values collected in KEYNOTE-006 and reported by Wang and colleagues [12]. Disutility adjustments were not made for adverse events, as these events were considered transitory and not anticipated to impact model results.

2.5. Healthcare Resource Utilization

Estimates of healthcare resource utilization associated with patient management in the pre-progression and post-progression health states were derived from the results of a US chart review study [13]. These estimates included oncologist visits, laboratory tests, and scans. Hospitalization costs for management of adverse events (grade 3 or higher) were estimated based on the proportions of patients experiencing grade 3 or higher adverse events reported in KEYNOTE-006 using Drug-Related Group (DRG) codes for gastrointestinal disorders, metabolism and nutrition disorders, and general disorders and administration site condition from the Centers for Medicare and Medicaid Services (CMS) final rule tables [14].

2.6. Costs

Unit costs for pembrolizumab and ipilimumab were based on the average sale price indicated in the 2023 Payment Allowance Limits for Medicare Part B Drugs sourced from CMS [15]. The average patient body weight used to model drug costs was back-calculated from the average dose of each drug reported in a previous cost-effectiveness analysis (98.7 kg for patients receiving ipilimumab, 112.0 kg for patients receiving pembrolizumab) [12]. Drug administration costs (per infusion) were derived from the CMS costs for hospital outpatient services list, and each drug infusion was assumed to incur a single administration cost [16]. The costs associated with patient management in the pre-progression and post-progression health states were estimated based on the results of a US chart review study, as were the costs associated with end of life [13]. The costs for subsequent therapies administered after progression were assumed to be the best supportive care in order to focus on comparative assessments between ipilimumab and pembrolizumab exclusively. No additional drug costs were modeled post-progression.

All costs were reported in 2023 USD, and where necessary, costs derived from previous studies were inflated to 2023 USD using the US consumer price index [17].

2.7. Statistical Analyses

To populate the model, transition probabilities were estimated based on KM curves from KEYNOTE-006 which were digitized using Webplotdigitizer software (Version 4.6), and individual patient-level data were reconstructed according to the Guyot algorithm [18] using the statistical package R Studio. Standard parametric distributions (exponential, log-normal, log-logistic, gamma, Weibull, and Gompertz) were fitted to the reconstructed patient-level data and the statistical fit was assessed based on maximum likelihood estimation. Curve selection was based on the Akaike information criterion (AIC) as well as a visual inspection of the curves to assess the face validity of the fit (Table 1).

Table 1. Akaike information criterion values for parametric curve fitting.

A—interim analysis data						
Parametric Curve Fits	**Weibull**	**Exponential**	**Log-Normal**	**Log-Logistic**	**Gamma**	**Gompertz**
Ipilimumab OS-AIC	933.1	932.7	921.7	927.6	932.0	934.6
Ipilimumab PFS-AIC	969.0	989.9	949.3	940.6	960.0	988.8
Pembrolizumab OS-AIC	818.4	821.0	813.6	816.1	817.6	822.2
Pembrolizumab PFS-AIC	964.9	965.1	947.8	947.4	962.4	965.6
B—long-term follow-up data						
Parametric Curve Fits	**Weibull**	**Exponential**	**Log-Normal**	**Log-logistic**	**Gamma**	**Gompertz**
Ipilimumab OS-AIC	1583.2	1603.6	1542.8	1554.2	1591.6	1542.1
Ipilimumab PFS-AIC	1460.1	1471.7	1374.6	1370.4	1471.0	1404.0
Pembrolizumab OS-AIC	3189.3	3205.0	3140.5	3158.8	3196.2	3150.1
Pembrolizumab PFS-AIC	3277.5	3355.7	3178.0	3202.0	3304.0	3193.9

Abbreviations: AIC, Akaike information criterion; OS, overall survival; PFS, progression-free survival.

Validation of the extrapolated survival curves was undertaken through comparing estimated the life expectancy, hazard ratios, and the number of clinical events with the KEYNOTE-006 data (Table 2).

Table 2. Comparison of reconstructed Kaplan–Meier data versus KEYNOTE-006 data.

A—interim analysis data					
Data Source	**Ipilimumab**		**Pembrolizumab**		**HR (95% CI)**
	Median Survival	**Events (N)**	**Median Survival**	**Events (N)**	

Data Source	Ipilimumab		Pembrolizumab		HR
	Median Survival	Events (N)	Median Survival	Events (N)	(95% CI)
KEYNOTE-006 trial (OS) 2nd interim analysis	not reached	NR	not reached	NR	0.69 (0.52–0.90)
Reconstructed (OS)	15.7	112	24.0	90	0.67 (0.50–0.88)
KEYNOTE-006 trial (PFS) 2nd interim analysis	2.8	NR	4.1	NR	0.58 (0.47–0.72)
Reconstructed (PFS)	3.3	190	5.3	154	0.58 (0.47–0.72)

| B—long-term follow-up data | | | | | |

Data Source	Ipilimumab		Pembrolizumab		HR
	Median Survival	Events (N)	Median Survival	Events (N)	(95% CI)
KEYNOTE-006 trial (OS) long-term follow-up	15.9	172	32.7	324	0.73 (0.61–0.88)
Reconstructed (OS)	22.1	171	35.2	171	0.73 (0.59–0.90)
KEYNOTE-006 trial (PFS) long-term follow-up	3.4	217	8.4	411	0.57 (0.48–0.67)
Reconstructed (PFS)	4.8	222	10.9	402	0.55 (0.47–0.65)

Abbreviations: CI, confidence interval; HR, hazard ratio; N, number; OS, overall survival; PFS, progression-free survival; NR, not reported.

For the cost-effectiveness analysis, the primary outcome of the analysis was calculated as the

incremental cost per quality-adjusted life-year (QALY) gained, and both costs and outcomes were

discounted at 3% annually as recommended by the Institute for Clinical and Economic Review (I.C.E.R.)

[19]. Probabilistic sensitivity analyses were conducted using 1000 Monte Carlo simulations to account for parameter uncertainty. Cost-effectiveness acceptability curves [20] were generated to assess the probability of being cost-effective at varying willingness-to-pay (WTP) thresholds. Scenario analyses were also conducted to address structural uncertainty in the model. These scenario analyses included using the best fitting parametric functions identified in a previous cost-effectiveness analysis, a scenario in which all PFS and OS curves across both treatment arms were fitted with the Gompertz distribution, as was done in the pembrolizumab reimbursement submission to NICE. In addition, to address the difference in the average patient weight between the ipilimumab and pembrolizumab trial arms, an additional scenario analysis was run using the equal average patient weight across both treatment arms. Finally, we varied the proportion of patients remaining progression free who would receive pembrolizumab re-challenge after 2 years according to values cited in HTA agency recommendations.

3. Results

3.1. Survival Analysis of Interim Data [5] versus Long-Term Follow-Up Data [7]

Compared to the KEYNOTE-006 data, the number of events, hazard ratios, and median OS for pembrolizumab from the reconstructed data were found to closely replicate the trial results. However, the reconstructed data of the interim data overestimated the median PFS for both treatment arms and the median OS for ipilimumab (Table 3A). For the 5 year data, hazard ratios and the number of clinical events were consistent with the trial data. However, median survival outcomes were over-estimated in the long-term reconstructed data (Table 3B).

To extrapolate the PFS and OS data over the 20 year time horizon using the interim data, the best-fitting curves for pembrolizumab and ipilimumab PFS and OS were the log-logistic and log normal distributions for both treatment groups, respectively, and log-normal for both treatment arms in the interim data (Table 2A). For KEYNOTE-006 long-term follow-up data, the best-fitting parametric survival curves were

found to be log-logistic for pembrolizumab PFS, log-normal for pembrolizumab OS, Gompertz for

ipilimumab PFS, and log-normal for ipilimumab OS (Table 2B).

3.2. Cost-Effectiveness Analysis of Interim Data [5] versus Long-Term Follow-Up Data [7]

Using the interim data, pembrolizumab generated a total of 3.99 undiscounted life-years (LYs) compared

to 2.85 undiscounted LYs for ipilimumab. The probabilistic ICUR was USD 100,293 per QALY gained

(deterministic ICUR: USD 111,861). Pembrolizumab was found to be cost-effective in 97% of the

simulations at a commonly cited WTP threshold of USD 150,000/QALY gained and 51% of simulations at

a WTP threshold of USD 100,000/QALY gained (Figure 4A). The model was most sensitive to cost-related

inputs, including the discount factor, cost of pembrolizumab, and average patient weight (Figure S1 in

the Supplemental Materials). The results of the scenario analyses are presented in Table S2 in the

Supplemental Materials.

Figure 4. Cost-effectiveness acceptability curves. (A)—cost-effectiveness acceptability curves based on

interim data from KEYNOTE-006; (B)—cost-effectiveness acceptability curves based on long-term follow-

up data from KEYNOTE-006.

(A)

(B)
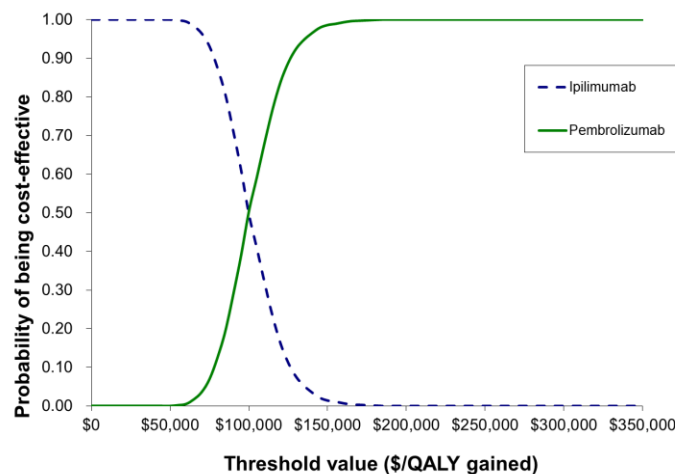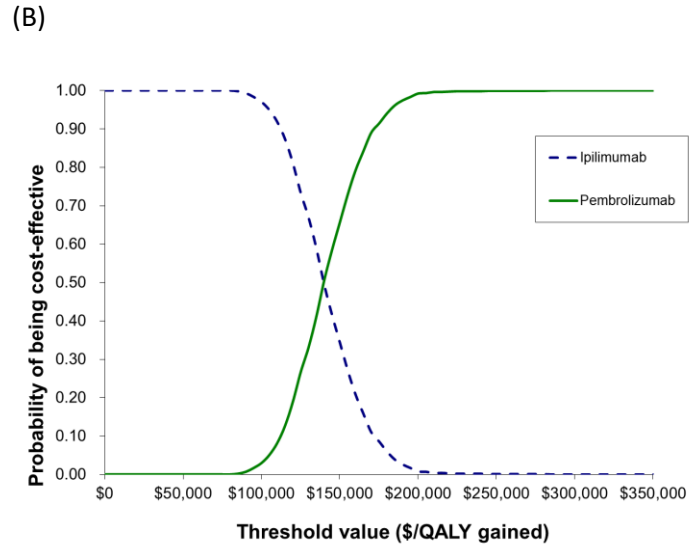
Incorporating the reconstructed long-term follow-up data into the model generated 5.91 undiscounted

LYs for pembrolizumab, and 4.31 undiscounted LYs for ipilimumab. The probabilistic ICUR was estimated

to be USD 139,583 per QALY gained (deterministic ICUR: USD 156,829/QALY gained). At a WTP of USD

150,000/QALY gained, pembrolizumab was cost-effective in 66% of the simulations and was cost-

effective in 3% of simulations at a WTP threshold of USD 100,000/QALY gained (Figure 4B). The ICUR

was found to be sensitive to the clinical inputs, including the shape parameter for both pembrolizumab

and ipilimumab, but was also sensitive to the discount factor (Figure S1 in the Supplemental Materials).

The scenario analysis results are presented in Table S2 in the Supplemental Materials.

4. Discussion

To our knowledge, this is the first reassessment of a drug reimbursed based on interim data and for

which long-term data have been published thereafter. Here, we conducted an economic evaluation of

pembrolizumab versus ipilimumab for advanced melanoma using two reconstructed datasets generated

from the KEYNOTE-006 trial interim analysis and a post hoc long-term follow-up analysis. We used

commonly reported modeling approaches and standard parametric curve fitting techniques to

extrapolate the clinical trial results over a 20 year time horizon and to estimate the cost-effectiveness of pembrolizumab versus ipilimumab from a US payer perspective. Our re-analysis using the long-term follow-up data generated an ICUR that was 42% higher than the ICUR based on the interim trial data. There are a number of reasons that may explain this result. First, the shapes of the KM survival curves are only partially known with interim data, and the shape of the survival curves used for long-term extrapolated outcomes may change as more data are collected, as shown in KEYNOTE-006. This additional data could explain why different best-fitting parametric curves were observed in our analysis based on long-term data compared with the analysis based on interim data and why the long-term survival extrapolations vary between our two models.

As clinicians gain experience with new drug indications, such as pembrolizumab for advanced melanoma, clinical management is likely to incrementally improve, and such improvements might not be fully reflected in KM survival curves based on interim trial data. In KEYNOTE-006, a notable shift outward in the KM PFS curves was reported in the long-term follow-up data (mPFS 8.4 months) compared with the published interim data (mPFS 4.1 months) (Figure 2). This outward shift also impacted our scenario analyses for pembrolizumab as treat-to-progression, showing a much larger percent change from the base case in the analysis based on long-term follow-up data (Table S2 in the Supplementary Materials). While the exact reasons for this outward shift may not be clear, improvements in clinical management could be an explanatory factor. The long-term OS data for ipilimumab presented in KEYNOTE-006 (31% patients alive at 5 years) were also notably higher than the OS data (22% patients alive at 5 years) from a long-term pooled analysis [11] based on 10 previous ipilimumab phase 2 and phase 3 trials (Figure S2 in the Supplementary Materials). Data from the pooled analysis were based on trials published between 2010 and 2013, whereas the interim data from KEYNOTE-006 were published in 2015. The outward-shifted PFS curves presented in the long-term

KEYNOTE-006 data could be at least partially explained by improvements in the clinical management of advanced melanoma patients.

The differences we observed in our interim data- and long-term data-based cost-effectiveness analyses suggest a tradeoff between the need to make recommendations for patient access based on uncertain clinical benefits from shorter term (immature) data, and the delay that would be required to make decisions based on longer-term clinical data in which uncertainty is substantially reduced. While coverage with evidence development has been considered by some HTA bodies as a means of addressing this inherent tradeoff, the availability of long-term data may be sufficiently impactful to warrant an HTA reassessment.

4.1. Previous Studies

A previous cost-effectiveness analysis [12] used the KM curves from the trial for the first 60 weeks, then used parametric extrapolations based on a previous study by Schadendorf and colleagues [11] from 20 to 260 weeks for the ipilimumab arm. For the pembrolizumab arm, the authors used trial data for the first 60 days, then applied a time-varying hazard ratio versus ipilimumab between week 60 and 260 based on a previous study [11], and then used data from a US melanoma registry by Balch and colleagues [21] thereafter. In the pembrolizumab recommendation from NICE for treatment of advanced melanoma, the Evidence Review Group stated that there was a risk of selection bias in using data from the Schadendorf study for extrapolation, as well as limitations in the algorithm used to adjust for patient characteristics and the long-term survival data from Balch and colleagues [21] used to project long-term survival. In contrast, we used trial data only to perform survival curve extrapolations, a simplified approach which nevertheless closely replicated the incremental LY estimates reported in the previous cost-effectiveness study [12] over a 20 year time horizon. Our estimate of an undiscounted incremental gain of 1.15 LYs for treatment of pembrolizumab over ipilimumab was very close to the

number reported in a previous cost-effectiveness study (1.14 LYs). These findings suggest that our survival curve extrapolation approach may be appropriate.

However, our base case probabilistic ICUR estimate based on interim data from KEYNOTE-006 (USD 111,861/QALY gained) was higher than in the previous CEA based on the same interim data (USD 81,091/QALY gained). The most likely explanation for this discrepancy may be found in the differences in how utility values were applied. We applied EQ-5D utility values to the progression-free (0.83) and progressed (0.78) health states, whereas the authors of the previous analysis calculated utility scores based on multiple time-to-death categories: 360 days or more (0.85), 270–360 days (0.78), 180–270 days (0.74), 90–180 days (0.75), 30–90 days (0.69), and under 30 days (0.48) to death. Given that our estimates of incremental undiscounted LYs (1.15 vs. 1.14) and incremental discounted total costs (USD 59,023 vs. USD 63,680) were nearly identical to those of Wang et al. (2015), we conclude that the differences in our results can be largely accounted for by the differences in how utility values were applied in the models.

4.2. Strengths

A number of strengths can be identified in our approach and results. First, the modeling approach and non-clinical input parameters were identical for both the analyses based on interim data and analyses based on long-term data, allowing us to isolate the impact of survival data on the model results. Second, we utilized the same partitioned survival modeling approach reported in HTA agency appraisals that have reviewed KEYNOTE-006 data for advanced melanoma, which helps to support the external validity of our results. We also derived relevant cost inputs from published real-world data [13] and up-to-date CMS costing databases [14–16], as well as a previously published economic evaluation [12] in order to enhance external validity and comparability with previously conducted studies and HTAs. These inputs and methods allowed us to produce an updated estimate of the cost-effectiveness of pembrolizumab

versus ipilimumab for advanced melanoma based on more recent clinical and economic evidence than has been previously published.

Another advantage of our analysis concerns the use of long-term follow-up data in which a large number of patients at risk was retained throughout the vast majority of the long-term follow-up period. In general, it is more challenging to validate results based on small sample sizes, and our use of long-term follow-up data provided a sufficient sample size to have confidence in the results despite a high degree of censoring in the tail of the survival curves.

When modeling chronic conditions (such as cancer), or when treatments have differential effects on mortality, a lifetime horizon is most appropriate. Our survival extrapolations, based on published clinical data from the KEYNOTE-006 trial, indicate that approximately 10% of patients are expected to be alive at the 20-year timepoint (Figure 3B). Using a shorter time horizon would result in important clinical events (e.g. disease progression and death) being missed, and the full costs and clinical benefits would not be captured. While a time horizon of 20 years may seem long, it is consistent with the published literature [12] and health technology appraisal documents from NICE [9] and CADTH [10]. A shorter time horizon would therefore not be appropriate.

In addition, we based our extrapolation approach on the NICE DSU TECHNICAL SUPPORT DOCUMENT 21: Flexible Methods for Survival Analysis [22]. The long-term hazards are expected to follow a simple shape in KEYNOTE-006, exemplified in the simple shape (no kinks and no inflection points) of the long-term OS curve reported by Robert 2019 [7] (5 years follow-up), indicating that the standard parametric curve-fitting approach is appropriate. In addition, although other approaches could be used to model outcomes from KEYNOTE-006, our model follows methodological recommendations from the NICE DSU 21 and is aligned with models from the UK [9], Canada [10], and the US [12].

4.3. Limitations

Our study is not without limitations. First, we did not have access to the trial patient level data and we relied on reconstructing KM curves using digitization techniques. However, our reconstructed survival data closely replicated the hazard ratios and number of clinical events reported in the interim and long-term data from KEYNOTE-006. While the reconstructed data consistently over-estimated median survival outcomes, the survival over-estimation was consistent when using either the interim or long-term data (Table 2), and therefore it should not affect our primary conclusions. Second, several of our model input parameters, including healthcare resource utilization and utility values, were derived from model parameters presented in work by Wang et al. [12]. However, we did not have access to their model, which could explain why the base case ICUR in our cost-effectiveness analysis based on interim data was higher than what has been reported in previous research [12]. In addition, we based our analyses on clinical trial data, and health technology reassessments using RWE could be used as an additional confirmatory source of evidence.

Another limitation of our approach is that during the time period between publication of interim trial data and the subsequent availability of long-term follow-up data, new comparators may have arisen in the clinical environment which may render the results of a re-assessment using the same comparison less clinically relevant. Finally, two different doses of pembrolizumab were studied in the interim analysis of KEYNOTE-006, whereas we modeled only the 2 mg/kg every 3 weeks dose in order to align with the FDA-approved dose. Outcomes from both doses were reported in a combined KM curve for PFS and OS in the long-term pembrolizumab data, but since the two doses studied in the interim analysis had overlapping PFS and OS curves, the impact of this discrepancy is expected to be minimal.

Although HTAs are not commonly used in the US for decision making compared to countries with well-established HTA systems such as the UK or Canada, we replicated a US cost-effectiveness model in the

absence of published cost-effectiveness studies of pembrolizumab for treatment of patients with advanced melanoma in Canada and the UK. Using this US study as a foundation for model inputs, we replicated the US model as closely as possible and therefore we also used a US payer perspective. While the US publication provided a lot of information on the methods and model inputs (e.g., OS extrapolation), some details regarding some model parameters (e.g., utility data) were not available in the US publication, which explains why we were not able to replicate the exact results of the US study. Nonetheless, our primary aim was to replicate the survival reported in the previous study, which is a critical validation step before conducting our re-assessment using long-term data, and we achieved this aim very closely; the number of incremental LYs estimated in our study (1.15) was nearly identical to the value reported in the Wang 2017 study (1.14). Our incremental cost estimate (USD 59,023) also closely matched the previous study (USD 63,680), providing additional validation for our replication approach.

4.4. Future Research

We have demonstrated the impact of long-term clinical trial data on the results of a cost-effectiveness analysis for a single drug in a single therapeutic indication. While the interim analysis from KEYNOTE-006 provided promising preliminary data, the latter half of the KM curves were heavily censored, leading to clinical uncertainty and rendering validation difficult. We addressed this uncertainty by conducting a re-assessment based on long-term follow-up data. However, since the clinical environment of phase 3 trials is highly controlled, our trial-data-based results could be validated in future studies through incorporation of real-world clinical evidence.

Standard parametric curve fitting and extrapolation were used in our analyses. However, a plateau was observed in the long-term follow-up data from both arms of KEYNOTE-006, as well as from long-term pooled data from previous ipilimumab studies [11], implying a potentially curative effect for a small but defined proportion of the advanced melanoma patient population treated with either pembrolizumab

113

or ipilimumab. To account for these observed survival plateaus in the long-term data, a mixture–cure modeling approach [23], in which the patient populations are stratified into 'cured' and 'non-cured' groups to better capture their respective outcomes, could be more methodologically appropriate than standard parametric curve fitting. Future research is encouraged to investigate the impact of using this approach. In addition, our reconstructed data, while very accurately replicating the hazard ratios and number of events reported in KEYNOTE-006, nevertheless overestimated the median PFS and the median OS for the ipilimumab treatment arm (Table 2). This may be at least in part due to the shape of the KM curves which had a steep drop around month 3, resulting in an "s"-shaped curve that is challenging to fit with a single parametric distribution. As a result, fitting spline models with several knots to the reconstructed trial data could potentially be a reasonable alternative methodology to consider in future research.

5. Conclusions

While clinical and economic uncertainty may be reduced with longer-term follow-up data, the results of our analysis suggest that this reduction may come at a cost: decreased cost-effectiveness. Our findings suggest that there may be good reason to consider conducting health technology re-assessments of certain oncology products on the basis of longer-term data availability, especially for those health technology adoption decisions made based on immature clinical data. A lifecycle or health technology management approach could be a practical solution for decision makers to ensure that decision making remains informed by the most appropriate, relevant, and up-to-date evidence. Future research comparing cost-effectiveness models based on interim and final data would be required to generalize the results of our study to other settings.

Chapter 4 references

1. Eckermann, S.; Karnon, J.; Willan, A.R. The Value of Value of Information. PharmacoEconomics 2010, 28, 699–709.
2. 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. 2017, 28, 1738–1750.
3. NHS England. National Cancer Drugs Fund List. 2023. Available online: https://www.england.nhs.uk/wp-content/uploads/2017/04/NationalCDF-List-ver1.265.pdf (accessed on May 23, 2023).
4. National Institutute for Health and Care Excellence. NICE strategy 2021 to 2026. 2021. Available online: https://static.nice.org.uk/NICE%20strategy%202021%20to%202026%20-%20Dynamic,%20Collaborative,%20Excellent.pdf (accessed on Jan 6, 2023).
5. Robert, C.; Schachter, J.; Long, G.V.; Arance, A.; Grob, J.J.; Mortier, L.; Daud, A.; Carlino, M.S.; McNeil, C.; Lotem, M.; et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N. Engl. J. Med. 2015, 372, 2521–2532.
6. Schachter, J.; Ribas, A.; Long, G.V.; Arance, A.; Grob, J.J.; Mortier, L.; Daud, A.; Carlino, M.S.; McNeil, C.; Lotem, M.; et al. Pembrolizumab versus ipilimumab for advanced melanoma: Final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006). Lancet 2017, 390, 1853–1862.
7. Robert, C.; Ribas, A.; Schachter, J.; Arance, A.; Grob, J.-J.; Mortier, L.; Daud, A.; Carlino, M.S.; McNeil, C.M.; Lotem, M.; et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): Post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. Lancet Oncol. 2019, 20, 1239–1251.
8. Merck Press Release. FDA Approves Expanded Indication for Merck's KEYTRUDA® (Pembrolizumab) for the Treatment of Patients with Advanced Melanoma. 2015. Available online: https://www.merck.com/news/fda-approves-expanded-indication-for-mercks-keytruda-pembrolizumab-for-the-treatment-of-patients-with-advanced-melanoma/ (accessed on November 4, 2022).
9. National Institute for Health and Care Excellence. Technology Appraisal Guidance—Pembrolizumab for Advanced Melanoma Not Previously Treated with Ipilimumab; National Institute for Health and Care Excellence: London, UK, 2015.
10. pan-Canadian Oncology Drug Review Economic Guidance Panel. pCODR Final Economic Guidance Report—Pembrolizumab (Keytruda) for Metastatic Melanoma; Canadian Agency for Drugs and Technologies in Health: Canada, 2015.
11. Schadendorf, D.; Hodi, F.S.; Robert, C.; Weber, J.S.; Margolin, K.; Hamid, O.; Patt, D.; Chen, T.-T.; Berman, D.M.; Wolchok, J.D. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. J. Clin. Oncol. 2015, 33, 1889–1894.
12. Wang, J.; Chmielowski, B.; Pellissier, J.; Xu, R.; Stevinson, K.; Liu, F.X. Cost-Effectiveness of Pembrolizumab Versus Ipilimumab in Ipilimumab-Naïve Patients with Advanced Melanoma in the United States. J. Manag. Care Spec. Pharm. 2017, 23, 184–194.
13. Tarhini, A.; Corman, S.L.; Rao, S.; Margolin, K.; Ji, X.; Mehta, S.; Botteman, M.F. Healthcare resource utilization and associated costs in patients with advanced melanoma receiving first-line ipilimumab. J. Cancer Ther. 2015, 6, 833.

14. Center for Medicare & Medicaid Services. FY 2015 Final Rule Tables (Tables 1, 5 and 7); Center for Medicare & Medicaid Services: Baltimore, MD, USA, 2023.
15. Center for Medicare & Medicaid Services. April 2023 Medicare Part B Drug and Biological Average Sales Price; Center for Medicare & Medicaid Services: Baltimore, MD, USA, 2023. Available online: https://www.cms.gov/files/zip/april-2023-asp-pricing-file.zip (accessed on March 18, 2023).
16. Center for Medicare & Medicaid Services. Costs for Hospital Outpatient Services by HCPCS Code for CY 2023; Center for Medicare & Medicaid Services: Baltimore, MD, USA, 2023.
17. U. S. Bureau of Labor Statistics. Consumer Price Index for All Urban Consumers (Not Seasonally Adjusted); U. S. Bureau of Labor Statistics: Washington, DC, USA, 2023. Available online: https://www.bls.gov/regions/mid-atlantic/data/consumerpriceindexhistorical_us_table.htm (accessed on April 12, 2023).
18. Guyot, P.; Ades, A.E.; Ouwens, M.J.N.M.; Welton, N.J. Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. BMC Med. Res. Methodol. 2012, 12, 9.
19. Institute for Clinical and Economic Review. ICER's Reference Case for Economic Evaluations: Principles and Rationale; Institute for Clinical and Economic Review: Boston, MA, USA, 2020.
20. Briggs, A.C.K.; Sculpher, M. Decision Modelling for Health Economic Evaluation; Oxford University Press: Oxford, UK, 2006.
21. Balch, C.M.; Buzaid, A.C.; Soong, S.-J.; Atkins, M.B.; Cascinelli, N.; Coit, D.G.; Fleming, I.D.; Gershenwald, J.E.; Houghton, A.; Kirkwood, J.M.; et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. J. Clin. Oncol. 2001, 19, 3635–3648.
22. Rutherford, M.J.; Lambert, P.C.; Sweeting, M.J.; Pennington, R.; Crowther, M.J.; Abrams, K.R. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. Available online: http://www.nicedsu.org.uk (accessed on February 26, 2023).
23. Lambert, P.C. Modeling of the Cure Fraction in Survival Studies. Stata J. 2007, 7, 351–375.

Chapter 4 supplementary materials

Table S1. Model input parameter values

Input parameter	Base case value	Source
Modeling approach		
Discount Factor (costs)	3%	I.C.E.R. guidelines
Discount Factor (outcomes)	3%	
Starting age of cohort (years)	63	Robert 2015
Average body weight (kg), pembrolizumab	112.0	Wang 2017
Average body weight (kg), ipilimumab	98.7	Wang 2017
Time Horizon (months)	240	Wang 2017
Fitted parametric curve selections		
Pembrolizumab PFS	Log-logistic	Calculated
Pembrolizumab OS	Log-normal	Calculated
Ipilimumab PFS	Loglogistic	Calculated
Ipilimumab OS	Log-normal	Calculated
Unit costs		
Pembrolizumab (per mg)	$55.42	CMS 2023
Ipilimumab (per mg)	$165.45	CMS 2023
Chair time unit cost	$144.39	CMS 2023
Pre-progression patient management (monthly)	$873.37	Tarhini 2015, CPI-adjusted
Post-progression patient management (monthly)	$2,722.63	Tarhini 2015, CPI-adjusted
End of life costs	$6,484.45	Tarhini 2015, CPI-adjusted
Adverse event management costs (grade 3+)		
Fatigue	$4,350.38	CMS 2023
Diarrhea	$7,159.38	CMS 2023
Rash	$4,350.38	CMS 2023
Pruritus	$4,350.38	CMS 2023

Nausea	$7,159.38	CMS 2023
Asthenia	$0.00	Assumption
Arthralgia	$4,350.38	CMS 2023
Vitiligo	$0.00	Assumption
Colitis	$7,159.38	CMS 2023
Hepatitis	$4,350.38	CMS 2023
Hypophysitis	$7,159.38	CMS 2023
Pneumonitis	$4,350.38	CMS 2023

Adverse event rates	Pembrolizumab arm	Ipilimumab arm	
Fatigue	0.4%	1.2%	Robert 2015
Diarrhea	1.1%	3.1%	Robert 2015
Rash	0.0%	0.8%	Robert 2015
Pruritus	0.0%	0.4%	Robert 2015
Nausea	0.0%	0.8%	Robert 2015
Asthenia	0.4%	0.4%	Robert 2015
Arthralgia	0.4%	0.8%	Robert 2015
Vitiligo	0.0%	0.0%	Robert 2015
Colitis	2.5%	7.0%	Robert 2015
Hepatitis	1.8%	0.4%	Robert 2015
Hypophysitis	0.4%	1.6%	Robert 2015
Pneumonitis	0.4%	0.4%	Robert 2015

Utility values		
Progression-free health state	0.83	Wang 2017
Post-progression health state	0.78	Wang 2017

Abbreviations: CMS, Centers for Medicare and Medicaid Services; CPI, consumer price index; I.C.E.R., Institute for Clinical and Economic Review; kg, kilogram; mg, milligram

Table S2. Scenario analyses

A – Scenario analyses based on interim data from KEYNOTE-006

Scenario analysis (based on interim data)	ICUR ($/QALY)	% change from base case
BASE CASE (deterministic)	*$111,861*	-
Best-fitting parametric functions from Wang et al	$109,715	-2%
Parametric distributions from NICE (all Gompertz)	*survival curves cross*	
Average patient weight set equal	$101,914	-10%
Proportion of progression-free pembrolizumab patients receiving re-challenge (CADTH: 25%)	$107,792	-4%
Proportion of progression-free pembrolizumab patients receiving re-challenge (Robert 2019: 19%)	$106,731	-5%
Pembrolizumab as treat-to-progression	$175,300	36%

Abbreviations: CADTH, Canadian Agency for Drugs and Technologies in Health; ICUR, incremental cost-utility ratio; NICE, National Institute for Health and Care Excellence; QALY, quality-adjusted life-year

B – Scenario analyses based on long-term follow-up data from KEYNOTE-006

Scenario analysis (based on long-term data)	ICUR ($/QALY)	Percent change from base case
BASE CASE (deterministic)	*$156,829*	-
Best-fitting parametric functions from Wang et al	*survival curves cross*	
Parametric distributions from NICE (all Gompertz)	$157,932	1%
Average patient weight set equal	$147,272	-6%
Proportion of progression-free pembrolizumab patients receiving re-challenge (CADTH: 25%)	$146,604	-7%
Proportion of progression-free pembrolizumab patients receiving re-challenge (Robert 2019: 19%)	$143,936	-8%
Pembrolizumab as treat-to-progression	$425,298	171%

Abbreviations: CADTH, Canadian Agency for Drugs and Technologies in Health; ICUR, incremental cost-utility ratio; NICE, National Institute for Health and Care Excellence; QALY, quality-adjusted life-year

Figure S1. Deterministic sensitivity analyses

A – One-way sensitivity analyses based on interim data from KEYNOTE-006
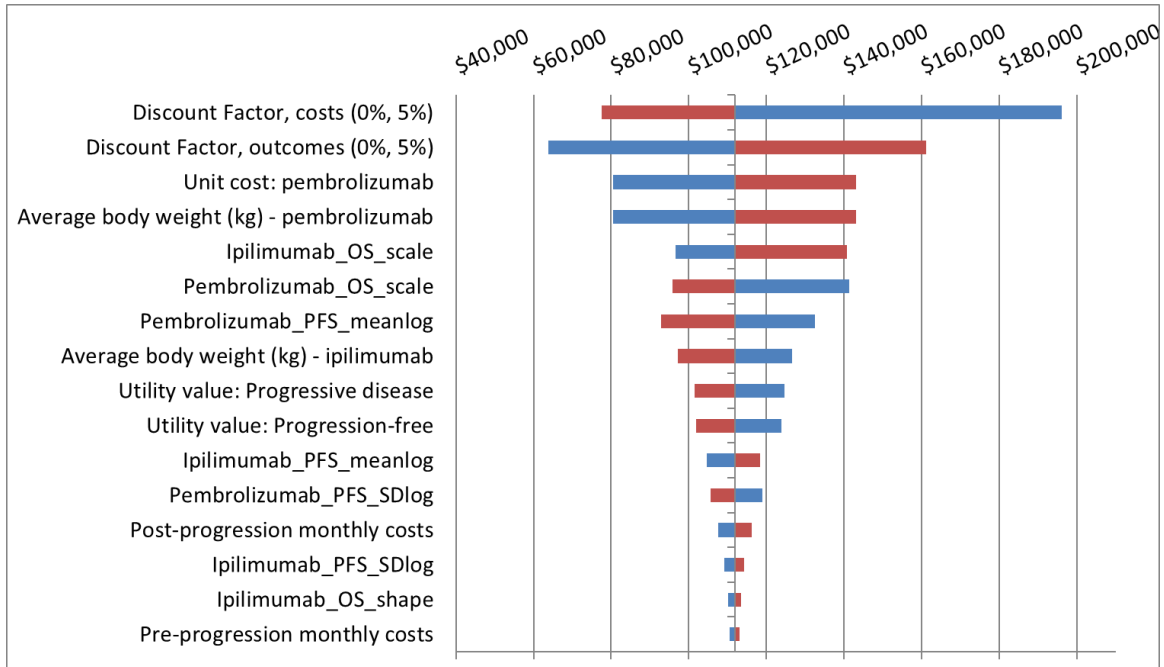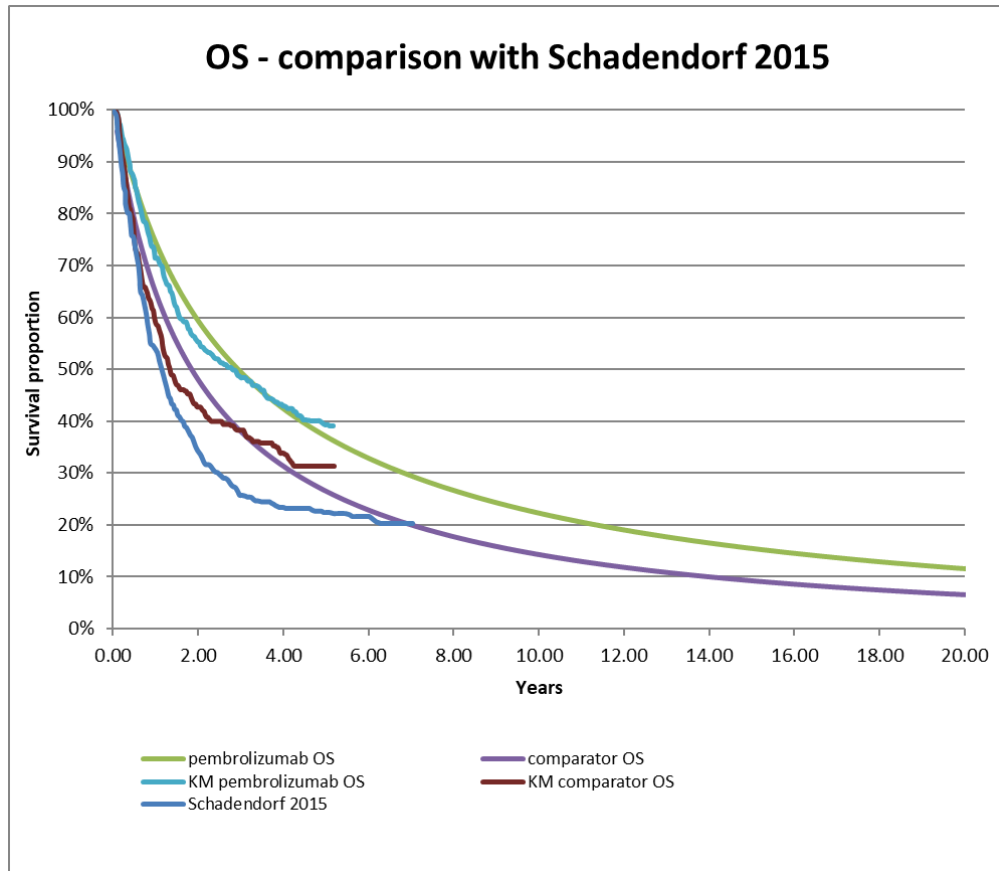

Abbreviations: kg, kilogram; OS, overall survival; PFS, progression-free survival

B – One-way sensitivity analyses based on long-term follow-up data from KEYNOTE-006
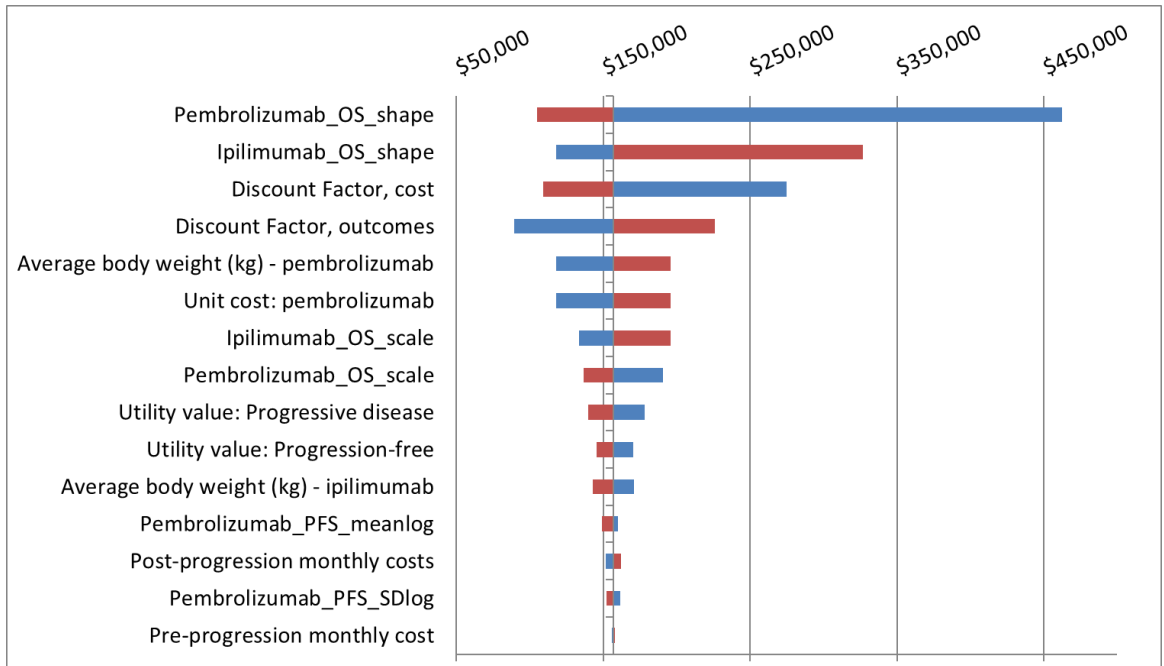
Abbreviations: kg, kilogram; OS, overall survival; PFS, progression-free survival

Figure S2. Overall survival based on long-term follow-up data from KEYNOTE-006 and pooled long-term ipilimumab data from Schadendorf et al. 2015

Abbreviations: KM, Kaplan-Meier; OS, overall survival; PFS, progression-free survival

Chapter 5: Conclusion/Discussion

This thesis explored economic evaluation methods in oncology, both in published literature and HTA agency reports, as well as in a novel model-based application using longer-term follow-up data to investigate the impact and importance of length of follow-up in clinical data and transparency of reporting methods. Broadly, the findings from this thesis suggest that there is a need to improve transparency through better documentation of methods used in economic evaluation, both in the published oncology literature and in HTA agency appraisal documents. Discrepancies observed in published studies and HTA agency reports also suggest the need for reporting guidelines for economic evaluation, in addition to currently existing conduct guidelines. This concluding chapter highlights the key findings and methodological contributions of this thesis, suggests potentially fruitful avenues for future research, and makes a number of concluding statements.

Key conclusions

The past decade has seen many advances in economic evaluation methods diffuse into common usage, including deterministic and probabilistic sensitivity analysis, extrapolation of clinical outcomes, cost-effectiveness acceptability curves, modeling techniques (Markov models, partitioned survival models, mixture-cure models), survival analysis techniques, and the cost-utility analytical structure. However, despite the continuing evolution and methodological advancement of economic evaluation techniques and associated guidelines, little research has been dedicated to investigating uptake of methods over time and across geographies. The objective of this thesis research was to fill these gaps in the literature through investigating published oncology literature, reimbursement submissions to HTA agencies, and adapting existing HTA frameworks to include a life-cycle approach using long-term clinical trial data.

The primary objectives of chapter 2 were to identify, examine, and describe the methods used in economic evaluations in oncology over a 10-year period, while secondary objectives included examining

the use of identified methods across different geographic regions. The research conducted in chapter 2 demonstrated that based on a review of 76 papers (20% random sample of 378 studies included for extraction), some methodological techniques, including testing of the proportional hazards assumption, assessments of statistical fit for extrapolated survival curves, statistical curve-fitting techniques, and validation procedures are inconsistently utilized and reported across different regions, despite economic evaluation guidelines recommending and supporting their use. On the basis of the findings reported in chapter 2, it was suggested that greater detail in reporting of extrapolation methods, statistical analyses, and validation is warranted.

Chapter 3 investigated the question of whether HTA agencies in Canada, the UK, and Australia are consistent in their reporting and appraisal of the economic evaluations submitted by drug manufacturers for reimbursement of oncology medications. Publicly available HTA recommendations and reports for oncology drugs issued by CADTH over a 2-year period, 2019–2020, were identified and compared with the corresponding HTA documents from NICE and the PBAC. Baseline characteristics of the HTAs from CADTH, NICE, and the PBAC were found to be consistently reported, though NICE provided more comprehensive and detailed reports, especially in terms of extrapolation methods, despite the requirements of drug manufacturers to follow the same survival analysis guidelines. The reported criticisms of manufacturer-submitted cost-effectiveness models and the extent of reanalysis undertaken was observed to be different across the 3 HTA agencies, including substantial variation in the level of detail provided to support funding recommendations. In addition, the reanalyzed ICER values reported by the HTA agencies were observed to be double, on average, compared with the ICER values submitted by manufacturers. Given the discrepancies across the HTA agencies as reported in chapter 3, it was suggested that in addition to guidelines for HTA submissions for manufacturers, common reporting standards should be established for the results of HTA agency appraisals.

Standardized reporting is particularly important to consider in the context of time-limited health technology decision making where there is urgency to make funding decisions based on clinical data that are not yet mature. In the absence of common reporting requirements it may be difficult to assess consistency of decision making across multiple appraisals, whereas implementation of common reporting standards can provide a way to enhance transparency and credibility of analyses. Common reporting standards would also allow analyses that incorporate new techniques and methods to be compared with previous analyses in a transparent way, further enhancing credibility.

Since there is no guarantee that the clinical and economic benefits identified at a given moment will persist over time, health technology re-assessment forms a critical component of a life-cycle approach which can be used to improve both patient care and system efficiency. The objectives of Chapter 4 were to determine the impact of using trial data of different maturity (long term versus short term) on survival curve extrapolations, and then assess the impact of these different data on the results of a cost-effectiveness analysis using the example of a pembrolizumab (which was reimbursed based on interim trial data). In this chapter, a partitioned survival model was used to assess the cost-effectiveness of pembrolizumab versus ipilimumab for treatment of advanced melanoma from a US payer perspective over a 20 year time horizon.

This chapter presented a health technology re-assessment using longer-term clinical follow-up data of an oncology drug in the context of economic evaluation. Drawing from a previously published cost-effectiveness analysis from the US[1] which utilized interim clinical data, the health technology re-assessment presented in chapter 4 used long-term clinical data to generate an incremental cost-utility ratio (ICUR) that was 42% higher than the ICUR value from the previous US study. These results also demonstrated that while clinical and economic uncertainty may be reduced in long-term follow-up data, a tradeoff exists between the need to make coverage decisions based on short-term data (with substantial clinical and economic uncertainty), and the delay that would be required to make coverage

decisions based on the longer-term data (with substantially reduced uncertainty). Chapter 4 concludes that adoption of a life-cycle approach to HTA, in at least some cases where long-term follow-up data from clinical trials are available, is needed.

Overall impact of results

This thesis explored economic evaluation methods in oncology, and in doing so has filled several existing gaps in the literature. The research conducted in chapters 2, 3, and 4 demonstrates that non-standardized use of economic evaluation methods in oncology can result in differing assessments of cost-effectiveness, and ultimately, reimbursement decisions being made based on inconsistent information despite a similar or identical clinical evidence base. The impact of these inconsistencies may be especially pronounced in cases where survival analysis and extrapolation methods differ, as small methodological differences can result in large divergences in results.

In the published oncology literature over a 10-year period, basic characteristics of economic evaluation methods have been reported with reasonable consistency, demonstrating that many methodological advancements in economic evaluation have diffused into common usage. However, extrapolation methods and validation of results have been inconsistently utilized or reported. The impact of these inconsistencies in the utilization and reporting of methods may indicate different rates of diffusion of new techniques across geographies, entrenched ways of conducting economic evaluation in particular regions, but may also lead readers/stakeholders to form differing opinions of the value-for-money for identical medications. Greater emphasis on reporting the details of extrapolation procedures, testing statistical fit, and validation of results is needed in order to provide transparency into how these procedures have been carried out, as well as to avoid misinterpretation due to omission of details.

In HTAs recently published by Canadian, UK, and Australian HTA agencies between 2019-2020 (which was the most available data at the time of analysis) the reporting of methods was generally well-

documented for common economic evaluation attributes, but significant differences were observed in the reporting of methodological criticisms and survival analysis methods. These discrepancies suggest that HTA agencies emphasize different methodological areas for critique in their appraisals, and while some degree of difference in reporting should not be surprising, greater consistency and transparency of reporting would assist the public and other stakeholders to better understand the rationale behind funding recommendations.

To address the reporting shortcomings demonstrated in chapters 2 and 3, standardized way of reporting is needed. HTA agency guidelines and good practice guidelines should provide, at minimum, an appendix or ideally fully integrated guidance for what and how methods should be reported in economic evaluation. For example, CADTH should provide specific methods reporting guidance for manufacturers, rather than deferring to existing NICE DSU documents.

In addition, survival analysis techniques continue to evolve over time, and despite the impact that extrapolation can have on results, insufficient attention has been allocated to ensuring consistency in describing and reporting how extrapolation techniques and statistical curve-fitting procedures are used in practice, both in the published oncology literature (as shown in chapter 2) and in appraisals from HTA agencies (as shown in chapter 3). The observed inconsistencies in the reporting of the technical aspects of economic evaluation methods suggest that greater attention should be focused on ensuring transparent reporting of methods based on common reporting standards, both in the literature and in reports published by HTA agencies. Prior to the publication of chapters 2 and 3, very little effort has been applied to comprehensively describe and assess the consistency of methods used and reported.

An example of the potential impact of inconsistency is presented in chapter 4, based on a comparison of interim RCT data with long-term follow-up data, and incorporating commonly used methodological frameworks and analytical techniques detailed in chapters 2 and 3. The duration of follow-up in clinical

data may have sizable impacts on the results of cost-effectiveness analysis due to early clinical trial data being subject to transient effects that may not be representative of the long-term hazards. As a result, the differences in the shapes of the survival curves based on interim and long-term data can lead to outsized differences when the clinical data is extrapolated for use in economic evaluation. This chapter applies the learnings from chapters 2 and 3, focusing on transparency of the methods used including details and discussion of extrapolation techniques, statistical curve-fitting procedures, and multiple assessments of validity to support the main results and conclusions. This example of transparent reporting of economic evaluation methods also serves to quantify the impact of length of follow-up data on the results of cost-effectiveness analysis, highlighting the importance of detailed and transparent reporting of the methods used.

Methodological contributions

Through examinations of published oncology studies, HTA agency reports, and a model-based health technology re-assessment of a drug for treating advanced melanoma, the results of this thesis made several important methodological contributions to the literature.

Using a systematic literature review approach, the results of chapter 2 indicate that while economic evaluation methods are routinely reported in the literature, the degree of detail reported is often insufficient to meet guideline recommendations, especially in terms of survival analysis methods which are often critical to the conclusion of the analysis. In addition to documenting which methods are most commonly used, chapter 2 also suggests what is missing which has rarely been done. As shown by our findings, researchers need to report details of how survival analysis was carried out, and must also seek to validate the results they publish, which will improve the transparency and generalizability of the results. Appropriately detailed reporting and validation are important methodological steps for establishing credibility of the results of economic evaluations.

Similarly, the results of chapter 3 demonstrated varying degrees of transparency and comprehensiveness in published HTA agency reports, signifying that there is room for HTA agencies to better ensure that the rigorous methodological requirements for drug reimbursement submissions are reflected in equally rigorous reporting requirements. While several papers have compared the funding recommendations of HTA agencies, no studies have taken a methodological lens when comparing HTA recommendations. Given the discrepancies in reporting observed between CADTH, NICE, and PBAC, submission guidelines alone appear insufficient for this task: reporting guidelines are also needed in order to harmonize the structure and content of HTAs issued by publicly funded HTA agencies.

Chapter 4 provides what may be the first concrete example of the impact that health technology re-assessment using long-term clinical trial data can have on the results of economic evaluation, potentially expanding the scope of HTA towards incorporation of a lifecycle approach. Through examining the impact of using short-term versus long-term clinical data in economic evaluations, the results of chapter 4 signal that methods, and the transparent reporting of these methods, are critically important for inspiring confidence in the results: even small omissions or unclear reporting of methods can make it impossible to replicate the results of a given economic evaluation. Methodologically, there may also be an inherent tradeoff between uncertainty and the value of the ICER, and HTM may be a useful approach for reducing the inherent uncertainty in the results of economic evaluation.

Limitations

There are a number of limitations identified in chapters 2, 3, and 4 that merit attention. For the systematic literature survey conducted in chapter 2, there are several sources of potential selection bias. These potential biases may be due to limiting searches to English language only, and also due to the 20% random sample taken for analysis which may not be fully reflective or generalizable to the full dataset. It should also be noted that not all published economic evaluations in oncology include survival curve

extrapolation, and the number of studies included in cross-regional comparisons was small. Finally, there may be potential for publication bias as oncology models submitted to HTA agencies for reimbursement were not included in the analysis (this is remedied through the investigations conducted in chapter 3).

In chapter 3, a key limitation was the limited scope of analysis: 2 years. There is potential for publication bias due to this limited timeframe of analysis, and different conclusions could result if conducting a similar analysis using data over a longer timeframe. There is also a potential lack of generalizability to other therapeutic areas due to the exclusive focus on oncology. In addition, reporting of study characteristics of interest was coded as binary (either "yes" or "no") regardless of the quantity of information reported. It was also assumed that manufacturer reimbursement submissions to the HTA agencies studied were similar in structure and content since they were based on the same clinical studies conducted by the same company, but since direct access to the submission documents themselves is not possible this assumption can be neither confirmed nor denied.

Limitations encountered in chapter 4 include that fact that patient-level clinical trial data was not available and instead Kaplan-Meier curve reconstruction techniques[2] were utilized for modeling clinical outcomes in the cost-effectiveness model. The previously published cost-effectiveness model[1] upon which chapter 4 was based also could not be directly procured and therefore it was not possible to identically replicate the median survival values and the approach taken to model utility values. In part due to the rapid evolution of cost-effectiveness modeling methods and constant barrage of new innovative therapies being constantly introduced, the clinical relevance of chapter 4 could also be potentially diminished due to diffusion of new therapeutic comparators during period between publication of interim data and long-term follow-up data used in the analyses.

Considered together, the limitations associated with chapters 2, 3, and 4 present several interesting avenues of potential inquiry for future studies, both to corroborate and further validate the findings as well as to potentially extend the study of economic evaluation methods to therapeutic areas outside of oncology.

Future research

The key findings and conclusions of this thesis were focused on oncology. Future research could be conducted to expand the findings of this thesis through explorations outside of oncology, in other geographic jurisdictions, and over additional time points in order to establish generalizability beyond economic evaluation in oncology.

Expanding from the results presented in chapter 2, additional research into the use of novel modeling techniques such as discrete event simulation, multistate modeling, and mixture cure models, which are more frequently used to overcome specific limitations inherent in more rudimentary analytical approaches, would be useful in order to build on the findings from the 2010-2019 data from published literature used in this chapter.

The results from chapter 3, based on data from 2019-2020, suggest that future studies could seek to conduct similar analyses for HTA recommendations from additional years in order to account for any recent changes in HTA processes or turnover of HTA agency personnel. Efforts could also be put into expanding comparisons beyond CADTH, NICE, and PBAC, in order to include other countries that have adopted HTA processes such as Sweden, the Netherlands, South Korea, Taiwan, and, more recently, Japan.

The preliminary results presented in chapter 4 could be extended using similar methodology applied to therapeutic areas outside of melanoma in order to establish generalizability. In addition, these trial-data-based results could and should be validated by future studies incorporating real-world clinical

evidence to account for the highly controlled environment inherent in clinical trials. In addition, for clinical data associated with observed survival plateaus in the long-term data, a mixture–cure modeling approach could be more methodologically appropriate than standard parametric curve fitting, and future studies could assess the impact long-term versus short-term data in those patient populations. Investigating the robustness of the results by fitting spline models with several knots to the reconstructed trial data, rather than conventional parametric curves, could potentially also be an interesting alternative methodology to consider in future research.

Conclusions

The research conducted for this thesis addresses several important knowledge gaps. First, methods used in economic evaluation, especially for the extrapolation of survival curves, can have profound impacts on the results of cost-effectiveness analysis (chapter 4). Second, the technical aspects of economic evaluation, including extrapolation and statistical curve-fitting, are inconsistently reported in appraisal documents from HTA agencies (chapter 3) and in the published oncology literature (chapter 2). Considered together, these chapters suggest a clear need for greater transparency and common standards for the reporting of economic evaluation methods in order to enhance consistency and avoid misleading conclusions or misinterpretations due to insufficient or omitted details on methods.

Chapter 5 references

1. Wang J, Chmielowski B, Pellissier J, Xu R, Stevinson K, Liu FX. Cost-Effectiveness of Pembrolizumab Versus Ipilimumab in Ipilimumab-Naïve Patients with Advanced Melanoma in the United States. J Manag Care Spec Pharm 2017;23(2):184-194. (In eng). DOI: 10.18553/jmcp.2017.23.2.184.
2. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology 2012;12(1):9. DOI: 10.1186/1471-2288-12-9.