Evaluating Heterogeneity in Treatment Effects and Economic Value of Tumor-Agnostic Drugs

Yilin Chen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Josh J. Carlson, Chair

Anirban Basu

Lurdes Y.T. Inoue

Program Authorized to Offer Degree:

School of Pharmacy

University of Washington

Abstract

Evaluating Heterogeneity in Treatment Effects and Economic Value of Tumor-Agnostic Drugs

Yilin Chen

Chair of the Supervisory Committee:
Josh J. Carlson
Department of Pharmacy

Tumor-agnostic drugs (TAD), also known as histology-independent treatments, have the potential to benefit patients who currently have limited therapeutic options. TAD typically received accelerated approvals based on basket trials which included a small number of multi-cohort, single-arm studies. However, the evaluation of TAD poses major challenges for health technology assessment agencies, such as the potential heterogeneity in treatment effect by tumor type, the lack of comparative data due to single-arm studies, and variable standard of care (SoC) across tumor types. Consequently, these challenges create significant uncertainty regarding the expected clinical and economic impact of TAD.

In Aim 1, I used Bayesian hierarchical models (BHM) to assess heterogeneity in treatment outcomes across tumor types and improve estimates of tumor-specific treatment outcomes from Phase II basket trials, which are crucial for healthcare decision-making. My analysis revealed high

heterogeneity and uncertainty in survival endpoints, including median progression-free survival (PFS) and median overall survival (OS), although the treatment effects are more similar when judged by surrogate endpoint at approval. Metrics such as intra-class correlation can be used to quantify the variation between groups, which could inform a recommendation of TAD for use in all tumor types or a restricting subset of patients. The findings from our study are important because they demonstrated that BHM could reduce uncertainty of estimates derived from basket trial evidence, potentially improving confidence in tumor-agnostic decision making, despite small sample sizes in some tumor types. The methods presented in this study can be applied to the future assessment of TAD.

In Aims 2 and 3, I address comparative effectiveness and economic value of TAD with single-arm trial evidence. To overcome the challenge of lacking comparators, I created eight external controls using observational data from the TriNetx electronic health databases. The Copula method was employed to simulate correlated trial samples while matching the dependence structure of the trial baseline covariates to that in the real-world population [1]. Additionally, an inverse odds weighting approach was used to further balance the baseline characteristics between the trial and external control arms [2]. Weighted Cox regressions showed that patients with MSI-H/dMMR advanced/metastatic colorectal and endometrial cancers receiving pembrolizumab were associated with significant prolonged PFS but not OS than real-world patients receiving chemotherapies. This analysis demonstrated that incorporating external control data in early phase trials may provide a more comprehensive understanding of treatment effects of tumor-agnostic drugs than relying solely on single-arm trials.

Finally, using adjusted efficacy inputs from Aim 1 and external controls from Aim 2, I assessed the economic value of pembrolizumab compared to SoC across 8 tumor types to inform

coverage and reimbursement decisions in the United States. A partitioned survival model with three health states (i.e., progression-free, post-progression, and death) was developed to evaluate the cost-effectiveness of pembrolizumab for previously treated patients with advanced or metastatic MSI-H/dMMR cancers [3]. We found substantial variation in the economic value across tumor types, with pembrolizumab being a cost-effective strategy in treating colorectal and endometrial cancers at $150,000 willingness-to-pay per quality adjusted life years threshold, compared to SoC chemotherapies. However, pembrolizumab was not found to be cost-effective in treating other assessed cancers. The main findings from value of stratification estimates suggest that recommendations for using pembrolizumab in specific patient populations, based on comparative effectiveness or net health benefit, could result in greater overall value to the healthcare system compared to a tumor-aggregated recommendation.

TABLE OF CONTENTS

List of Figures iii

List of Tables iv

Chapter 1. 7

Tumor-Specific Decisions Using Tumor-Agnostic Evidence from Basket Trials: A Bayesian Hierarchical Approach 7

1.1 Abstract 7

1.2 Introduction 8

1.3 Methods 11

1.4 Results 16

1.5 Discussion 19

1.6 Limitations 25

1.7 Conclusion 25

1.8 Tables & Figures 27

1.9 Supplement 33

Chapter 2. 37

Generating External Controls for Single-Arm Trials in Tumor-Agnostic Indications Leveraging Electronic Health Record Data Using Copula and Propensity Score Methods 37

2.1 Abstract 37

2.2 Introduction 38

2.3 Methods 41
2.4 Results 47
2.5 Discussion 50
2.6 Limitations 54
2.7 Conclusion 55
2.8 Tables & Figures 56
2.9 Supplement 64

Chapter 3. 75

Tackling Challenges in Evaluating Economic Value of Tumor-Agnostic Therapies: A Cost-Effectiveness Analysis of Pembrolizumab 75

3.1 Abstract 75
3.2 Introduction 76
3.3 Methods 79
3.4 Results 85
3.5 Discussion 89
3.6 Limitations 94
3.7 Conclusion 95
3.8 Tables & Figures 96
3.9 Supplement 102

Bibliography 118

LIST OF FIGURES

Figure 1.1. Summary of posterior distributions by tumor types for objective response rate

Figure 1.2. Summary of posterior distributions by tumor types for median progression-free survival

Figure 1.3. Summary of posterior distributions by tumor types for median overall survival

Figure 1.4. Summary of posterior distributions by tumor types for median overall survival, excluding small intestine cancer

Figure 1.5. Posterior density plots by tumor types and treatment outcomes obtained by fitting a BHM

Figure 2.1. Real-world chemotherapy survival curves by tumor types

Figure 2.2. Adjusted and unadjusted Kaplan-Meier curves for colorectal cancer

Figure 2.3. Adjusted and unadjusted Kaplan-Meier curves for endometrial cancer

Figure 3.1. Cost effectiveness acceptability curves

LIST OF TABLES

Table 1.1. Trial-reported clinical outcomes among included patients

Table 2.1. KEYNOTE trial inclusion/exclusion criteria applied to the real-world dataset to identify eligible patients

Table 2.2. Baseline characteristics for the trial cohorts versus real-world external control arms

Table 2.3. Comparison of median PFS and OS between pembrolizumab and chemotherapies

Table 2.4. Estimated survival rates and hazard ratios for colorectal and endometrial cancers

Table 3.1. Model parameters: base-case values and sources

Table 3.2. Stratified base case results

Table 3.3. Value of stratification: outcome estimates based on tumor-aggregated versus tumor-specific recommendations

Table 3.4. Aggregated incremental cost-effectiveness ratios and value-based prices

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the following individuals and organizations for their invaluable support and assistance throughout my doctoral journey.

First and foremost, I am immensely grateful to my supervisor, Dr. Josh J. Carlson, for your guidance, expertise, and unwavering support. Your insightful feedback and continuous encouragement have been instrumental in shaping the direction of my research.

I extend my heartfelt appreciation to the members of my dissertation committee, Dr. Anirban Basu and Dr. Lurdes Y.T. Inoue. Their valuable insights, constructive feedback, and dedication to academic excellence have significantly enhanced the quality of my work.

I would like to sincerely thank other faculty members who I have collaborated or interacted with: Monisha, Sean, Aasthaa, Ryan, Beth, Lou, and Doug. Thank you for sharing your knowledge, resources, and research opportunities that has been invaluable in broadening my understanding of the field. Thanks to staff and faculty at the CHOICE Institute for providing me with a stimulating and supportive academic environment.

I would also like to acknowledge the financial support I received from the Magnuson Family and Rubenstein Family. Your generous scholarships have motivated and supported me to attend relevant conferences and courses, enriching my research experience and advance my personal and professional growth.

Furthermore, I would like to express my gratitude to my dear friends and colleagues at CHOICE Institute, especially Enrique, Nathaniel, Lizzy, Meng, Woojung, and Boshen. Thank you for answering so many of my questions and giving me support in need. Thanks to all my friends

at CHOICE, especially my cohort friends Sara and Jacinda. You have made this journey more enjoyable.

Lastly, I want to thank my family for their unconditional love, encouragement, and patience throughout this demanding process. Mom and Dad, thanks for your understanding and support for whatever I choose to do. Even during challenging times, you have been crucial in keeping me motivated and focused through my PhD journey. I also want to express my gratitude to Jingcheng: thank you for always being by my side, believing in me, encouraging me when I fail, and cheering for me when I succeed.

In conclusion, the completion of this dissertation would not have been possible without the collective efforts and support of these individuals and organizations. I am truly grateful for your presence in my life.

Chapter 1.

Tumor-Specific Decisions Using Tumor-Agnostic Evidence from Basket Trials: A Bayesian Hierarchical Approach

1.1 ABSTRACT

INTRODUCTION: Heterogeneity in treatment effect across tumor types remains a challenge to evidence interpretation and implementation of tumor-agnostic drugs (TADs), which are typically approved based on basket trial evidence. We sought to use Bayesian hierarchical models (BHM) to assess heterogeneity and improve estimates of tumor-specific treatment outcomes, which are crucial for healthcare decision-making.

METHODS: We fitted BHMs and Bayesian fixed-effect models to evaluate the objective response rate (ORR), the median progression-free survival (mPFS), and the overall survival (mOS). We estimated the posterior distribution of outcomes for each tumor type, the pooled effects, and intra-class correlations (ICC). Using published trial evidence for pembrolizumab (KEYNOTE-158/KEYNOTE-164), we obtained the predictive outcomes in a new cancer type drawn from the same population. In the base case, we assumed non-informative priors with uniform distributions for between-tumor standard deviation. We performed sensitivity analyses with various priors to account for uncertainty in the prior specification.

RESULTS: The BHMs shrunk the original tumor-specific estimates toward a pooled treatment effect. The borrowing of information across tumor types resulted in less variability in the posterior tumor-specific estimates compared to the original trial estimates, reflected in narrower 95% credible intervals (CrLs). We found low heterogeneity for ORR but high heterogeneity for mPFS and mOS across cancers (ICC: 0.22, 0.87, 0.70). The predicted posterior medians and

95%CrLs were 0.37 (0.15-0.64) for ORR, 3.82 months (0.24-50.45) for mPFS, and 14.87 months (0.42-276.49) for mOS, respectively.

CONCLUSIONS: Borrowing information through BHM can improve the precision of tumor-specific estimates, thereby facilitating more robust policy decisions regarding TADs. Our analysis revealed high heterogeneity and uncertainty in survival endpoints. Both pooled and tumor-specific estimates are informative for clinical and coverage decision making.

1.2 INTRODUCTION

Precision medicine has enabled the development of drugs that target specific molecular targets. In oncology, tumor-agnostic drugs (TADs) have begun to shift the therapeutic paradigm from organ-based treatment to a molecular target-based approach [4-6]. In 2017, the Food and Drug Administration (FDA) granted an accelerated approval for the first TAD, pembrolizumab, for the treatment of adult and pediatric patients with unresectable or metastatic, microsatellite instability–high (MSI-H) or mismatch repair–deficient (dMMR) solid tumors that have progressed after prior treatment [7]. As of 2022, FDA has approved five medicines for tumor-agnostic indications, with many more in the drug development pipeline [7][8-11]. This paradigm shift in drug development has driven the field of precision medicine to consider the molecular details of cancer biology to identify opportunities to improve the treatment of patients with cancer [12, 13]. Histology-independent treatments have the potential to provide substantial benefits to patients who currently have few or no therapeutic alternatives.

The accelerated approvals for pembrolizumab and other TADs were based on basket trials characterized by a small number of multi-cohort, single-arm studies [14]. Basket trials simultaneously evaluate the effect of a targeted agent in various disease subtypes that harbor the

same genetic or molecular alteration [15]. Planning a basket trial is administratively more efficient than conducting separate trials for each cancer type. It has predominantly been used for early-phase oncology drug development, with a recent review finding 37 ongoing basket trials registered in ClinicalTrials.gov [5, 16]. The emergence of TADs and basket trials has created substantial innovation and unique challenges for drug approval and value assessment for regulatory and health technology assessment (HTA) agencies. One major concern of analyzing early phase basket trials is the potential for heterogeneity of the treatment effect in various tumor types [17, 18]. For example, the Phase II KEYNOTE-158 study enrolled 233 patients with advanced MSI-H or dMMR cancers and 27 different tumor types previously treated [19, 20]. The study showed the objective response rate (ORR) ranging from 18.2% in pancreatic cancer to 48% in endometrial cancer. Moreover, even though it is approved for use in all solid tumors, it is difficult to extrapolate these results to a tumor type not included in the basket trial with the same biomarker/mutation.

Bayesian hierarchical models (BHM), or Bayesian random-effects models, are particularly well-suited for handling issues related to heterogeneity and information from varying sample sizes. They are adept in borrowing information on treatment effects across tumors, thereby increasing the precision of tumor-specific estimates while down-weighting extreme results in cohorts with small sample sizes [21-23]. It has advantages over conventional stand-alone subgroup analyses (also known as the approach of no borrowing), which often lack sufficient power to detect treatment effects because of the small sample size of each stratum [24]. At the other extreme, a simple pooled analysis, i.e., complete pooling, would not account for systematic patient-to-patient differences across tumor types [23, 25-27]. BHMs are unique in that they are able to simultaneously account for heterogeneity in treatment outcomes between tumor

types and patient-to-patient variability. The uncertainty in between-tumor heterogeneity can also be directly modeled [28]. BHMs are justified under the assumption of exchangeable treatment effects, which means that it is unknown a priori in which tumor types the treatment performs better or that the tumor-specific treatment effects are drawn from the same underlying overall distribution [29].

BHMs have served an important role in multiple settings in health sciences research, including basket trial design, trial efficacy evaluation, observational studies, comparative effectiveness research, meta-analysis, and evaluation of disease subtypes [30-34]. Prior studies using BHMs have only evaluated ORR and no other commonly used primary endpoints in clinical trials, such as median progression-free survival (mPFS) and median overall survival (mOS). The objective of this study was to estimate the treatment outcomes for each tumor type based on data from the KEYNOTE-164 [NCT02460198] and KEYNOTE-158 [NCT02628067] trials, using Bayesian random-effects models [19, 20, 35]. Compared to a random-effects model that allow the true effect sizes to differ, a fixed-effect model assumes that there is one true effect size, and that all difference in observed effects is due to within-group error, and thus no between-group variance [36]. We implemented BHMs and two types of Bayesian fixed-effect models (whether or not assuming more than one true effect) to obtain the tumor-specific and pooled treatment effects across all tumor types, and the predictive distributions in an unrepresented, new cancer type from the same overall population. The study findings will provide improved tumor-specific estimates for pembrolizumab using early-phase basket trial evidence and insights into the methods to consider when assessing the clinical benefits of TADs. The methods presented in this study can be applied to the future assessment of TADs.

1.3 METHODS

Overview of study methods and outcome measures

We fitted three types of models to evaluate the objective response rate (ORR), mPFS, and mOS: (1) Bayesian hierarchical model, which assumes varying true effect sizes, (2) Bayesian fixed-effect model, which assumes one true effect size, and (3) Bayesian fixed-effects model, assuming no between-tumor variation but with varying mean effects (hence the plural term fixed-effects). Posterior distributions summarized with the posterior mean, median, and 95% posterior credible intervals (CrLs) were obtained for treatment effects in each tumor type. The posterior distribution can be expressed as: $p(\theta_1, \dots, \theta_N, \phi|Data) = \prod_{i=1}^{N} p(Data_i|\theta_i, \phi)p(\theta_i|\phi)p(\phi)$. θ_i represents treatment effect by tumor type and ϕ is the overall mean treatment effect. *Data*$_i$ refers to data from tumor type i and *Data* refers to the data from all tumor sites.

Metrics of heterogeneity were calculated to quantify total variation across tumor types [28]. Posterior distributions for the between-tumor standard deviation (SD) and 95% CrLs were computed and summarized. Intra-class correlation (ICC) provides a measure of the proportion of total variation that is explained by between-tumor variability, expressed as $ICC = \frac{\tau^2}{\tau^2+\tau_\varepsilon^2}$. ICC is the ratio of the between-cluster variance (τ^2) to the total variance ($\tau^2 + \tau_\varepsilon^2$; τ_ε^2 is the within-cluster variance), which ranges from 0 to 1 [37, 38]. Therefore, low ICC indicates less variation in treatment effects between tumor types, which means a tumor-aggregated decision may be more appropriate, and vice versa. The calculation of ICC for each assessed outcome was outlined in the Intra-class Correlation section below.

Setting

Pembrolizumab was approved for use among previously treated adult and pediatric patients with unresectable or metastatic MSI-H/dMMR (i.e., MSI-H in this paper) cancers based on tumor response rate and durable clinical benefit in two Phase II studies. A total of 63 patients (cohort B) with MSI-H colorectal cancer (CRC) and 233 patients (cohort K) with MSI-H non-CRC were included in this analysis. The data cut-off date was September 4, 2018 for KEYNOTE-164 and October 5, 2020 for KEYNOTE-158 studies [19, 20, 35, 39]. The ORR, mPFS, and mOS data for eight tumors (colorectal, endometrial, gastric, cholangiocarcinoma, pancreatic, small intestine, ovarian, and brain) are summarized in Table 1.1. Due to the nature of the study, which involved using de-identified secondary data and no direct contact or interaction with human participants, the ethical approval was not obtained.

Model specification

Objective response rate

We fitted a Bayesian hierarchical model to the ORR data. The first unit of the hierarchical model is the aggregated data with the counts of responses per tumor type, and the second unit of analysis is the tumor type, with patients grouped or nested within tumor types. For each tumor type, the observed response rates are modeled using a binomial distribution, where y_j is the number of responses for tumor j, n_j is the number of patients for tumor type j, and p_j is the probability of response. A logit transformation is applied to the probability of response, expressed as $\theta_j = \ln\left(\frac{pj}{1-pj}\right)$, and it is assumed to be normally distributed. The hyperparameters are the overall mean log-odds of response (ϕ), average across all possible tumor types, and the between-tumor SD (τ). The two hyperparameters are treated as unknown random variables. Specifically, we assumed the following probability distributions.

$$y_j \sim Binomial(p_j, n_j)$$

$$logit(p_j) = \theta_j$$

$$\theta_j \sim Normal(\phi, \tau^2)$$

$$\phi \sim Normal(0,\ 10)$$

$$\tau \sim Uniform(0,\ 3)$$

In sensitivity analysis, we repeated our analysis using other prior distributions for the between-tumor SD (τ): (i) the exponential (0.3) distribution (that is, exponential with rate 0.30), another common non-informative prior distribution but with longer tail compared to a uniform distribution; (ii) a truncated normal (0, 100)T(0,) distribution (100 is the variance), which restricted all values to be positive; (iii) a uniform (0, 5) distribution, which assumes equally likely values between 0 and 5; and (iv) an exponential (1) distribution.

All tumor types were included in the analysis except for brain tumor due to the small cohort sample size (i.e., 13 patients).

Median Progression-free Survival and Median Overall Survival

Two hierarchical models were fitted separately to model mPFS and mOS. Data from all eight tumor types were included. Due to the positive nature of survival endpoints, we performed a log transformation to median PFS and OS data in the base case and assumed normal distribution for the log-transformed data. The model structure of mPFS and mOS are outlined below, where y_j is the log-transformed mPFS or mOS for tumor j and is assumed to be normal.

Due to the lack of individual-level survival data, we computed the within-tumor variance using the following methods. Note first that under an exponential distribution for the PFS and OS at the individual level data in each tumor type with parameter λ (with density function $f(x;\ \lambda) =$

$\lambda \exp(-\lambda x)$), we can re-express this parameter considering the median survival as $\lambda = \frac{\log(2)}{\text{Median}}$. According to Laplace, the distribution of the sample median from a population with a density function $f(x)$ is asymptotically normal with variance $\frac{1}{4nf(m)^2}$, where m is the median and n is the sample size [40, 41]. Thus, we were able to compute the sample variance for each tumor type by using this result. Finally, we use the delta method to obtain the variance of $\log(y_j)$, which were denoted as σ_j^2 in the model specification above.

The hyperparameters are the overall mean log-mPFS or log-mOS (ϕ), average across all possible tumor types, and the between-tumor SD (τ). The two hyperparameters are assumed to be normally and uniformly distributed, respectively.

$$\log(y_j) \sim Normal(\theta_j, \sigma_j^2)$$

$$\theta_j \sim Normal(\phi, \tau^2)$$

$$\phi \sim Normal(0,\ 2)$$

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

Due to unobserved median OS among small intestine cancer by the end of the study period, our model also accounted for right-censored observations. This required us to assume that the unobserved outcome value would be greater than a certain value (e.g., end of the study period). In addition, we initialized our MCMC algorithms by setting a value that was larger than the censoring limit. We conducted additional sensitivity analysis by excluding small intestine cancer in the median OS models to assess the impact of considering right-censored observations.

To define the prior distributions of parameters ϕ and τ, we assume a normal distribution for ϕ and uniform distribution for τ in base-case analyses. In addition, sensitivity analyses assumed the following distributions for τ: exponential, uniform, and half-normal distributions

[37, 42]. A smaller value of τ implies that the parameters are more similar and results in greater shrinkage, and vice-versa [43]. The priors were selected to be relatively non-informative in the base case to allow the data to primarily determine the degree of similarity among the different tumor types. Weakly informative priors based on scientific knowledge were explored in the sensitivity analyses.

Intra-class Correlation

Different methods were used to compute ICC for ORR and survival endpoints. To compute ICC for ORR, we adopted the analysis of variance (ANOVA) estimator approach, which appropriately accounted for the variable size of clusters [44, 45]. The ANOVA estimator is given by $\widehat{\rho_A} = \frac{MSB-MSW}{MSB+(n_A-1)MSW}$, where MSB and MSW are, respectively, the between-group and the within-group mean squares. Suppose that there are k clusters and that the i^{th} cluster has n_i individuals. Then

$$MSB = \frac{1}{k-1}\left\{\sum \frac{Y_i^2}{n_i} - \frac{(\sum Y_i)^2}{N}\right\},$$

$$MSW = \frac{1}{N-k}\{\sum Y_i - \sum \frac{Y_i^2}{n_i}\}, \text{ and}$$

$$n_A = \frac{1}{k-1}\left(N - \frac{\sum n_i^2}{N}\right).$$

where $Y_i = \sum X_i$ is the total number of successes in cluster i, and $N = \sum n_i$ is the total number of observations in the study [44, 45]. For mPFS and mOS models, between-tumor variance was directly obtained from the posterior summaries of τ^2 after applying the exponentiation to transform back the posterior distributions to their original scales. Within-tumor variance was estimated using the estimated standard error for the estimated treatment effect by tumor type and the number of patients as described in the sections above.

Bayesian Estimation and Inference

A Markov-chain Monte Carlo (MCMC) algorithm was used to generate samples from the posterior distribution, using the likelihood and priors as discussed in previous sections. For each model, 3 parallel chains containing 65,000 samples from the posterior distribution were obtained after a burn-in of 5,000, using a thinning factor of 5. Convergence was assessed using the Gelman-Rubin statistic which evaluates MCMC convergence by comparing multiple Markov chains. We also inspected the trace plots and autocorrelation plots for diagnostic checks. The autocorrelation measures how linearly correlated the current value of the chain is to the past values. In addition, prior predictive checks were conducted to assess the appropriateness of the priors and the uncertainty. All Bayesian analyses were performed using R (version 4.1.2) and the package R2JAGS.

1.4 RESULTS

A total of 290 patients with eight types of metastatic or advanced cancers were included in the analysis, including endometrial (n=79), colorectal (n=63), gastric (n=42), small intestine (n=25), ovarian (n=24), cholangiocarcinoma (n=22), pancreatic (n=22), and brain cancers (n=13). The clinical outcomes of ORR, mPFS, and mOS from the KEYNOTE-164 and KEYNOTE-158 trials were summarized in Table 1.1. We reported the mean and 95%CrLs for the posterior distribution of parameters of interest across three outcomes from Figure 1.1 to Figure 1.4. The posterior distributions for tumor-specific and pooled outcomes were plotted in Figure 1.5.

Posterior Distributions of Objective Response Rate

The highest posterior median ORR was observed in endometrial (0.44, 95%CrL: 0.34-0.55), followed by small intestine (0.41, 95%CrL: 0.29-0.58), cholangiocarcinoma (0.38, 95%CrL: 0.25-0.54), ovarian (0.36, 95%CrL: 0.22-0.49), colorectal (0.35, 95%CrL: 0.25-0.45), gastric (0.34, 95%CrL: 0.22-0.45), and pancreatic cancers (0.31, 95%CrL: 0.14-0.43). Using the random-effects model, the pooled mean posterior ORR across seven cancers was 0.37 (95%CrL: 0.26-0.47) and the predictive probability of response in a new, unrepresented MSI-H cancer was 0.37 (95%CrL: 0.16-0.63). Similar posterior ORR was obtained from the fixed-effect model with narrower 95% credible intervals (0.38, 95%CrL: 0.32-0.44), assuming no between-tumor heterogeneity (Figure 1.1).

The sensitivity of the BHM results was assessed using alternative prior distributions for the between-tumor SD (τ). The models rendered similar estimates to those obtained with a uniform (0, 3) distribution in the base case. The results showed low estimated heterogeneity between tumors irrespective of the prior distribution, with the intraclass correlation being 0.02 in Table S1.1.

Posterior Distributions of Median Progression-free Survival

When different types of tumors were compared for the mPFS, small intestine cancer had the longest posterior median estimate at 20.48 months (95%CrL: 11.53-35.94), followed by endometrial cancer (12.69, 95%CrL: 9.24-17.41), cholangiocarcinoma (4.18, 95%CrL: 2.34-7.42), colorectal (4.09, 95%CrL: 2.87-5.83), gastric (3.23, 95%CrL: 2.11-4.96), ovarian (2.3, 95%CrL: 1.3-4.02), pancreatic (2.21, 95%CrL: 1.22-3.95), and brain cancers (1.29, 95%CrL: 0.6-2.76). The pooled mean mPFS was 3.79 months (95%CrL: 1.46-8.19) with the random-effects model, with the predictive estimate of 3.82 months (95%CrL: 0.24-50.45). The fixed-

effect model assuming one mean effect across tumor types yielded higher pooled estimates than those obtained from the fixed-effect model assuming varying mean effects (5.35, 95%CrL: 4.53-6.31; vs. 4.38, 95%CrL: 3.73-5.17) (Figure S1.2).

The sensitivity of the BHM results was assessed by assuming various prior distributions for between-tumor variance (τ). The models rendered similar tumor-specific and pooled estimates to those obtained when assuming a uniform distribution (0, 5) in the base case. The results showed high estimated heterogeneity between tumors irrespective of the prior distribution. This can be seen in the intraclass correlation value (0.87) in Table S1.1 and the considerably wide 95%CrLs around the predicted estimate (Figure 1.2**)**.

Posterior Distributions of Median Overall Survival

Small intestine cancer was estimated to have the longest posterior mOS at 61.35 months with wide 95% credible intervals (25.57-782.82) due to its right censoring at the end of study period. The posterior median estimate for mOS was 45.38 months for endometrial cancer (95%CrL: 33.08-62.2), 32.36 months for colorectal cancer (95%CrL: 22.78-46.08), 31.99 for ovarian cancer (95%CrL: 18.2-56.07), 19.13 for cholangiocarcinoma (95%CrL: 10.7-34.37), 11.17 for gastric cancer (95%CrL: 7.2-17.2), 6.26 for brain cancer (95%CrL: 2.91-13.35), and 4.11 for pancreatic cancer (95%CrL: 2.22-7.49). The estimated pooled mOS was 14.7 months (95%CrL: 4.02-34.49) with the random-effects model, with the predictive estimate of 14.87 months (95%CrL: 0.42-276.49). Similar to mPFS models, the fixed-effect model with a single mean effect across tumor types provided higher pooled estimates than those obtained from the fixed-effect model with varying mean effects (23.43, 95%CrL: 19.8-27.74; vs. 19.05, 95%CrL: 16.14-22.57) (Figure 1.3). The posterior distributions obtained from models excluding small

intestine cancer were summarized in Figure 1.4, with similar tumor-specific estimates but lower pooled mean effects.

Sensitivity analysis results were summarized in Table S1.4. When compared to the base case model, some differences were manifested in the posterior estimates for small intestine cancer, the predicted and pooled outcomes, indicating that using BHM for evaluating median OS with censored observations is sensitive to the distribution of between-tumor variance under this setting with limited information about the censored data. The results also showed high estimated heterogeneity between tumors irrespective of the distribution of the data and the prior. The intraclass correlation value was 0.7 (Table S1.1) and the 95% credible intervals around the predicted estimate were wide in both models with and without small intestine cancer (Figure 1.3; Figure 1.4**)**.

Model diagnostics

Gelman-Rubin diagnostics estimates were near 1 for all parameters in all models. The trace plots or autocorrelation plots did not suggest convergence failure of the MCMC algorithm.

1.5 DISCUSSION

In this study, we used Bayesian random-effects and fixed-effect models to estimate the posterior distributions of tumor-specific and overall treatment outcomes of pembrolizumab, including ORR, median PFS, and median OS. The pooled ORR across seven tumor types was 35%-37% using Bayesian random-effects and fixed-effect models, meaning that at least 35% of cancer patients had tumor shrinkage. This value is greater than the commonly used regulatory target of 30% with tumor shrinkage used for single-agent anticancer therapies to demonstrate

breakthrough activity and for monitoring in phase II basket trials [46, 47]. The BHM approach shrunk the original tumor-specific estimates towards a pooled treatment effect and the resultant ORR estimates in all tumor types were above the 30% target. The revised estimates demonstrated markedly less variability than the original estimates, with narrower 95% credible intervals compared to the original CIs. We also found high heterogeneity in median PFS and median OS across eight tumor types and low heterogeneity in ORR across seven tumor types, excluding brain tumor. The BHM approach allows for prediction of outcomes in tumor types outside of the studied types; however, the wide 95%CrLs around these predicted estimates indicate substantial uncertainty, especially for the predicted median PFS and median OS estimates.

Implications

Our study demonstrated that a Bayesian hierarchical approach could reduce the uncertainty in estimates from available evidence, providing more confidence in tumor-agnostic decision making, given the small sample sizes in some tumor types. This flexible approach can incorporate censored observations of survival endpoints available at trial completion. Moreover, metrics such as ICC can be used to quantify the variation between groups, which can inform a recommendation of pembrolizumab for use in all tumor types or restricting it to a subset of patients. In addition, BHM's can be used to inform many different types of research questions. It has been used in meta-analysis and economic evaluation of healthcare interventions [43, 48]. For instance, Kwok et al. [43]demonstrated the use of BHM to incorporate all available information from multiple sources, including a meta-analysis of immunosuppressive therapy in idiopathic dilated cardiomyopathy using data from related trials, and a subgroup analysis of the National

Institute of Neurological Disorders and Stroke intravenous tissue plasminogen activator stroke trial. Many more interesting questions can be answered by applying such models in future studies, such as estimating the posterior probability that patients with each type of cancer demonstrate the largest treatment effect among all cancer patients or the probability of response rate higher than a certain threshold. Such information would provide additional evidence on tumor-specific treatment outcomes, which would inform the approval and coverage decisions.

A high ORR with a long duration of response has been used for accelerated approvals under the assumption that it is highly predictive of longer-term survival outcomes. Our analyses did not support this assumption for pembrolizumab. Specifically, despite the fact that the posterior mean ORR in all tumor types was above 30%, we observed low mPFS and mOS in certain tumor types, such as pancreatic and brain tumors. Furthermore, although homogenous ORR outcomes were shown across tumor types, there was high variability for the survival endpoints (i.e., median PFS and OS). In this context, confirmatory trials or post-marketing evidence generation should be considered following accelerated approvals and could be targeted towards tumor types with the greatest discordance and/or uncertainty. Future studies can model the correlation of median PFS and OS outcomes in the BHM and explore additional survival endpoints such as 12-month/24-month survival rates. In addition, alternative approaches that allow tailored or partial borrowing across strata can be employed when there is high between-tumor heterogeneity in the basket trial setting [49].

Considerations for choosing random-effects vs fixed-effect models

Similar to a meta-analysis, we derived the distribution of the pooled treatment effect for each outcome to quantify the overall clinical benefit across tumor types. We explored both

Bayesian random-effects and fixed-effect models to obtain a pooled effect size across subgroups, and thus, it is crucial to determine the optimal choice of model for a given scenario. The fixed-effect model assumes that there is no between-tumor heterogeneity, or that patients with different tumors are part of a homogenous population, and the only cause for differences in observed effects is the sampling error of tumor sites. In contrast, the random-effects model has the advantage of assuming there is more than one true effect size, but a pre-specified distribution of true effect sizes. The caveat is that this assumption may fail if an outlying group is inconsistent with a certain type of distribution for the between-group variability (τ), potentially resulting in biased estimates for that outlying group. Therefore, it is important to consider sensitivity analyses to assess the robustness of the results to changes in the choices of prior distributions.

The selection of a fixed-effect or a random-effects model depends on how heterogeneous the effect sizes are across tumor types. For example, the fixed-effect pooled estimate of the ORR model might be closer to the true effect size if we believe that patients with different tumor types respond similarly to the pembrolizumab. Then, we would focus on reporting the mean and its confidence interval. However, the assumptions of the fixed-effect model may seem too simplistic in many real-world populations and for long-term outcomes. When there is high heterogeneity across subgroups, for outcomes such as median PFS and OS, it may be more plausible to use the pooled effect size from a random-effects model and report tumor-specific estimates rather than focusing on the common effect from the fixed-effect model.

Degree of shrinkage

Shrinkage in Bayesian hierarchical models helps to mitigate the impact of extreme estimates in subgroups and leads to more stable and conservative estimates. The original tumor-

specific estimates were shrunken towards a pooled treatment effect in BHMs which is consistent with prior studies. For example, Chugh and others evaluated the ORR of a phase II trial of Imatinib in 10 histologic subtypes of sarcoma using a BHM. The posterior estimates based on the BHM shrunk the estimates of the subgroup-specific ORR toward the population mean, thus reducing extreme estimates for subtypes with few patients (e.g., only two patients with rhabdomyosarcoma) [47]. We found that greater shrinkage of posterior estimates toward the overall mean occurred for tumor subgroups with relatively smaller sample sizes or less precision. For example, posterior ORR of ovarian cancer (raw: 0.33, 95%CrL: 0.16-0.55 vs revised: 0.36, 95%CrL: 0.22-0.49; n=24 patients) has shrunken to a greater degree than that of colorectal cancer (raw: 0.33, 95%CrL: 0.22-0.46 vs revised: 0.35, 0.25-0.45; n=63 patients). In other words, because there is fewer underlying data (i.e., when there is greater uncertainty), the ovarian cancer estimate borrows more information from the other tumor subgroups than the colorectal cancer estimate. In addition, the degree of shrinkage also depends on how far the estimate is from the overall mean. For example, more shrinkage was observed for cholangiocarcinoma (raw: 0.41, 95%CrL: 0.21-0.64 vs. revised: 0.38, 95%CrL: 0.25-0.54) than for pancreatic cancer (raw: 0.18, 95%CrL: 0.05-0.4 vs. revised: 0.31, 95%CrL: 0.14-0.43), with equal sample size (n=22 patients). Similar findings are shown in mPFS and mOS models. Prior studies have demonstrated the benefit of shrinkage and borrowing of information across subgroups, including improving the accuracy of estimates overall than if we were to use the naïve estimates from individual subgroups [24, 50].

Prior distributions

Between-tumor variance (τ^2), the shrinkage parameter, estimates the variation across subgroups and controls the strength of information borrowing, was treated as an unknown parameter following a non-informative prior in our study. A small value of τ^2 indicates less heterogeneity across subgroups and thus induces strong information borrowing, whereas a large value of τ^2 induces little information borrowing [51]. Therefore, it is important to consider prior distributions for τ^2. In our study, we chose non-informative uniformly distributed priors in the base case to allow data rather than prior knowledge to primarily determine the degree of borrowing. This is consistent with previous investigation around the priors in the literature. Furthermore, when more prior information is desired, for instance to restrict τ away from very large values, Gelman et al. recommended working within the half-t family of prior distributions, which are more flexible. However, the inverse-gamma family of non-informative prior distributions is not recommended due to the sensitive inferences when τ is estimated to be near zero [37, 42].

In a fully Bayesian approach, the third level of the hierarchical model consists of the prior distributions of the hyperparameters which quantify knowledge based on external information. Vague, non-informative prior distributions may be used in the absence of relevant external information, or to permit inferences to depend exclusively on the present data [43]. In an empirical Bayes approach, in contrast, the hyperparameters are derived from the data and then treated as known values [52]. However, this often results in estimates of the first-level parameters that are too precise as it ignores the uncertainty in the values of the hyperparameters [53].

1.6 LIMITATIONS

Our study is not without limitations. The main assumption of this approach is exchangeability that efficacy is similar (rather than equal or independent) across baskets, with different tumors not determining a particular ordering of effectiveness a priori. It is unknown whether the exchangeability assumption will apply to unrepresented tumors. Furthermore, our results may be dependent on the data distribution and the priors. To address the uncertainty, we used existing literature and knowledge to come up with the ones that make clinical sense and tested a few to examine whether the results change significantly. In the absence of individual patient data, we computed the within-tumor variance for mPFS and mOS models using above-mentioned methods. We also used published Kaplan-Meier curves to estimate median OS for colorectal and endometrial cancers, assuming an exponential distribution. The assumption of exponential distribution for time-to-event outcomes is also a potential limitation since it is assuming constant hazard. Our proposed model, however, could be extended to consider other parametric distributions such as the Weibull distribution and eliciting other quantiles of the Kaplan-Meier curves instead of just the median OS. Future studies should validate these findings when more data becomes available. Finally, we only derived predictive distributions from the BHM and not from the fixed-effects model considering its capability to account for the between-tumor heterogeneity.

1.7 CONCLUSION

Assuming exchangeable effects of tumor-agnostic therapies across cancers, BHM can be useful for studying treatment effects by borrowing of information, compared to stand-alone subgroup analyses with limited sample sizes in some strata. Our findings suggest that making

tumor-agnostic drug indication and coverage decisions simply based on the pooled treatment outcomes in a basket trial may not be appropriate in some tumor types. Both pooled and tumor-specific ORR, mPFS, and mOS posterior estimates may be useful in informing decision making. Finally, this approach permits quantifying the between-tumor heterogeneity across subgroups, characterizing the proportion of variance explained by the between-tumor variability, and offers further evidence on survival endpoints in an early-phase basket trial.

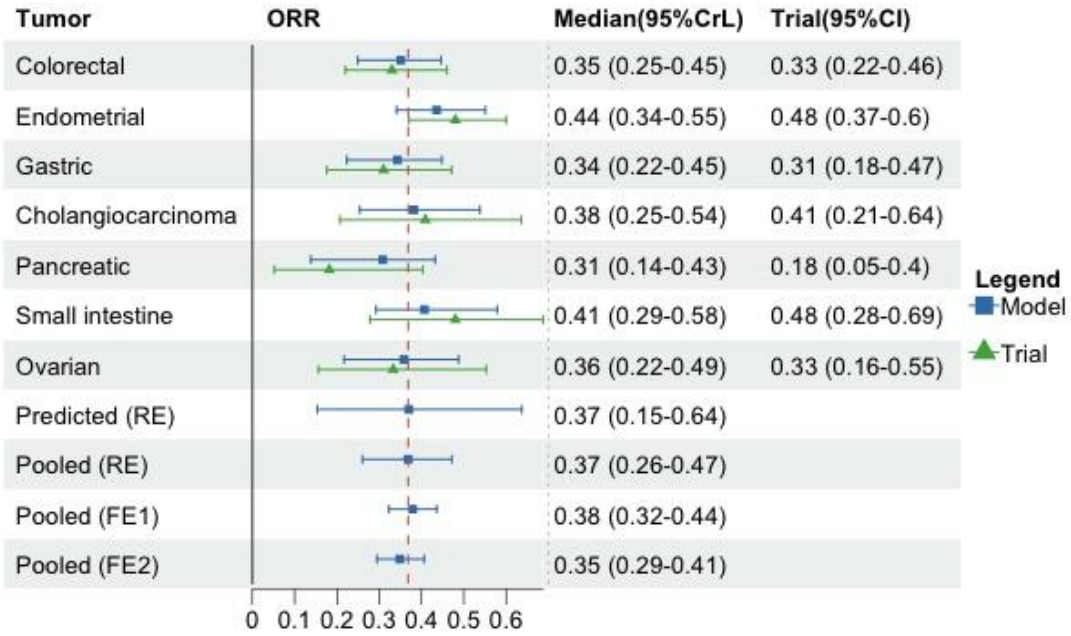
1.8 TABLES & FIGURES

Table 1.1. Trial-reported clinical outcomes among included patients

Tumor type	**Number of patients**	**ORR, % (95%CI)**	**Median PFS, months, (95%CI)**	**Median OS, months, (95%CI)**
Colorectal	63	33 (22, 46)	4.1 (2.1, 18.9)	33* (19.2, NR)
Endometrial	79	48.5 (36.2, 61)	13.1 (4.3, 34.4)	46.5* (27.2, NR)
Gastric	42	31 (17.6, 47.1)	3.2 (2.1, 12.9)	11 (5.8, 31.5)
Cholangiocarcinoma	22	40.9 (20.7, 63.6)	4.2 (2.1, 24.9)	19.4 (6.5, NR)
Pancreatic	22	18.2 (5.2, 4.03)	2.1 (1.9, 3.4)	3.7 (2.1, 9.8)
Small intestine	25	48 (27.8, 68.7)	23.4 (4.3, NR)	NR (16.2, NR)
Ovarian	24	33.3 (15.6, 55.3)	2.2 (2.0, 6.2)	33.6 (11.0, NR)
Brain	13	0 (0, 24.7)	1.1 (0.7, 2.1)	5.6 (1.5, 16.2)

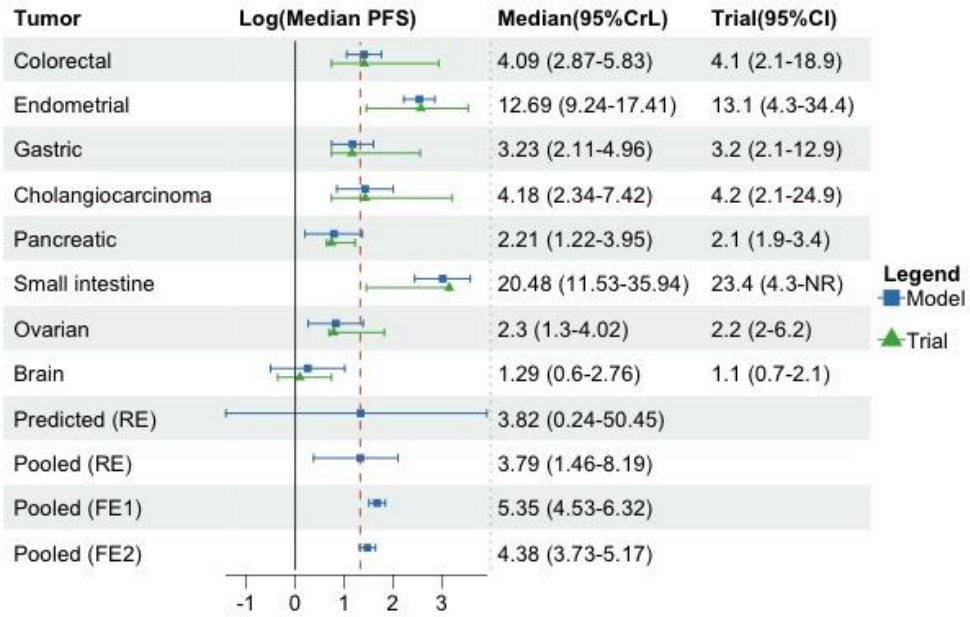
Abbreviations: CI, confidence interval; ORR, objective response rate; PFS, progression-free survival; OS, overall survival; NR, not reached at the end of study follow-up period.
* Estimated from the fitted exponential distribution to the digitized KM survival curve.

Figure 1.1. Summary of posterior distributions by tumor types for objective response rate



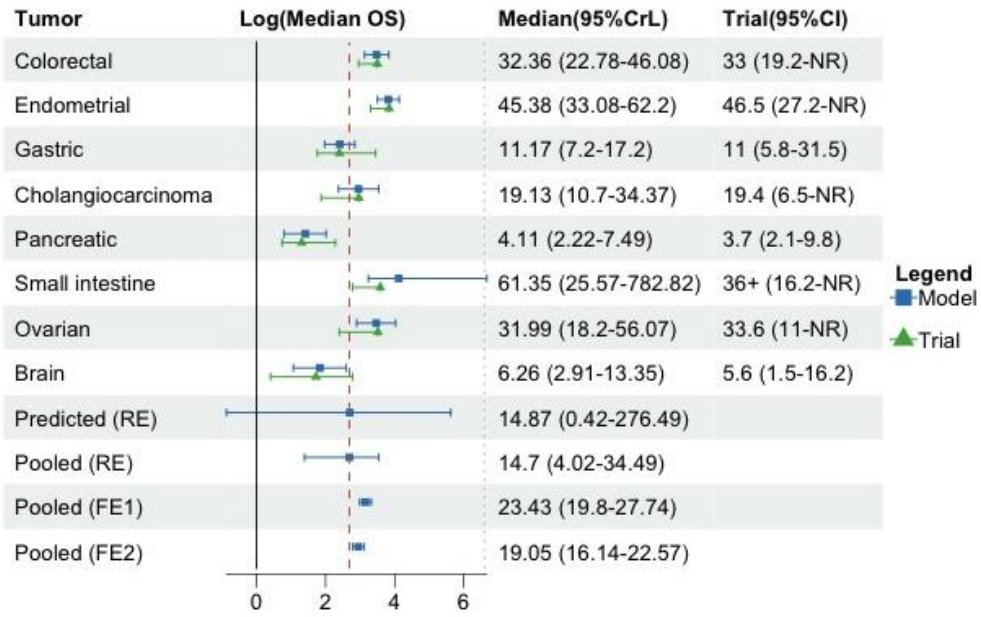
Abbreviations: ORR, objective response rate; RE, random-effects model; FE1, fixed-effect model with single mean effect; FE2, fixed-effects model with varying mean effects; CrL, credible interval; CI, confidence interval.
The red dashed line represents the mean pooled effect from the RE model.

Figure 1.2. Summary of posterior distributions by tumor types for median progression-free survival



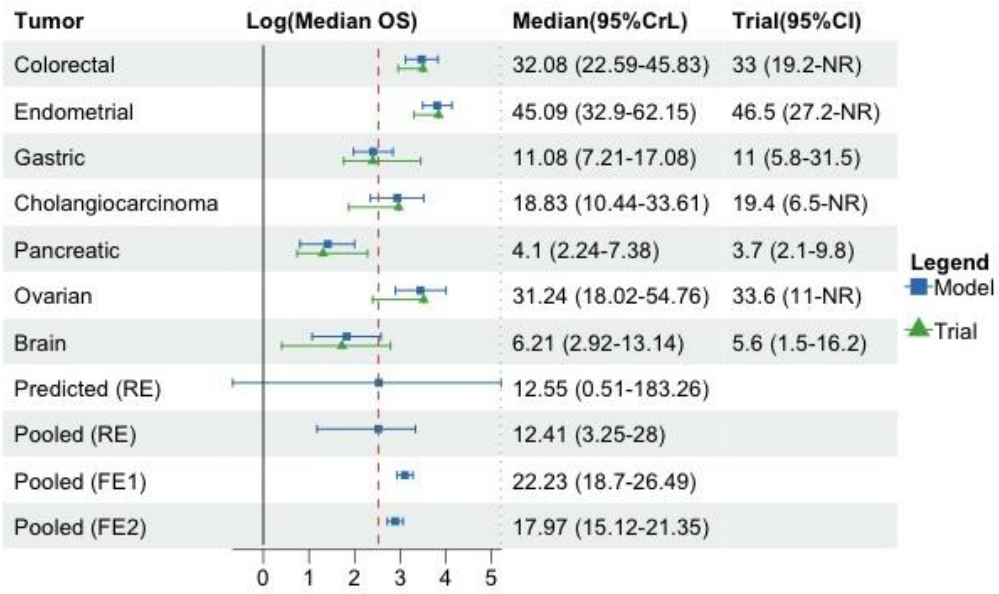
Abbreviations: PFS, progression-free survival; RE, random-effects model; FE1, fixed-effect model with single mean effect; FE2, fixed-effects model with varying mean effects; CrL, credible interval; CI, confidence interval; NR, not reached.
The red dashed line represents the mean pooled effect from the RE model.

Figure 1.3. Summary of posterior distributions by tumor types for median overall survival



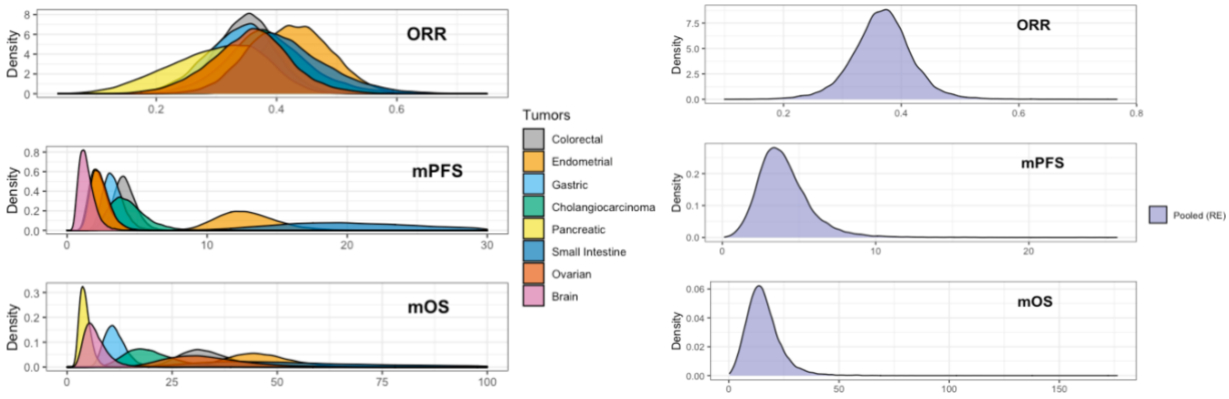
Abbreviations: OS, overall survival; RE, random-effects model; FE1, fixed-effect model with single mean effect; FE2, fixed-effects model with varying mean effects; CrL, credible interval; CI, confidence interval; NR, not reached; +, censored at the end of the study period.
The red dashed line represents the mean pooled effect from the RE model.

Figure 1.4. Summary of posterior distributions by tumor types for median overall survival, excluding small intestine cancer



Abbreviations: OS, overall survival; RE, random-effects model; FE1, fixed-effect model with single mean effect; FE2, fixed-effects model with varying mean effects; CrL, credible interval; CI, confidence interval; NR, not reached.
The red dashed line represents the mean pooled effect from the RE model.

Figure 1.5. Posterior density plots by tumor types and treatment outcomes obtained by fitting a BHM



Note: The upper limit of median PFS and OS outcomes for small intestine cancers were not shown in the figure to allow better visualization for the other tumor types.
Abbreviations: ORR: objective response rate; mPFS: median progression-free survival; mOS: median overall survival; RE: random-effects model.

1.9 SUPPLEMENT

Table S1.1. Estimated intraclass correlation metrics of random-effects models

Outcome	**Value**
Objective response rate	0.02 (0, 0.18)
Median progression-free survival	0.87 (0.42, 0.97)
Median overall survival	0.70 (0.33, 0.89)

Table S1.2. Sensitivity analysis for modeling the objective response rate under different choices for the prior distribution on the between-tumor SD

Median (95%CrL)	**ORR**	**Uninformative prior**			**Weakly informative prior**	
		Base case τ~Uniform(0, 3)	τ~Uniform(0, 5)	τ~Truncnorm(0, 100)	τ~Exp(1)	τ~Exp(0.3)
Colorectal	0.33	0.35 (0.25-0.45)	0.35 (0.25-0.45)	0.35 (0.26-0.45)	0.35 (0.25-0.45)	0.35 (0.25-0.45)
Endometrial	0.48	0.44 (0.34-0.55)	0.44 (0.34-0.55)	0.43 (0.34-0.54)	0.44 (0.34-0.55)	0.43 (0.34-0.55)
Gastric	0.31	0.34 (0.22-0.45)	0.34 (0.22-0.45)	0.35 (0.23-0.45)	0.34 (0.22-0.45)	0.34 (0.22-0.45)
Cholangiocarcinoma	0.41	0.38 (0.25-0.54)	0.38 (0.26-0.54)	0.38 (0.26-0.53)	0.38 (0.25-0.54)	0.38 (0.26-0.53)
Pancreatic	0.18	0.31 (0.14-0.43)	0.31 (0.14-0.43)	0.32 (0.15-0.44)	0.31 (0.14-0.43)	0.31 (0.14-0.43)
Small intestine	0.48	0.41 (0.29-0.58)	0.41 (0.29-0.58)	0.4 (0.29-0.57)	0.41 (0.29-0.58)	0.41 (0.29-0.58)
Ovarian	0.53	0.36 (0.22-0.49)	0.36 (0.22-0.49)	0.36 (0.23-0.48)	0.36 (0.22-0.49)	0.36 (0.22-0.49)
Predicted (RE)		0.37 (0.15-0.64)	0.37 (0.15-0.64)	0.37 (0.18-0.59)	0.37 (0.15-0.63)	0.37 (0.16-0.61)
Pooled (RE)		0.37 (0.26-0.47)	0.37 (0.26-0.47)	0.37 (0.27-0.46)	0.37 (0.26-0.47)	0.37 (0.27-0.47)
Between-tumor SD		0.34 (0.02-1.11)	0.35 (0.03-1.13)	0.29 (0.02-0.9)	0.35 (0.03-1.13)	0.33 (0.03-1.05)
DIC		40.1	40	40.2	40.2	40.2

Abbreviations: RE, random-effects model; SD, standard deviation; Exp, exponential distribution; Truncnorm, truncated normal distribution; ORR, objective response rate; DIC, Deviance Information Criterion.

Table S1.3. Sensitivity analysis for modeling median progression-free survival under different choices for the prior distribution on the between-tumor SD

Median (95%CrL)	**Median PFS**	**Likelihood: Lognormal Distribution**			
		Base case τ~Uniform(0, 5)	τ~Exp(1)	τ~Truncnorm(0, 4)	τ~Uniform(0, 3)
Colorectal	4.1	4.09 (2.87-5.83)	4.1 (2.88-5.8)	4.1 (2.9-5.83)	4.09 (2.87-5.82)
Endometrial	13.1	12.69 (9.25-17.41)	12.66 (9.22-17.33)	12.69 (9.26-17.36)	12.73 (9.31-17.5)
Gastric	3.2	3.23 (2.11-4.96)	3.24 (2.11-4.98)	3.22 (2.1-4.95)	3.23 (2.1-4.94)
Cholangiocarcinoma	4.2	4.18 (2.34-7.42)	4.18 (2.36-7.41)	4.16 (2.32-7.39)	4.16 (2.34-7.5)
Pancreatic	2.1	2.21 (1.22-3.95)	2.23 (1.25-4.01)	2.21 (1.23-3.97)	2.21 (1.23-3.97)
Small intestine	23.4	20.48 (11.53-35.94)	20.07 (11.4-35.11)	20.28 (11.62-35.75)	20.52 (11.75-35.86)
Ovarian	2.2	2.3 (1.3-4.02)	2.33 (1.33-4.03)	2.31 (1.31-4.05)	2.3 (1.31-4.02)
Brain	1.1	1.29 (0.6-2.76)	1.33 (0.62-2.84)	1.31 (0.61-2.77)	1.3 (0.6-2.8)
Predicted (RE)		3.82 (0.24-50.45)	3.83 (0.37-36.83)	3.83 (0.28-44.13)	3.82 (0.25-51.09)
Pooled (RE)		3.79 (1.46-8.19)	3.87 (1.7-7.87)	3.81 (1.55-7.99)	3.79 (1.51-8.23)
Between-tumor SD		3.23 (1.87-10.12)	2.83 (1.8-6.22)	2.86 (1.85-7.97)	2.97 (1.86-9.14)
DIC		9.6	10	9.6	9.5

Abbreviations: RE, random-effects model; SD, standard deviation; Exp, exponential distribution; Truncnorm, truncated normal distribution; PFS, progression-free survival; DIC, Deviance Information Criterion.

Table S1.4. Sensitivity analysis for modeling median overall survival under different choices for the prior distribution on the between-tumor SD

Median (95%CrL)	**Median OS**	**Likelihood: Lognormal Distribution**			
		Base case τ~Uniform(0, 5)	τ~Exp(1)	τ~Truncnorm(0, 4)	τ~Uniform(0, 3)
Colorectal	33*	32.36 (22.78-46.08)	32.17 (22.58-45.83)	32.3 (22.67-46.22)	32.31 (22.67-45.98)
Endometrial	46.5*	45.38 (33.08-62.2)	45.11 (32.96-62.13)	45.33 (33.04-62.48)	45.38 (32.93-62.35)
Gastric	11.0	11.17 (7.2-17.2)	11.22 (7.29-17.17)	11.16 (7.22-17.25)	11.16 (7.27-17.18)
Cholangiocarcinoma	19.4	19.13 (10.7-34.37)	19.12 (10.67-34.02)	19.06 (10.6-34.34)	19.11 (10.56-34.65)
Pancreatic	3.7	4.11 (2.22-7.49)	4.22 (2.32-7.65)	4.13 (2.27-7.52)	4.09 (2.23-7.49)
Small intestine	36+	61.35 (25.57-782.8)	55.41 (24.69-372.1)	59.53 (25.36-536)	60.72 (25.32-620.1)
Ovarian	33.6	31.99 (18.2-56.07)	31.8 (17.88-54.94)	31.93 (18.28-55.76)	31.85 (18.08-55.97)
Brain	5.6	6.26 (2.91-13.35)	6.51 (3.04-13.54)	6.35 (2.95-13.39)	6.27 (2.92-13.24)
Predicted (RE)		14.87 (0.42-276.5)	15.84 (1.13-172.4)	14.89 (0.72-222)	14.92 (0.59-254.5)
Pooled (RE)		14.7 (4.02-34.49)	15.55 (5.8-32.71)	15.03 (4.95-33.77)	14.79 (4.68-33.82)
Between-tumor SD		3.38 (1.92-19.78)	2.87 (1.82-8.31)	3.18 (1.9-11.47)	3.32 (1.91-12.75)
DIC		7.8	8.0	8.1	8.0

Abbreviations: RE, random-effects model; SD, standard deviation; Exp, exponential distribution; Truncnorm, truncated normal distribution; OS, overall survival; DIC, Deviance Information Criterion.

+, censored at the end of the study period; * Estimated from the fitted exponentiation distribution to the digitized KM survival curve.

Chapter 2.

Generating External Controls for Single-Arm Trials in Tumor-Agnostic Indications Leveraging Electronic Health Record Data Using Copula and Propensity Score Methods

2.1 ABSTRACT

OBJECTIVES: Single-arm basket trials represent significant challenges for assessing comparative effectiveness. We compared progression-free survival (PFS) and overall survival (OS) of patients with eight metastatic cancers who received pembrolizumab versus patients receiving NCCN-guideline-based chemotherapies.

METHODS: We applied inclusion/exclusion criteria from microsatellite instability–high (MSI-H) trials (Keynote-164 and Keynote-158) to TriNetx electronic health record database to derive chemotherapy control cohorts. Gaussian copulas were fitted to extract the dependence structure from the real-world controls, and trial samples were simulated by combining with marginal distributions of trial baseline covariates. Imbalances of known covariates were adjusted with inverse odds weighting. Median OS and PFS along with 95%CIs were compared between trial and real-world cohorts for all tumor types. Unweighted and weighted survival analyses were conducted for colorectal cancer (CRC) and endometrial cancer (EC) given available Kaplan-Meier survival curves.

RESULTS: Heterogenous median real-world OS (mrwOS, in months) and median real-world PFS (mrwPFS, in months) were observed among 9,267 patients with metastatic cancers who progressed on ≥1 line of systemic therapies. Compared to patients receiving pembrolizumab, patients receiving chemotherapy for colorectal (37.5 months extrapolated and not reached at 39

months vs. 34.3), endometrial (51.7 months extrapolated and not reached at 56 months vs. 36.3), cholangiocarcinoma (19.4 vs. 11.7), small intestine (not reached at 56 months vs. 21.8), and ovarian cancers (33.6 vs. 28.5) had shorter mrwOS, whereas patients with gastric (11 vs. 16.5), pancreatic (3.7 vs. 10), and brain cancers (5.6 vs. 10.3) had longer mrwOS. Similar patterns were observed for mrwPFS, which ranged from 2.8 to 4.7 months. Pembrolizumab was associated with more favorable PFS than chemotherapy for treating CRC and EC, respectively (CRC: weighted Hazard Ratio [HR], 0.46; 95%CI, 0.33-0.64; EC: weighted HR, 0.35; 95%CI, 0.27-0.46). However, MSI-H patients with CRC and EC receiving pembrolizumab did not demonstrate significantly better overall survival than patients receiving SoC chemotherapies (CRC: unweighted HR, 1.01; 95%CI, 0.68-1.51; weighted HR, 1.18; 95%CI, 0.78-1.80; EC: unweighted HR, 0.80; 95%CI, 0.55-1.17; weighted HR, 0.85; 95%CI, 0.57-1.28). CONCLUSIONS: Pembrolizumab was associated with significant prolonged PFS but not OS versus chemotherapy in patients with metastatic CRC or EC. Incorporating external control data in early phase trials may provide a more comprehensive understanding of treatment effects of tumor-agnostic drugs than relying solely on single-arm trials. This study provided external comparators for efficacy evidence from single-arm tumor-agnostic trials, which can be used to assess the cost-effectiveness of TAD.

2.2 INTRODUCTION

Randomized controlled trials (RCTs) are considered the “gold standard” when evaluating the efficacy and safety of a new intervention. Oncology drug approvals in the United States are increasingly supported by single-arm clinical trials, which facilitates faster patient access to novel therapies. The Accelerated Approval program by US Food and Drug Administration

(FDA) has allowed approvals of precision oncology drugs based on surrogate endpoints from single-arm trials [54]. Among the approved 563 new oncology drug indications by FDA between 2002 and 2021, 31% were based on data from single-arm trials. There was a substantial increase in approvals supported by single-arm trials in the years 2017-2021 compared with previous years [55]. Tumor-agnostic drugs (TAD), typically approved based on single-arm basket trials, represent significant challenges for assessing comparative effectiveness in the absence of comparators. Such trials typically require comparators from external data sources to demonstrate comparative efficacy and safety of an intervention. The assessment of value is further compounded by the fact that the standard of care (SoC) varies by tumor type, therefore the relative effect of a TAD against SoC will likely differ for each tumor type.

Intrapatient comparisons and external controls have been used to demonstrate comparative effectiveness of indications approved based on single-arm trials. In an intrapatient comparison, patients with prior systemic therapies within the same trial can serve as their own controls, allowing direct comparison of different treatments within the same patient over time. However, such comparisons can only assess intermediate outcomes, such as objective response rate and time to discontinuation on the most recent prior therapy [56]. Conversely, an external control involves a sample of patients from another source that is comparable to the clinical trial population's treatment arm. Historical data from earlier clinical trials (historical control), or from similar patient cohorts to single-arm trial cohorts, have been utilized to assess comparative effectiveness outcomes among a "real-world" control [57]. For instance, researchers have compared efficacy or safety of a single-arm interventional oncology trial with outcomes in historical patient cohorts previously treated with a comparator drug for the same indication at the

same institution [58-61]. Nevertheless, both intrapatient comparison and historical control from earlier clinical trials may suffer from small sample sizes after applying patient selection criteria.

External controls from real-world data (RWD) sources reflect patient treatment journey in routine clinical practice, including electronic health record (EHR) or registry data [62]. It can be a useful approach to evaluating the comparative effectiveness of an intervention when clinical trial control data are not available. Compared to clinical trial data, RWD can select a contemporaneous cohort from a relatively large pool of patients. Additionally, RWD can provide valuable information regarding the effectiveness of therapies administered during routine care, which can be used as a benchmark or a comparator for the target population of a single-arm trial. This is especially useful for rare and novel indications with little historical or published evidence. Furthermore, the use of external controls in early-phase single-arm trials can provide a better understanding of treatment effect, to inform assumptions for later-phase or confirmatory trials [63].

RWD derived external cohorts is increasingly included in regulatory or health technology assessment submissions in the U.S. and Europe as complementary to clinical trial data [64-68]. Under the 21st Century Cures Act, the FDA has developed a RWD regulatory framework to support the approval of new indications for previously approved drugs and to support or fulfill post approval requirements [69]. A recent systematic review has summarized 43 cases in which external controls were used to support regulatory approval decisions with non-randomized studies designs [70]. In addition to the industry guidance on external controls [71, 72], other working groups have provided guidance and suggestions for improving the methods to address the challenges in selecting external controls, such as potential bias related to patient selection,

confounding arising from the lack of randomization, and missing data elements from typical patient EHRs [62, 70, 73-75].

The objective of this chapter was to construct external controls for eight tumor types to be used for assessing comparative effectiveness and cost-effectiveness. We aimed to apply the strictest possible criteria to both data and methodology, to enhance comparability of the external control cohort to the clinical trial cohort and minimize important biases. Furthermore, we aimed to address the issue of varied SoC therapies for a variety of tumor types in a basket trial. We compared progression-free survival (PFS) and overall survival (OS) of patients with eight metastatic cancers who received pembrolizumab in KEYNOTE-164 and KEYNOTE-158 versus patients receiving NCCN-guideline-based chemotherapies in the clinical practice.

2.3 METHODS

Figure S2.1 provides an overview of the study processes.

Data Sources

A total of 282,135 patients with a diagnosis of colorectal cancer (CRC) and 402,059 non-CRC patients aged 18 years or older were identified in the TriNetx database before applying any cohort selection criteria. At the data cut-off, TriNetx EHR databases cover over 70 healthcare organizations providing continuous, comprehensive, up-to-the-month data with more than 116 million patient lives in the U.S. In addition to data from TriNetx EHR databases, we extracted aggregated summary data and available Kaplan-Meier (KM) survival curves from KEYNOTE-164 and KEYNOTE-158 trials.

Generate External Controls

The eligibility criteria of the KEYNOTE 164 and KEYNOTE 158 trials were compared and applied if available in the TriNetx databases (Table 2.1). Filtering the RWD by the trial inclusion and exclusion criteria aims to mimic the clinical trial population and reduce baseline imbalances between the data sources. Eight external controls for eight tumor types were obtained after applying trial inclusion and exclusion criteria to the TriNetx data. The tumor types were brain, cholangiocarcinoma, CRC, gastric, ovarian, small intestine, pancreatic, and brain tumors. Patients were included if they: 1) had at least two separate ICD-9-CM or ICD-10-CM diagnosis codes of included cancers. The diagnosis date was between January 1, 2010, and March 31, 2019, for CRC, and between January 1, 2010, and December 31, 2019 for patients with non-colorectal cancers (non-CRCs); 2) had a metastasis code on or after the initial diagnosis date. The metastasis code was defined by secondary malignant neoplasm ICD-9-CM or ICD-10-CM codes; 3) had at least one line of standard of care systemic therapies after their first metastasis date (index metastasis date). In addition, CRC patients must have previously failed treatment with fluoropyrimidine, oxaliplatin, and irinotecan. The SoC systemic therapies were defined by reviewing NCCN guidelines for the treatment of eight metastatic cancers and further validated by an experienced oncology pharmacist (Table S2.1). We included all eligible chemotherapies on the market by the time of the analysis; and 4) aged ≥18 years old at the initial diagnosis date.

To derive time to progression, we used an iterative process to apply pre-defined line of therapy algorithms in the medication dataset. The algorithms were refined based on a targeted literature review of existing line of therapy identification algorithms in oncology [76-80]. The details of the algorithms can be found in Figure S2.2. The start date of the first line was defined as the medication start date of any systemic treatment following the metastatic diagnosis. The medication start date was the date of medication order, prescription, or administration. The end

date of a line was defined as either the most recent medication dispense date plus one drug administration cycle, based on dosing schedules, before advancing to another line of therapy or loss to follow-up. To define index date, we randomly selected a line of therapy (the index line) from the available lines of therapy for each patient [81]. We defined time to progression as the period from index date to either treatment discontinuation, the day before the next line of therapy, or the date of death, whichever occurred first.

The exclusion criteria included: 1) patients who were not on chemotherapies at the index line of therapy (e.g., targeted therapies, immunotherapies); 2) patients who did not initiate their first-line metastatic treatment within 90 days following their index metastasis date; 3) patients who had any prior treatment with anti-PD-1, -PD-L1, or -PD-L2 therapies; 4) patients with certain comorbidities at baseline, including active CNS metastases, carcinomatous meningitis, pneumonitis, HIV, Hepatitis B or C infections, or autoimmune diseases. To exclude active comorbidities that require treatments, the baseline period was defined as within one month prior to the first metastasis treatment date. Diagnosis of comorbidities were defined by ICD-9-CM or ICD-10-CM codes. A final cohort of 9,267 patients with metastatic cancers were identified as comparator groups after applying all eligibility criteria. The cohort attrition details can be found in the Figure S2.3.

Copula Models and Simulation

After cohort selection, we fitted a Gaussian Copula and estimated the copula parameters to extract the dependence structure from the eight real-world controls (rwControls), respectively. A copula is a function which joins or "couples" a multivariate distribution function to its one-dimensional marginal distributions [1]. Variable correlations are specified via the covariance

matrix. To generate multivariate pseudo trial samples using the inverse cumulative density function (CDF) transformation, it works by first transforming the normal marginals to the uniform distributions by applying the normal CDF transformation to each component. Finally, the uniform marginals could be transformed to any distributions by applying the inverse CDF. The Sklar's theorem states that every joint distribution can be written as a copula and a set of marginal distributions. If the marginals are continuous, then the copula is unique.

Age, sex, race, prior (neo)adjuvant therapy, and prior lines of therapy were included in the Copula to extract the dependence structure in the eight real-world controls, respectively. Since discrete variables (i.e., binary or categorical) cannot be directly modeled by the Copula, a continuous extension was applied to discrete random variables before fitting the Copula.

Let $V_1, \dots, V_m$ be independent $Uniform(0,1)$ random variables, independent of X. Consider $Y_j = X_j + V_j - 1, Y = (Y_1, \dots Y_m)$, marginals of Y_j will be uniformly distributed. When adding -1 or another value will generate the same Copula model and correlation structure [82]. Figure S2.3 showed the CDF of a binary variable before and after its continuous extension. The details of the continuous extension process can be found in the Supplement File. We then fitted the Gaussian Copula and estimated the copula parameters to extract the dependence structure from the rwControls. To avoid the noise introduced by adding the V_j, the parameter estimates were averaged over a set of estimates from 100 runs [83].

Finally, combined with marginal distributions of baseline covariates in the trial (the same covariates used for Copula estimation), we were able to simulate correlated datasets with the sample size equal to that in the trial while maintaining the dependence structure of patient characteristics in the real-world.

Propensity Scores

Logistic regression was applied to the combined datasets (the simulated trial samples and the rwControls) to predict the probability of being enrolled in the clinical trial given individuals' characteristics. Propensity scores were estimated by regressing covariates against the data source indicator (i.e., clinical trial vs TriNetx EHR). We then used the nearest neighbor approach with replacement to find the closest match for each treated patient from the untreated group by using their propensity scores and a caliper equal to one-fifth of the standard deviation (SD) of the logit of the propensity score. The covariates were selected based on scientific knowledge on potential factors that are associated with the outcome of interest, or with both the treatment assignment and the outcome [84]. We included polynomial term of age to capture non-linear relationships between age and the treatment assignment. The specification of the propensity score model is:

$$logit(\Pr(Trt)) = \beta_0 + \beta_1 * sex + \beta_2 * age + \beta_3 * age^2 + \beta_4 * priorAdjuvantTherapy + \beta_5 * priorLinesOfTherapy$$

A weight was estimated for each individual in the rwControls as equal to the inverse odds of being enrolled in the external control vs. trial groups, using estimated propensity scores. For an individual patient i who is assigned to treatment T_i (0=external control and 1=trial), the inverse odds weight is represented by: $w_i = T_i + (1 - T_i)\frac{e_i}{1-e_i}$, where e_i is the estimated propensity score [85]. The weighting scheme aims to make patients in the rwControls more similar to those in the trial [2].

We examined the overlap in the distribution of propensity scores for the simulated trial samples and the rwControls. The covariates pre- and post-matching were compared between the treated and untreated groups using standardized differences (Figure S2.5). Applying weights

produced mean estimates of the baseline characteristics of rwControls that more closely resembled those of the trial groups.

In addition, we performed a repeated sampling process to account for the uncertainty in the simulation for the above-mentioned steps, including: 1) Copula fitting and parameter estimation; 2) simulation of one trial sample; (3) generate inverse odds weights for each patient. These steps were repeated for 500 times, and we then the set of 500 weights were averaged to obtain the final weights. The entire process was performed for eight tumor types to better balance covariates between simulated trial samples and rwControls, respectively.

Survival Analysis

To reconstruct individual patient survival data from the trials, we first digitized the available KM survival curves from KEYNOTE 164 and KEYNOTE 158 trial publications using previously developed methods [86]. The input included numbers at risk R(t) at time t, the total number of patients, and survival probability S(t), while the output generated time and status of being censoring or survival. This produced a sample of survival data with the same size as the treated group, with individuals assigned a time of death or censoring based on the approximated counts of events and patients at risk. We compared the reconstructed KM survival curves and the original published KM survival curves. Median survival outcomes, and survival rates at 12 months and 24 months of the reconstructed survival data were compared to those in the trials.

PFS was not available in the TriNetx data, so we defined time to progression as the period from the start date of the index line to either the end date of that line plus one drug cycle, the day before the next line of therapy, or the date of death, whichever occurred first. Furthermore, we defined time to death as the time elapsed from the start date of the index line to

the date of death due to any cause. Patients who were lost to follow-up or without the relevant event by the end of the study period were censored at the date of their last confirmed activity date.

Median OS and median PFS were compared between pembrolizumab and rwControls for all tumor types. Cox proportional hazard (PH) models was fitted to the combined survival data, including an indicator for pembrolizumab or control to capture the relative effect. Imbalances of known covariates were adjusted with inverse odds weighting. Specifically, the pseudo individual survival data from the pembrolizumab group were assigned a weight of 1 and combined with patients from the rwControl which was assigned the inverse odds weights derived from the previous step. We obtained adjusted and unadjusted hazard ratio (HR), 95%CIs, and robust standard errors. Proportional hazard assumption was assessed by examining whether there was a similar trend without crossing in the KM survival curves and testing of scaled Schoenfeld residuals. Weighted Cox proportional hazard regressions and accelerated time failure (AFT) models were only conducted for CRC and endometrial cancer (EC) given available Kaplan-Meier survival curves. All analyses were conducted in R 4.1.2 with copula, WeightIt, survival, and flexsurv packages (The R Core Team).

2.4 RESULTS

Baseline Characteristics of External Controls

After applying inclusion and exclusion criteria, a total of 9,267 patients who were treated by chemotherapies and have progressed on at least one line of therapy were included in the final analytical sample. The majority of patients were diagnosed with colorectal cancer (n=3,104, 33.5%), followed by ovarian (n=2,178, 23.5%), pancreatic (n=1,514, 16.3%), endometrial

(n=1,089, 11.8%), cholangiocarcinoma (n=421, 4.5%), gastric (n=524, 5.7%), brain (n=224, 2.4%), and small intestine cancer (n=213, 2.3%). The mean age was 62.5 years old (SD=11.8, median=63.0, min=18.0, max=90.0) among the overall patients. Most of patients were female (n=5,912, 63.8%), white (n=6,965, 75.2%), and only had one prior line of therapy (n=7,593, 81.9%). Less than a quarter of patients had prior neoadjuvant or adjuvant therapies (n=1,803, 19.5%). Table 2.2 compared the differences in baseline characteristics between trial cohorts and rwControls. The details of baseline characteristics by tumor types are summarized in Table S2.2.

Simulating Trial Samples and Propensity Score Estimation

The correlation matrices of the covariates were shown in Figure S2.6. Most variables did not demonstrate strong correlations between each other. After performing the above-mentioned steps for trial simulation and propensity score estimation for 500 times, we obtained 500 sets of inverse odds weights. The average weights and the weights from the last iteration were saved for later use. Balance diagnostics showed substantial improvement after inverse odds weighting, with similar distributions of covariates and propensity scores between simulated trial and rwControls. The covariate balance plot illustrates whether weighting reduces baseline imbalances between the two data sources. After weighting, baseline imbalance was improved for all covariates with the standardized mean differences less than 0.1 (Figure S2.5).

Comparative Median PFS and Median OS

Heterogenous median real-world OS (mrwOS, in months) and median real-world PFS (mrwPFS, in months) were observed among 9,267 patients with metastatic cancers who progressed on ≥1 line of systemic therapies. Compared to patients receiving pembrolizumab in

the trial, patients receiving chemotherapy for colorectal (37.5 months extrapolated and not reached at 39 months vs. 34.3), endometrial (51.7 months extrapolated and not reached at 56 months vs. 36.3), cholangiocarcinoma (19.4 vs. 11.7), small intestine (not reached at 56 months vs. 21.8), and ovarian cancers (33.6 vs. 28.5) had shorter mrwOS. On the contrary, patients with gastric (11 vs. 16.5), pancreatic (3.7 vs. 10.1), and brain cancers (5.6 vs. 10.3) had longer mrwOS. The 95%CIs of such estimates are shown in Table 2.3.

The median PFS ranged from 2.8 to 4.7 months across eight metastatic cancers in the rwControls compared to 1.1 to 23.4 months in the trial. Similarly, patients receiving chemotherapy for endometrial (13.1 vs. 3.2) and small intestine (not reached at 56 months vs. 21.8) cancers had shorter mrwPFS, compared to patients receiving pembrolizumab in the trial. Patients with other cancers showed similar median PFS between pembrolizumab vs real-world chemotherapies, except for pancreatic (2.1 vs. 4.2) and brain cancers (1.1 vs. 4.7) (Table 2.3).

The tumor specific survival curves for the probability of survival and the probability of progression-free within 5 years from index date among real-world patients receiving chemotherapies are shown in Figure 2.1.

Survival Rates and Survival Analysis for Colorectal and Endometrial Cancers

The survival rates for OS and PFS at 12, 24, and 36 months are shown in Table 2.4. Compared to rwControls, patients receiving pembrolizumab for both colorectal and endometrial cancers had similar overall survival rates at 12 months but surpassed at 24 months. More patients receiving pembrolizumab with endometrial cancer survived at 36 months compared to rwControl. The PFS rates were consistently higher at 12, 24, and 36 months for both colorectal and endometrial cancers (Table 2.4).

In addition, estimates from both weighted and unweighted Cox PH models have suggested that pembrolizumab was associated with more favorable PFS than chemotherapy for treating CRC among the treated (unweighted HR, 0.68; 95%CI, 0.49-0.93; weighted HR, 0.46; 95%CI, 0.33-0.64) and EC (unweighted HR, 0.54; 95%CI, 0.40-0.71; weighted HR, 0.35; 95%CI, 0.27-0.46), respectively. However, MSI-H patients with CRC and EC receiving pembrolizumab did not demonstrate significantly better overall survival than patients receiving real-world chemotherapies (CRC: unweighted HR, 1.01; 95%CI, 0.68-1.51; weighted HR, 1.18; 95%CI, 0.78-1.80; EC: unweighted HR, 0.80; 95%CI, 0.55-1.17; weighted HR, 0.85; 95%CI, 0.57-1.28) (Figure 2.2, Figure 2.3). Estimates from accelerated failure time models were similar to those obtained from Cox PH models (Table 2.4).

2.5 DISCUSSION

This study provided external comparators for efficacy evidence from single arm pembrolizumab trials. Creating external controls using RWD can be a viable approach to evaluate the comparative effectiveness of tumor-agnostic drugs assessed through single-arm, non-randomized basket trials while addressing the issue of differing standard of care treatments. In the absence of individual-level trial data, we simulated trial samples including baseline covariates using the same dependence structure as the eight real-world controls. The main finding of our survival analyses suggests that MSI-H patients receiving pembrolizumab in the trials had more favorable PFS but not prolonged OS than real-world patients receiving chemotherapy for treating CRC and endometrial cancer. We also derived time to progression and time to death for external controls by tumor types, which could serve as comparators in the cost-effectiveness analysis of pembrolizumab in a tumor-agnostic indication.

After adjusting using the inverse odds weights, the Kaplan-Meier survival curves for OS in CRC has shifted upwards compared to unadjusted OS curves. This is because the real-world control population consisted of CRC patients who had only one prior line of therapy, were male, and older compared to those who had more than one line of therapy, were female, and younger. Before any adjustment, these covariates had skewed distributions towards the groups with worse prognosis (Figure S2.5). Our study, consistent with other studies, found that RWD patients were shown to be older, which is known to be associated with worse prognosis for cancer patients in general [87].

Challenges of Establishing External Control Arms

Our study faced several challenges in constructing the external control arms. The first challenge involved selecting the appropriate SoC arms. To establish comparative effectiveness in real-world clinical practice, we compared patients receiving pembrolizumab in the KEYNOTE 158 and KEYNOTE 164 trials to a real-world cohort receiving all available chemotherapies. Two alternative approaches have been used in similar studies, including a specific regimen approach and an approach that includes all available therapies. For instance, Coleman et al. compared the overall survival of patients with MSI-H/dMMR receiving dostarlimab in the GARNET trial with a comparable real-world cohort receiving non-anti-PD-L1 treatments [88]. However, the authors did not report the distribution of chemotherapies versus targeted therapies. Moreover, some researchers have employed a RCT approach by selecting patients from the EHR cohort who received standard-of-care treatment as identified in the trial. Carrigan et al. explored this approach using Flatiron Health Databases to replicate control arms for nine lung cancer RCTs [89]. Two studies established comparative effectiveness for head-to-head comparison

trials between alectinib and ceritinib using data extracted from clinical trials and an EHR database, respectively [90, 91]. The choice of comparator should depend on the study's objectives and generalizability, which should be planned ahead. Additionally, researchers should keep in mind that treatment practices may change over time, necessitating contemporaneous comparators.

The second challenge we encountered was missing variables in RWD and the development of key variables. Some essential variables such as lines of therapy, ECOG score, and PFS were not available in TriNetx EHR. Although ECOG score could not be constructed from existing data elements, we adapted algorithms from previous literature to develop lines of therapy and time to progression variables. Notably, prior studies applied different algorithms to define the lines of therapy variable, which often depend on the tumor type and the assumption about the progression of therapy line [76-78, 80]. Our sensitivity analysis indicated that there was little impact of a 90-day and 120-day gap between medication dates on survival outcomes. To define the proxy for PFS, we used the time to discontinuation (TTD) of the most recent line of therapy or death, whichever occurs first. Similarly, there is no unified approach to define the proxy for time to progression in RWD. Previous studies comparing TTD, time to next line of treatment (TTNT), and PFS have found that TTD tends to be a lower-bound surrogate outcome for PFS, but TTNT usually exceeds PFS [92, 93]. For example, in assessing the comparative efficacy of another tumor-agnostic drug, entrectinib, TTD for any reason was defined as the time from the start of entrectinib until the end of the therapy, and PFS was defined as the time from the first dose of entrectinib to the first documented disease progression or death, whichever occurred first. The median PFS of patients receiving entrectinib was 11.2 months, slightly longer than TTD, 9.9 months (HR: 1.08 (95%CI: 0.6-1.9)) [56].

Alternative Designs for Comparative Effectiveness

In addition to external controls using real-world data, alternative approaches for comparative efficacy such as intrapatient comparisons and hybrid control can be expected to play a prominent role in future drug development. The hybrid control design incorporates external control data into an augmented or hybrid design, in which external data supplements the concurrent RCT control arm. [63, 94-96]. For example, Ventz et al. used both data from previously conducted clinical trials as well as RWD for an externally controlled trial to understand comparative efficacy on OS in glioblastoma multiforme [97]. The use of several sources of external data can reduce potential bias and retain sufficient power. Additionally, intrapatient comparisons have been used to establish comparative efficacy of tumor-agnostic therapies. For instance, entrectinib was approved by FDA and EMA with a tumour-agnostic indication for NTRK fusion-positive cancers based on single arm trials. Researchers used the patients' prior systemic therapies prior to commencing entrectninb as a comparator arm to generate evidence for comparative effectiveness. However, the sample size was small (n=51), which may not be representative of real-world patients with NTRK fusion-positive cancer. The median PFS on entrectinib was 11.2 months compared to median TTD (2.9 months) on most recent prior therapy [56]. Another intrapatient analysis compared Larotrectinib with prior lines of therapy among patients with NTRK fusion-positive cancers (n= 72) [98]. Intrapatient analyses have the advantage of eliminating between-patient variability by using each patient as their own control. However, the baseline clinical characteristics, such as comorbidities and stage, may differ by various lines of therapies.

2.6 LIMITATIONS

There are several limitations associated with this Aim. First, the lack of survival curves by tumor types in the KEYNOTE 158 trial has limited our capability to directly match controls for individual tumor types. However, since the inclusion/exclusion criteria are the same for non-CRC tumors in the KEYNOTE 158 trial, selecting patients at the individual tumor level will yield the same cohort as selecting at the aggregated level. Second, some clinical variables, such as the ECOG score, which are consistently used in the trial setting, may not be reliably recorded or used in routine clinical practice. The TriNetx dataset lacks information on the MSI-H/dMMR status due to the low prevalence of MSI-H and low uptake of testing in routine clinical practice. The MSI-H/dMMR status is known to have better prognosis only among Stage II CRC patients and has been found in 5~15% of these cancer types [99]. However, there is no conclusive evidence on whether MSI-H/dMMR status affects underlying prognosis among metastatic/advanced cancer patients. Prior literature showed that MSI-H/dMMR status is not a significant prognostic factor for non-anti-PD-(L)1/2 treatments in the advanced/recurrent setting [100]. Future studies need to validate the assumption that MSI-H/dMMR mutations do not significantly impact underlying prognosis among patients receiving chemotherapies in these tumor types. The effectiveness results need to be interpreted as the average treatment effect in the external control population. Although the line of therapy was not initially available in the TriNetx database, we derived the variable based on pre-defined algorithms. As prior lines of therapy distributions were not available by tumor types, the distribution in the external control population may not have precisely matched the KEYNOTE trial population. Finally, in trials, treatment for metastatic cancers is often administered until progression or toxicity, and a finite number of cycles of treatment are usually planned, which may lead to treatment stopping even in

responding patients before progression. In such cases, the TTD endpoint may underestimate PFS if discontinuation happens for reasons other than progression.

2.7 CONCLUSION

Patients with metastatic MSI-H CRC and endometrial cancers who received pembrolizumab in the trials demonstrated more favorable PFS, but not prolonged OS compared to real-world patients receiving chemotherapy. The methodology employed in this chapter has wide implications beyond tumor-agnostic drugs, such as in single-arm trials for rare diseases. Incorporating external control data in early phase trials may provide a more comprehensive understanding of treatment effects of tumor-agnostic drugs than relying solely on single-arm trials. Oncology marketing applications can leverage external control data to establish natural history of the disease, provide descriptive information regarding treatment effects on subgroups, and provide supportive evidence of direct comparisons to the experimental arm.

2.8 TABLES & FIGURES

Table 2.1. KEYNOTE trial inclusion/exclusion criteria applied to the real-world dataset to identify eligible patients

Categories	KEYNOTE 164 & KEYNOTE 158	TriNetx EHR
Diagnosis	Colorectal (CRC): previously treated histologically proven, locally advanced, unresectable or metastatic MSI-H CRC. Non-CRC: with a histologically/cytologically confirmed advanced—unresectable and/or metastatic—incurable MSI-H noncolorectal solid tumor	• Include patients diagnosed with one of the eight malignancies using ICD-9-CM/ICD-10-CM codes. • Patients were required to have at least two separate diagnosis codes in the EHR records. • Include patients with metastatic diseases identified via secondary malignant neoplasm ICD-9-CM/ICD-10-CM codes. • Include patients with 1+ metastasis codes on or after the diagnosis date.
Diagnosis		Restrict diagnosis date to before December 31, 2019 for non-CRC tumors and before Mar 31, 2019 for CRC to allow for at least two years of follow-up after initial diagnosis.
Prior treatment	CRC: include patients with >=1 prior line of therapy that has progressed following treatment with a fluoropyrimidine, oxaliplatin, and irinotecan (combination therapy). Non-CRC: include patients with disease progression on or intolerance to >=1 prior standard therapy	• Include patients with at least 1 prior standard of care systemic therapies after the first metastasis date. Prior systemic therapies include chemotherapies, targeted therapies, immunotherapies, and hormonal therapies. All drugs were identified from NCCN guidelines. • For CRC cohort, include patients with at least 1 prior systemic therapy after the index metastasis date including fluoropyrimidine. • After generating lines of therapies based on algorithms, randomly select a line of therapy as the follow-up starting line (treatment initiation date of that line as index date). • Exclude patients with only 1 prior line of systemic therapy. • Exclude patients using non-chemotherapies at the index date (chemotherapy as comparator). • Exclude patients who didn't initiate their 1st systemic therapy

		within 90 days following the first metastatic diagnosis.
Prior treatment	Exclude patients with prior anti-PD-1, programmed death ligand 1 (PD-L1), or PD-L2 therapy	Exclude patients with prior anti–PD-1, –PD-L1, or –PD-L2 therapy before the index date
Prior treatment	• No prior monoclonal antibody, chemotherapy, targeted small-molecule therapy, or radiation therapy within 2 weeks of study start • No live vaccine 30 days or less before study treatment	Not applicable
Age	Include patients aged >= 18 years	Exclude patients with age < 18 years old at index date
Comorbidities	Exclude certain comorbidities: Active autoimmune disease; active malignancy requiring treatment. active infection requiring systemic treatment; known history of HIV, interstitial lung disease, or active noninfectious pneumonitis; and active hepatitis B or C virus infection	Exclude comorbidities within one month before the index date, including diagnosis of: • CNS metastases • noninfectious pneumonitis, HIV, hepatitis B or C infection • Autoimmune diseases
Baseline Status	• >= 1 measurable lesion per RECIST version 1.1 (v1.1) an Eastern Cooperative Oncology Group performance status of 0 or 1 • Adequate organ function	Not applicable

Table 2.2. Baseline characteristics for the trial cohorts versus real-world external control arms

Characteristics	KEYNOTE 164 (CRC)	Real-world cohort (CRC)	KEYNOTE 158 (non-CRC)	Real-world cohort (non-CRC)
	N=63, n (%)	N=3,104, n (%)	N=233, n (%)	N=6,163, n (%)
Age, years				
Median (range)	59 (23-83)	61 (20-90)	60 (20-87)	65 (18-90)
Female	30 (52%)	1346 (43.4%)	137 (59%)	4566 (74.1%)
Performance status				
0	22 (35%)	NA	113 (49%)	NA
1	41 (65%)	NA	120 (52%)	NA
Prior lines of therapy				
0	0 (0%)	0 (0%)	7 (3%)	0 (0%)
1	24 (38%)	2,552 (82.2%)	87 (37%)	5,041 (81.8%)
2	20 (32%)	355 (11.4%)	61 (26%)	648 (10.5%)
≥3	19 (30%)	197 (6.3%)	78 (34%)	474 (7.7%)
Prior (neo)adjuvant therapy	17 (27%)	653 (21.0%)	55 (24%)	1,150 (18.7%)
Tumor types				
Colorectal	63 (100%)	3,104 (100%)	0 (0%)	0 (0%)
Endometrial	0 (0%)	0 (0%)	49 (21%)	1,089 (17.7%)
Gastric	0 (0%)	0 (0%)	24 (10%)	524 (8.5%)
Cholangiocarcinoma	0 (0%)	0 (0%)	22 (9%)	421 (6.8%)
Pancreatic	0 (0%)	0 (0%)	22 (9%)	1,514 (24.6%)
Small intestine	0 (0%)	0 (0%)	19 (8%)	213 (3.5%)
Ovarian	0 (0%)	0 (0%)	15 (6%)	2,178 (35.3%)
Brain	0 (0%)	0 (0%)	13 (6%)	224 (3.6%)
Other	0 (0%)	0 (0%)	69 (30%)	0 (0%)
Race				
Asian	14 (22%)	92 (3.0%)	NA	140 (2.3%)
Black/African American	7 (11%)	453 (14.6%)	NA	798 (12.9%)
White	42 (67%)	2,298 (74.0%)	NA	4,667 (75.7%)
Other	0 (0%)	261 (8.4%)	NA	558 (9.1%)

CRC, colorectal cancer; NA, not available.

Table 2.3. Comparison of median PFS and OS between pembrolizumab and chemotherapies

Tumor type	**N**	**Median OS (95%CI)**		**Median PFS (95%CI)**	
		Trial	**RW Control**	**Trial**	**RW Control**
	(RW Control)	**(Pembro)**	**(Chemo)**	**(Pembro)**	**(Chemo)**
Colorectal	3,104	37.5* (19.2, NR)	34.29 (31.53, 39.45)	4.1 (2.1, 18.9)	4.37 (4.14, 4.67)
Endometrial	1,089	51.7* (27.2, NR)	36.26 (29.42, NR)	13.1 (4.3, 34.4)	3.19 (2.99, 3.68)
Gastric	524	11 (5.8, 31.5)	16.54 (13.78, 20.32)	3.2 (2.1, 12.9)	3.88 (3.22, 4.5)
Cholangiocarcinoma	421	19.4 (6.5, NR)	11.67 (9.96, 13.64)	4.2 (2.1, 24.9)	4.37 (3.65, 5.13)
Pancreatic	1,514	3.7 (2.1, 9.8)	10.09 (9.44, 10.95)	2.1 (1.9, 3.4)	4.18 (3.85, 4.64)
Small intestine	213	NR (16.2, NR)	21.76 (15.55, 31.53)	23.4 (4.3, NR)	3.98 (3.22, 5.42)
Ovarian	2,178	33.6 (11.0, NR)	28.47 (25.78, 30.77)	2.2 (2.0, 6.2)	2.76 (2.66, 2.89)
Brain	224	5.6 (1.5, 16.2)	10.29 (8.55, 12.43)	1.1 (0.7, 2.1)	4.7 (3.72, 6.18)

NR, not reached at the end of study follow-up period.
* Estimated from the fitted log-normal distribution to the digitized KM survival curve.
Abbreviations: RW, real-world; Pembro, pembrolizumab; Chemo, chemotherapies; PFS, progression-free survival; OS, overall survival.

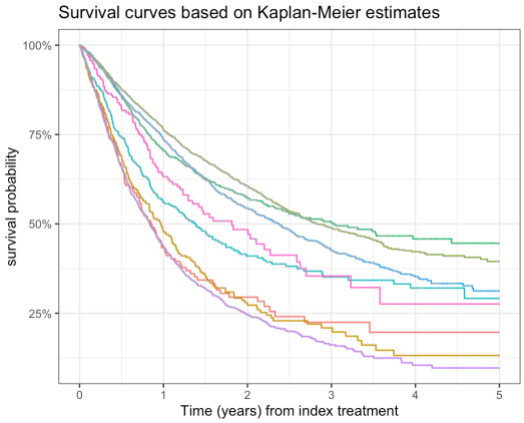
Table 2.4. Estimated survival rates and hazard ratios for colorectal and endometrial cancers

Outcomes	**Colorectal**				**Endometrial**			
	OS		**PFS**		**OS**		**PFS**	
	Trial	**RWC**	**Trial**	**RWC**	**Trial**	**RWC**	**Trial**	**RWC**
N	63	3,104	63	3,104	79	1,089	79	1,089
12-month survival, %	76	76.4	41	28	69	70.5	51	26.4
24-month survival, %	63	60.7	37	21.7	64	57.4	41	22.5
36-month survival, %	NA	48.9	NA	19.4	60	50.3	37	21.6
HR, unweighted – Cox (95%CI)	1.01 (0.68-1.51)	ref	0.68* (0.49-0.93)	ref	0.80 (0.55-1.17)	ref	0.54*** (0.40-0.71)	ref
HR, unweighted – AFT (95%CI)	1.05 (0.67-1.64)	ref	0.74 (0.51-1.06)	ref	0.89 (0.56-1.42)	ref	0.55*** (0.42-0.71)	ref
HR, the average weight – Cox (95%CI)	1.18 (0.78-1.80)	ref	0.46*** (0.33-0.64)	ref	0.85 (0.57-1.28)	ref	0.35*** (0.27-0.46)	ref
HR, the last weight – Cox (95%CI)	1.18 (0.78-1.80)	ref	0.49*** (0.36-0.68)	ref	0.84 (0.56-1.27)	ref	0.33*** (0.26-0.43)	ref

PFS, progression-free survival; OS, overall survival; HR, hazard ratio; Cox, Cox Proportional Hazard regression; AFT, accelerated failure time model; RWC, real-world cohort.
*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

Figure 2.1. Real-world chemotherapy survival curves by tumor types

a. Overall survival (OS)



b. Progression-free survival (PFS)

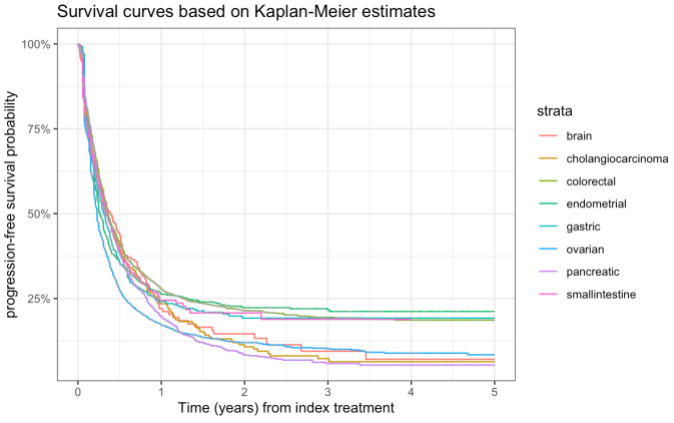
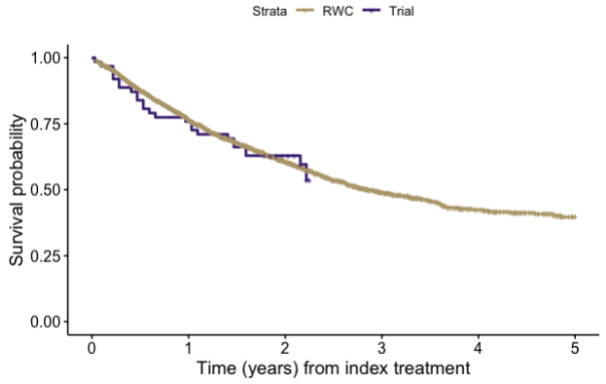
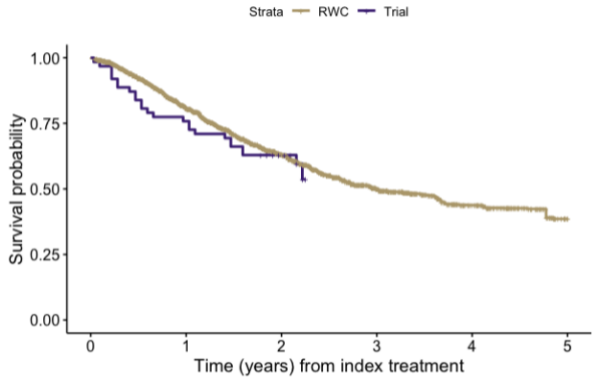


Figure 2.2. Adjusted and unadjusted Kaplan-Meier survival curves for colorectal cancer

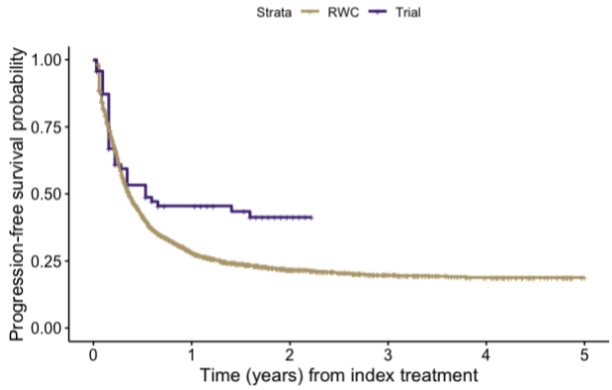
(1) Unadjusted OS curve



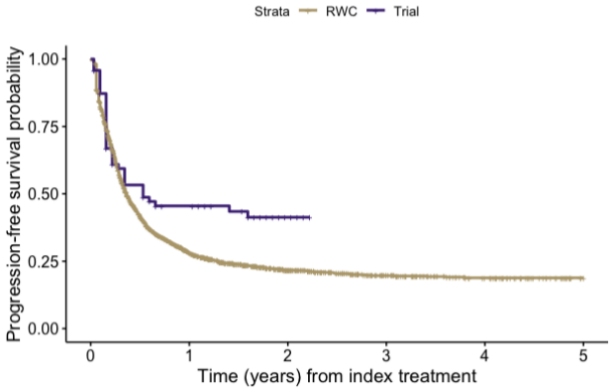
(2) Adjusted OS curve



(3) Unadjusted PFS curve



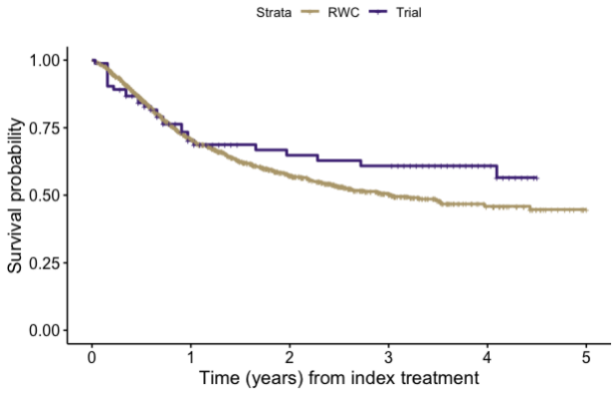
(4) Adjusted PFS curve



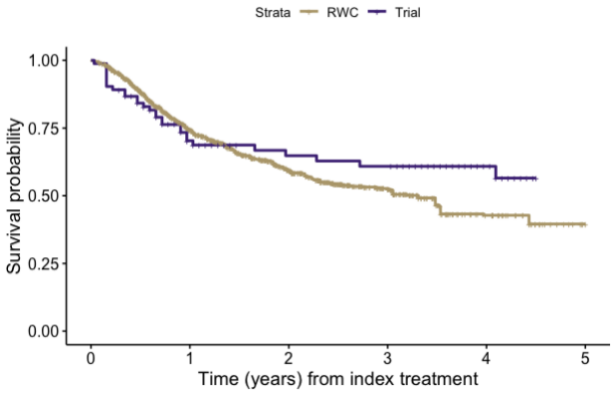
PFS, progression-free survival; OS, overall survival; RWC, real-world cohort.

Figure 2.3. Adjusted and unadjusted Kaplan-Meier survival curves for endometrial cancer

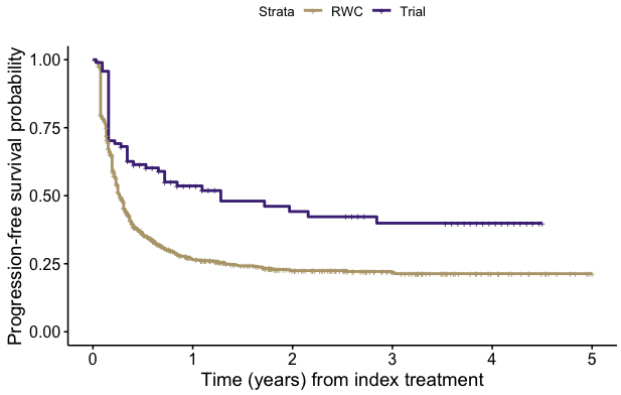
(1) Unadjusted OS curve



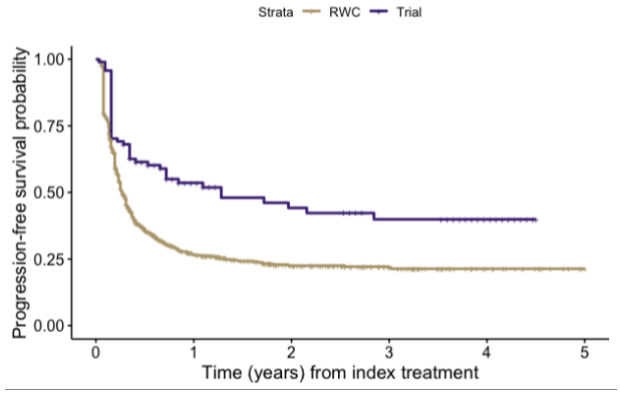
(2) Adjusted OS curve



(3) Unadjusted PFS curve



(4) Adjusted PFS curve



PFS, progression-free survival; OS, overall survival; RWC, real-world cohort.

2.9 SUPPLEMENT

Table S2.1. List of NCCN-guidelines based chemotherapy drugs for external controls

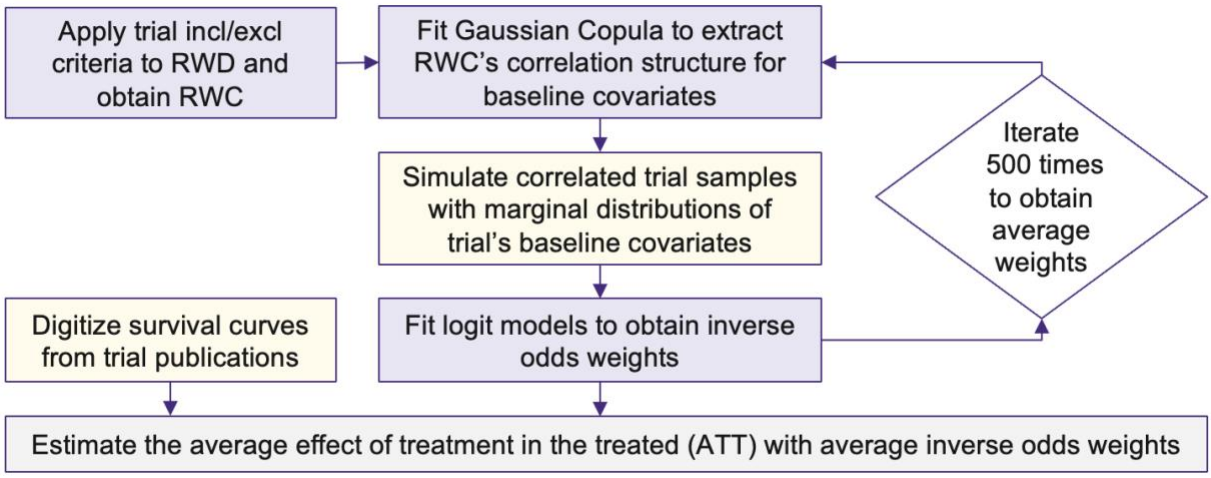
Tumor type	Generic name	RxNum
Brain	bevacizumab	253337
	carboplatin	40048
	carmustine	2105
	cisplatin	2555
	etoposide	4179
	lomustine	6466
	procarbazine	8702
	temozolomide	37776
	thioguanine	10485
	vincristine	11202
Cholangiocarcinoma	capecitabine	194000
	cisplatin	2555
	fluorouracil	4492
	gemcitabine	12574
	irinotecan	51499
	irinotecan hydrochloride	153329
	irinotecan liposome	1719768
	leucovorin	6313
	levoleucovorin	877015
	oxaliplatin	32592
	paclitaxel, albumin bound	486610
Colorectal	bevacizumab	253337
	capecitabine	194000
	fluorouracil	4492
	irinotecan	51499
	leucovorin	6313
	oxaliplatin	32592
	trifluridine and tipiracil	1670310
Endometrial	bevacizumab	253337
	carboplatin	40048
	cisplatin	2555
	docetaxel	72962
	doxorubicin	3639
	doxorubicin liposomal	214525
	gemcitabine	12574
	ifosfamide	5657
	ixabepilone	337523
	paclitaxel	56946
	paclitaxel, albumin bound	486610
	pemetrexed	68446
	temsirolimus	657797
	topotecan	57308
Gastric	capecitabine	194000
	carboplatin	40048
	cisplatin	2555
	docetaxel	72962
	fluorouracil	4492
	irinotecan	51499
	leucovorin	6313
	levoleucovorin	877015

	oxaliplatin	32592
	paclitaxel	56946
	trifluridine and tipiracil	1670310
Ovarian	bevacizumab	253337
	bleomycin sulfate	1621
	capecitabine	194000
	carboplatin	40048
	cisplatin	2555
	cyclophosphamide	3002
	dactinomycin	3100
	docetaxel	72962
	doxorubicin	3639
	doxorubicin liposomal	214525
	etoposide	4179
	fluorouracil	4492
	gemcitabine	12574
	ifosfamide	5657
	irinotecan hydrochloride	153329
	leucovorin	6313
	melphalan	6718
	melphalan hydrochloride	235857
	oxaliplatin	32592
	paclitaxel	56946
	paclitaxel, albumin bound	486610
	pemetrexed	68446
	topotecan	57308
	vinblastine	11198
	vincristine	11202
	vinorelbine	39541
Pancreatic	capecitabine	194000
	cisplatin	2555
	docetaxel	72962
	fluorouracil	4492
	gemcitabine	12574
	irinotecan	51499
	leucovorin	6313
	levoleucovorin	877015
	oxaliplatin	32592
	paclitaxel, albumin bound	486610
Small intestine	bevacizumab	253337
	capecitabine	194000
	carboplatin	40048
	docetaxel	72962
	fluorouracil	4492
	gemcitabine	12574
	irinotecan	51499
	leucovorin	6313
	oxaliplatin	32592
	paclitaxel	56946
	paclitaxel, albumin bound	486610

Table S2.2. Baseline characteristics of real-world chemotherapy cohorts

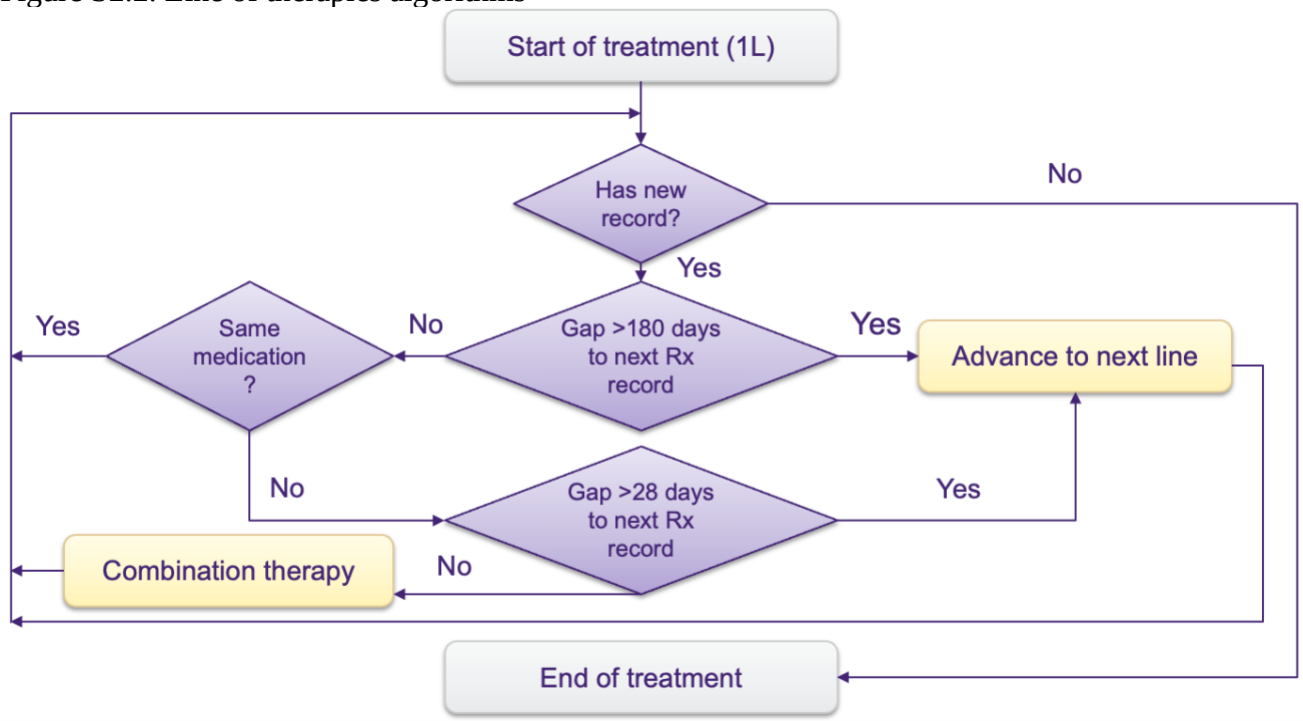
	Overall (N=9267)	Brain (N=224)	Cholangiocarcinoma (N=421)	Colorectal (N=3104)	Endometrial (N=1089)	Gastric (N=524)	Ovarian (N=2178)	Pancreatic (N=1514)	Small intestine (N=213)
Age at 1L									
Mean ± SD (Min, Max)	62.46 ± 11.82 (18, 90)	51.93 ± 16.08 (18, 86)	63.77 ± 10.48 (29, 86)	60.41 ± 11.99 (20, 90)	64.92 ± 10.35 (28, 88)	62.42 ± 11.40 (22, 87)	63.16 ± 12.02 (18, 90)	64.97 ± 10.20 (19, 90)	63.53 ± 12.26 (27, 86)
Age group									
<65	4959 (53.51 %)	171 (76.34 %)	209 (49.64 %)	1907 (61.44 %)	473 (43.43 %)	293 (55.92 %)	1103 (50.64 %)	702 (46.37 %)	101 (47.42 %)
>=65	4308 (46.49 %)	53 (23.66 %)	212 (50.36 %)	1197 (38.56 %)	616 (56.57 %)	231 (44.08 %)	1075 (49.36 %)	812 (53.63 %)	112 (52.58 %)
Sex									
Female	5912 (63.80 %)	95 (42.41 %)	215 (51.07 %)	1346 (43.36 %)	1089 (100.00 %)	173 (33.02 %)	2178 (100.00 %)	711 (46.96 %)	105 (49.30 %)
Male	3355 (36.20 %)	129 (57.59 %)	206 (48.93 %)	1758 (56.64 %)	0 (0.00 %)	351 (66.98 %)	0 (0.00 %)	803 (53.04 %)	108 (50.70 %)
Race group									
White	6965 (75.16 %)	183 (81.70 %)	322 (76.48 %)	2298 (74.03 %)	771 (70.80 %)	370 (70.61 %)	1669 (76.63 %)	1198 (79.13 %)	154 (72.30 %)
Asian	232 (2.50 %)	4 (1.79 %)	13 (3.09 %)	92 (2.96 %)	16 (1.47 %)	22 (4.20 %)	52 (2.39 %)	31 (2.05 %)	2 (0.94 %)
Black/African American	1251 (13.50 %)	16 (7.14 %)	41 (9.74 %)	453 (14.59 %)	210 (19.28 %)	86 (16.41 %)	238 (10.93 %)	175 (11.56 %)	32 (15.02 %)
Other	819 (8.84 %)	21 (9.38 %)	45 (10.69 %)	261 (8.41 %)	92 (8.45 %)	46 (8.78 %)	219 (10.06 %)	110 (7.27 %)	25 (11.74 %)
Ethnicity									
Hispanic or Latino	768 (8.29 %)	10 (4.46 %)	39 (9.26 %)	294 (9.47 %)	72 (6.61 %)	90 (17.18 %)	149 (6.84 %)	99 (6.54 %)	15 (7.04 %)
Not Hispanic or Latino	7055 (76.13 %)	196 (87.50 %)	325 (77.20 %)	2205 (71.04 %)	855 (78.51 %)	355 (67.75 %)	1759 (80.76 %)	1187 (78.40 %)	173 (81.22 %)
Unknown	1444 (15.58 %)	18 (8.04 %)	57 (13.54 %)	605 (19.49 %)	162 (14.88 %)	79 (15.08 %)	270 (12.40 %)	228 (15.06 %)	25 (11.74 %)
Prior lines of therapy									
1L	7593 (81.94 %)	200 (89.29 %)	376 (89.31 %)	2552 (82.22 %)	879 (80.72 %)	472 (90.08 %)	1556 (71.44 %)	1377 (90.95 %)	181 (84.98 %)
2L	1003 (10.82 %)	17 (7.59 %)	29 (6.89 %)	355 (11.44 %)	122 (11.20 %)	36 (6.87 %)	324 (14.88 %)	99 (6.54 %)	21 (9.86 %)
>=3L	671 (7.24 %)	7 (3.12 %)	16 (3.80 %)	197 (6.35 %)	88 (8.08 %)	16 (3.05 %)	298 (13.68 %)	38 (2.51 %)	11 (5.16 %)
Number of prior lines of therapy									
Mean ± SD (Min, Max)	1.31 ± 0.83 (1, 7)	1.15 ± 0.50 (1, 5)	1.16 ± 0.53 (1, 5)	1.29 ± 0.76 (1, 7)	1.34 ± 0.86 (1, 7)	1.17 ± 0.63 (1, 7)	1.55 ± 1.10 (1, 7)	1.13 ± 0.52 (1, 6)	1.24 ± 0.70 (1, 6)
Prior (neo)adjuvant ther	1803 (19.46 %)	59 (26.34 %)	25 (5.94 %)	653 (21.04 %)	334 (30.67 %)	51 (9.73 %)	521 (23.92 %)	136 (8.98 %)	24 (11.27 %)

Figure S2.1. Overview of study process



RWC, real-world cohort.

Figure S2.2. Line of therapies algorithms



Note: The role of drug interchangeability was incorporated. For example, 5-fluororuacil and capecitabine may be considered interchangeable and were categorized as ‘fluoropyrimidines’. Leucovorin and levoleucovorin were considered interchangeable, and the changing of a drug within these two would not suggest disease progression (hence, advancement of the line of therapy).

Figure S2.3. Cohort Attrition Chart

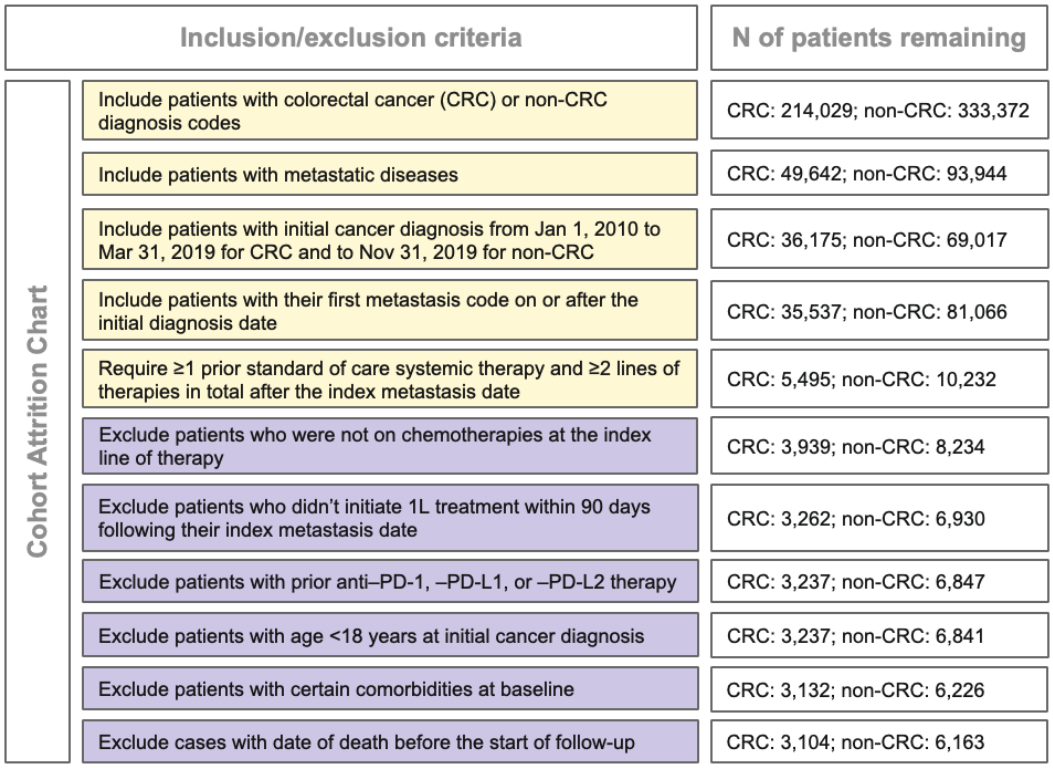
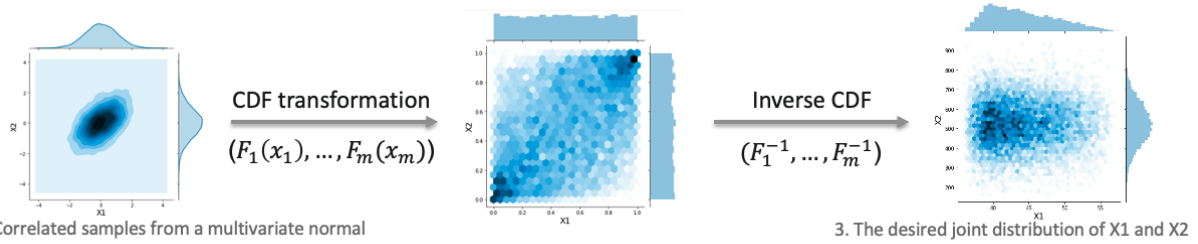


Figure S2.4. Illustration of a copula and applying continuous extension to sex variable

(1) CDF transformation and inverse CDF

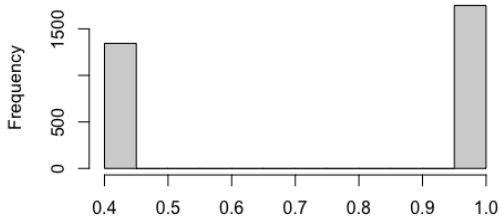


1. Correlated samples from a multivariate normal distribution with a covariance of 0.5.

2. The joint & marginal distributions of X1 and X2.

3. The desired joint distribution of X1 and X2 which are drawn from different distributions.

(2) CDF before transformation



(3) CDF after transformation

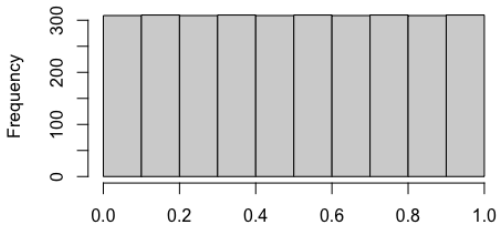
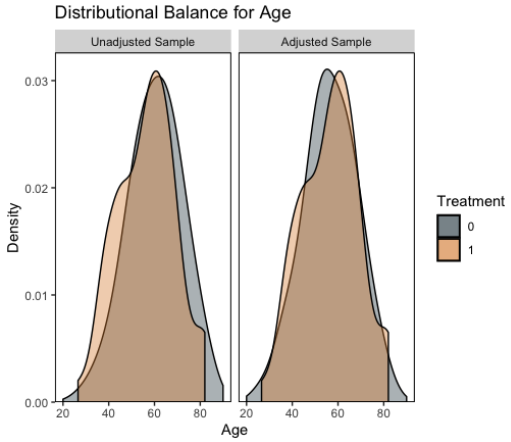
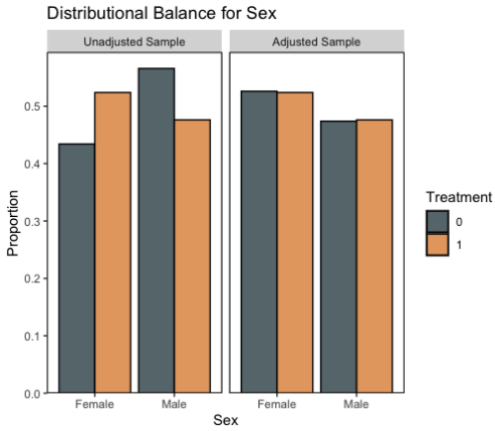


Figure S2.5. Covariate balance before and after propensity score weighting for colorectal cancer

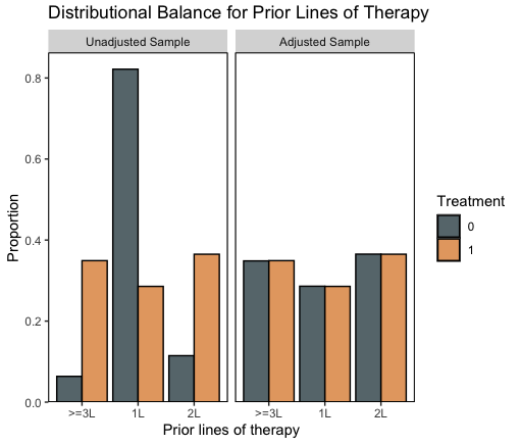
(1) Age



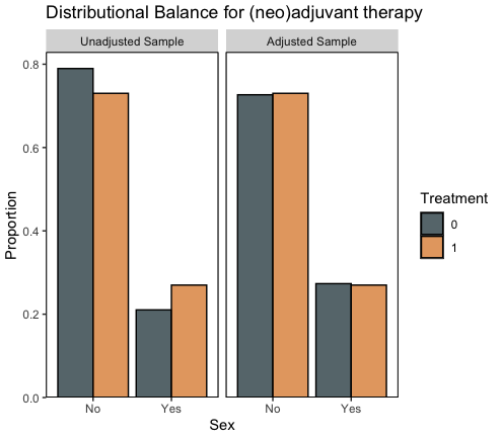
(2) Sex



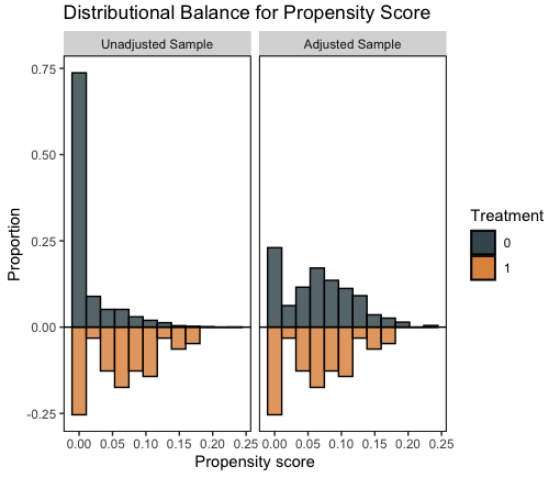
(3) Prior lines of therapy



(4) Prior (neo)adjuvant therapy



(5) Distribution of propensity score



(6) Covariate balance plot

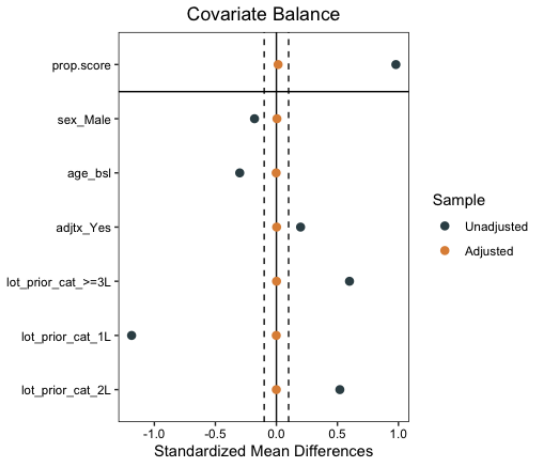
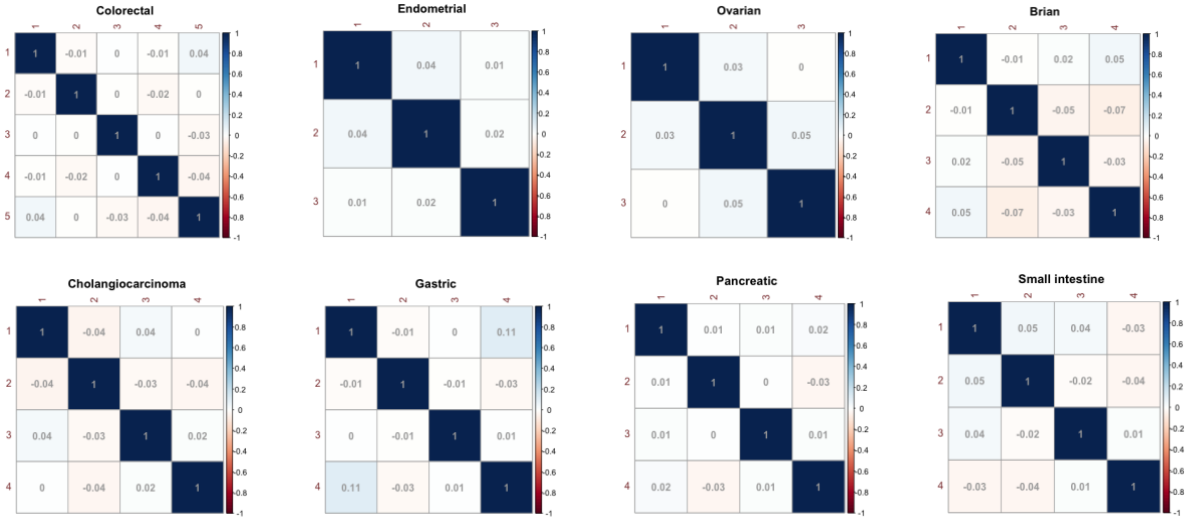


Figure S2.6. Correlation matrix of covariates by tumor types



Supplement File. Continuous Extension

(1) Rationale:

Sklar's Theorem states that an m-copula is an m-dimensional distribution function with all m univariate margins being $Uniform(0,1)$ [101]. In other words, Copula is a cumulative density function (CDF) defined over uniform marginals.

If y_m is a random variable with continuous CDF F, then the distribution of $U = F(y_m)$ is $Uniform(0,1)$.

Consider a continuous m-variate distribution function $F(y_1, \dots, y_m)$ with univariate marginal distributions $F_1(y_1), \dots, F_m(y_m)$ and inverse functions $F_1^{-1}, \dots, F_m^{-1}$. Then $y_1 = F_1^{-1}(u_1), \dots, y_m = F_m^{-1}(u_m)$, where $u_1, \dots, u_m$ are uniformly distributed variates. Hence,

$$F(y_1, \dots, y_m) = F(F_1^{-1}(u_1), \dots F_m^{-1}(u_m))$$
$$= \Pr[U_1 \le u_1, \dots, U_m \le u_m]$$
$$= C(u_1, \dots, u_m)$$
$$where\ U_j = Uniform(0,1), for\ j = 1, \dots, m$$

is **the unique copula** associated with the distribution function F.

Therefore, for an m-variate function F, the copula associated with F is a distribution function C that satisfies:

$$F(y_1, \dots, y_m) = C(F_1(y_1), \dots, F_m(y_m);\ \theta),$$

where θ is a parameter of the copula called **the dependence parameter**, which measures dependence between the marginals. If $F_1(y_1), \dots, F_m(y_m)$ are not all continuous, the joint distribution function can still be expressed as above, although in such a case, the copula is not unique [101]. **The lack of uniqueness of a copula for discrete distributions becomes an issue.**

In a discrete variable case, the margins are not uniform. Let $x = (x_1, \dots x_m)$ be a random vector with N^m values. The unique inverse transformation to generate the corresponding copula $F(F_1^{-1}(u_1), \dots F_m^{-1}(u_m))$ does not work anymore because its margins are not uniform. **So, to find a unique copula C, one solution is to first perform transformation on the discrete variables** [101].

In addition, implementation of the multivariate normal (MVN) copula for discrete data is possible, but not easy, because the MVN distribution as a latent model for discrete response requires rectangle probabilities based on high-dimensional integrations or their approximations [83]. Maximization of the likelihood with discrete margins often runs into computational difficulties, reflected in the convergence failure of the algorithm [101].

(2) Solution - Continuous extension:

Let $V_1, \dots, V_m$ be independent $Uniform(0,1)$ random variables, independent of $X = (X_1, \dots X_m)$. Consider

$$Y_j = X_j + V_j - 1,$$
$$Y = (Y_1, \dots Y_m)$$

Then, Y_j is continuous and

$$\{Y_j \leq x_j\} = \{X_j \leq x_j\},$$
$$x_j \in N.$$

The unique copula C of Y is also a copula of X.

Basically, what this means is that the discrete variables X_j are made continuous by adding a random independent perturbation term taking values in [0, 1] and then finding a unique copula based on Y_j [101]. The construction process is called continuous extension. This construction method was studied by Denuit and Lamber and they found that continuous extension preserves the concordance order [102]. This method was also used by Machado and Santos Silva to extend quantile regression to count data [103].

In this study, continuous extension transformation was applied to binary and categorical variables, including sex, race, prior (neo)adjuvant therapy, prior lines of therapy, and primary tumor site. After the continuous extension, one can base the estimation of Copula parameters on the likelihood function for a family of continuous copulas.

(3) Subsequent steps:
In general, there are three steps for constructing a Copula and generating samples from it.

First, marginal fitting. The first step is to specify and estimate univariate marginal distributions $(F_1(y_1), \dots, F_m(y_m))$ based on a statistical feature of the data. **Second, Copula fitting.** Methods include maximum likelihood estimation, semi-parametric approach, or the method of moments based on the invariance property of Kendall's tau under strictly increasing transformations of the marginals [104]. The Kendall's Tau method was used for estimating copula parameter. Kendall's Tau is a measure of dependence between random variables, which is particularly suitable when dealing with ordinal or non-parametric data. **Third, generate correlated random variables via the inverse of the joint probability function.** The inverse probability transformation $(F_1^{-1}, \dots, F_m^{-1})$ is applied to transform the Copula back to the natural distribution of the data [104].

Chapter 3.

Tackling Challenges in Evaluating Economic Value of Tumor-Agnostic Therapies: A Cost-Effectiveness Analysis of Pembrolizumab

3.1 ABSTRACT

OBJECTIVES: Assessing the value of tumor-agnostic therapies is challenging given the potential for treatment effect heterogeneity and variable standards of care (SoC) across tumor types, as well as small sample sizes and lack of comparative data. Our study developed and applied novel methods to assess the value of pembrolizumab compared to SoC considering its heterogeneous efficacy across eight tumor types to inform coverage and reimbursement decisions in the United States. We quantified the impact of making optimized coverage decisions based on a stratification approach by recommending coverage only for specific tumor subgroups compared to recommending pembrolizumab for all tumor types.

METHODS: We developed a partitioned survival model with three health states to evaluate the cost-utility of pembrolizumab for previously treated patients with advanced or metastatic MSI-H/dMMR cancers. Efficacy of pembrolizumab was based on data from basket trials directly or with adjustment using Bayesian hierarchical models (BHM). Tumor specific chemotherapy comparators were developed using the TriNetX databases. Outcomes included costs, life years (LYs), quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs) per QALY gained, and ICER per LY gained. Outcomes were stratified according to three decision criteria: tumor-aggregated recommendation (TAR) which covers pembrolizumab in all tumor types, tumor-specific recommendation based on undiscounted LYs gained or superior comparative effectiveness (TSR-CE), and tumor-specific recommendation based on incremental net health benefit (TSR-NHB).

RESULTS: Over a lifetime horizon, pembrolizumab was associated with increased incremental effectiveness compared to chemotherapies in colorectal (QALYs: 0.64, LYs: 0.64), endometrial (QALYs: 3.79, LYs: 5.47), and small intestine cancers (QALYs: 1.73, LYs: 2.48), but not for

patients with metastatic gastric, cholangiocarcinoma, pancreatic, ovarian, and brain cancers. Incremental cost-effectiveness ratio (ICER) values varied substantially across tumor types, with pembrolizumab being a cost-effective strategy in treating colorectal and endometrial cancer (ICER: $92,079 and $132,391, respectively) at $150,000 willingness-to-pay per QALY threshold, compared to SoC. Pembrolizumab was not found to be cost-effective in treating other assessed cancers. Recommending pembrolizumab for coverage with TSR-CE and TSR-NHB decisions increased the average incremental NHB by 13,998 ($2.10 billion NMB) and 15,983 ($2.40 billion NMB), respectively, compared to the TAR decision.

CONCLUSIONS: Value of tumor-agnostic therapies such as pembrolizumab can vary by cancer type. Using analytic tools such as external controls and BHMs can tackle some of the challenges in assessing the value of tumor-agnostic therapies and uncertainties arising from basket trials. Optimized recommendations for using pembrolizumab in specific patient populations with the greatest benefits can result in greater overall value in the healthcare system.

3.2 INTRODUCTION

In the era of precision medicine, tumor agnostic or histology-independent therapies have the potential to greatly benefit patients with limited therapeutic alternatives. Pembrolizumab was the first tumor-agnostic therapy approved by the U.S. Food and Drug Administration (FDA) in 2017 based on evidence from basket trials, i.e., clinical trials that simultaneously evaluate one drug across various disease subtypes that harbor the same molecular or genetic alteration [15]. Pembrolizumab was approved for the treatment of patients with unresectable or metastatic, microsatellite instability–high (MSI-H) or mismatch repair–deficient (dMMR) solid tumors that have progressed after prior treatment [7]. This new treatment paradigm focuses on specific genomic or molecular alterations of cancer cells rather than the tissue of origin. However, these types of therapies face significant challenges in adoption under the existing value assessment frameworks. Health technology assessment (HTA) agencies across health systems have detailed

major challenges of value assessment of tumor agnostic drugs (TADs), such as the potential for heterogeneity in treatment effect, the lack of comparative data due to single-arm studies, variable standards of care (SoC) across tumor types, and large uncertainty in the evidence base available to inform coverage and reimbursement decisions [6, 17, 18, 105, 106].

Traditionally, the SoC arm in a randomized controlled trial is used as the clinical and economic comparator for evaluating innovative therapies. However, TADs studied in single-arm basket trials do not allow for this approach because they lack a direct comparator arm. Furthermore, the variation in SoC options across different tumor types poses a challenge in identifying a universal comparator for all cases. In such situations, real-world data (RWD) has emerged as a valuable tool for assessing the efficacy of oncology drugs [62, 67, 107]. Creating external controls using RWD can be a viable approach to evaluate the clinical comparative and the cost-effectiveness of TADs assessed through single-arm, non-randomized basket trials, which will address the issue of differing SoC treatments and the resulting variable prognostic effects.

Another major concern of early phase basket trials is the heterogeneous treatment effect in various tumor types [17, 18]. Conventional stand-alone subgroup analyses often lack sufficient power to detect treatment effects because of the small sample size of each stratum [24]. For example, the Keynote 158 study enrolled 233 patients with 27 tumor types in a basket trial setting. The study reported a range of median progression-free survival (PFS) from 1.29 months in brain cancer (n=13) to 20.47 months in small intestine cancer (n=25) [19, 20]. Prior studies have proven utility of Bayesian hierarchical models (BHM) in improving the precision of estimates from such trials by borrowing information on treatment effects across clusters, especially in cohorts with small sample sizes [30, 43, 47].

Despite the marketing authorization of tumor-agnostic therapies for all tumor types, reimbursement may be appropriate only for specific histological subgroups due to the variability in the treatment effect and cost-effectiveness [108]. While a tumor-agnostic recommendation for all tumor types could expand patient access to new therapies, it also carries the risk of imposing significant financial burden on payers and wasting valuable resources that could be allocated more efficiently. A more effective approach would be to make optimized coverage decisions based on a stratification approach by recommending coverage only for specific tumor subgroups [109, 110]. These decisions could be based on superior comparative effectiveness or positive net health benefits (NHB) (considering both costs and benefits). This approach can improve population health and promote efficiency in the health system.

Several studies have attempted to address the challenges of evaluating cost-effectiveness of a TAD. However, most of these studies have used a naïve comparison approach for efficacy inputs based on targeted literature reviews or employed non-responders or an intracohort approach with prior line of therapy as controls [111, 112]. Additionally, some studies have applied generic model inputs across heterogenous cancers, such as utility values. In this study, we aim to estimate the cost-effectiveness of pembrolizumab compared with SoC among patients with 8 MSI-H/dMMR (i.e., MSI-H in this paper) metastatic cancers in the United States. This is the first cost-effectiveness model evaluating a tumor-agnostic indication from a U.S. payer perspective. We illustrate how some of the challenges associated with single-arm trial data may be overcome by leveraging BHM methods and external controls from real-world data for 8 different tumor types. Furthermore, we quantify the impact of stratified coverage decisions compared to a tumor-agnostic recommendation, using the incremental NHB framework.

3.3 METHODS

Model Overview

In this analysis, we estimated the health and cost outcomes of pembrolizumab for eight tumor types: colorectal, endometrial, gastric, cholangiocarcinoma, pancreatic, small intestine, ovarian, and brain cancers (Table S3.1). The target population was patients with advanced or metastatic MSI-H cancers who have failed at least one systemic anticancer therapy. For colorectal cancer, we considered patients who had recurrence after receiving treatment with fluoropyrimidine, oxaliplatin, and irinotecan.

We developed a partitioned survival model comprising three health states (i.e., progression free, post progression, and death). The model schema is presented in Figure S3.1. The intervention included treatment with pembrolizumab, and the comparators were tumor-specific SoC chemotherapies for eight tumor types. For the comparator arm, National Comprehensive Cancer Network (NCCN) guidelines were used to identify all available chemotherapy treatments for each tumor type. We excluded targeted/biologic drugs and experimental treatments outside current clinical guidelines.

The model estimated the lifetime costs, life years (LYs), quality-adjusted life years (QALYs) of each strategy and incremental cost-effectiveness ratios (ICERs) per QALY gained, and ICER per LY gained between the two strategies. All outcomes were estimated in a U.S. payer perspective. We used a willingness-to-pay (WTP) threshold of $150,000 per QALY gained. All costs and outcomes were discounted at an annual rate of 3% (Neumann, 2016).

Efficacy Inputs

<u>Intervention:</u>

We obtained efficacy inputs for pembrolizumab from two phase II MSI-H trials [19, 35], including Kaplan-Meier survival curves for colorectal cancer (CRC) and endometrial cancer, and tumor-specific median progression-free survival (PFS) and median overall survival (OS) endpoints for all included tumor types. Keynote 164 and Keynote 158 trials recruited a total of 63 patients (cohort B) with MSI-H CRC and 233 patients (cohort K) with MSI-H non-CRC, respectively [19, 20, 35, 39].

We used Engauge Digitizer software to extract the data points from available Kaplan-Meier survival curves in the trial publications. Commonly used parametric models were fitted to digitized data to extrapolate survival data beyond the trial observation period. The parametric distributions considered were exponential, Weibull, log normal, log logistic, gompertz, gamma, and generalized gamma. We selected the parametric models based on model fit statistics (i.e., Akaike information criterion and the Bayesian information criterion or Schwarz information criterion), visual inspection of the model fit to the Kaplan-Meier survival curve and whether they have similar hazard patterns, and clinical plausibility (Figure S3.2) [113].

For tumor types without Kaplan-Meier survival curves, we explored two sets of efficacy inputs: the trial-based estimates versus BHM-derived median PFS and OS estimates (Table S3.2). In the latter approach, we first fitted Bayesian hierarchical models to median PFS and OS estimates from the trials and obtained posterior mean estimates of median PFS and OS as efficacy inputs. Details of the modeling process can be found in Chapter 1. We then obtained the rates of the exponential models so that their medians are equal to the tumor-specific median PFS and OS estimates. Based on the chosen parametric models, we computed the probability of being alive (OS) or alive and progression-free at each cycle. The model used 60 years old as the

baseline age and we estimated the gender-weighted background mortality rate for each age group using the 2020 US life tables [114].

Comparator arms:

We used eight external controls derived from Chapter 2 as comparator arms. The Kaplan-Meier survival curves for the real-world PFS and OS are shown in Figure 2.1. We then fitted a range of parametric models to the survival data for all comparator arms to extrapolate the outcomes to a lifetime time horizon (Figure S3.3). The selection criteria for the distributional form were the same as those used for the intervention arms. Hazard functions of the parametric models can be found in Figure S3.4 & Figure S3.5.

Drug Acquisition and Monitoring Costs

Drug acquisition costs per unit were obtained from 2022 Veterans Affairs Federal Supply Schedule (VA FSS) prices, which reflect negotiated prices between drug companies and VA, one of the largest purchasers of pharmaceuticals in the U.S. Drug unit costs were applied to the utilization estimates to calculate monthly drug costs. Based on the trial protocol, patients receiving pembrolizumab were given 200 mg intravenously (IV) every 3 weeks and were assumed to continue for up to 35 cycles or until disease progression [19, 35]. Monthly costs of chemotherapy drugs were calculated using an average body weight of 70kg based on the median for cancer patients in the U.S. and were weighted by dosing regimens. If multiple dosing schedules were available, we chose the preferred regimen in the NCCN guidelines or the most used regimen. To calculate the total drug costs for each cycle, we multiplied the monthly cost of each chemotherapy drug by the proportion of patients receiving that drug and the proportion of

patients who remained on treatment during that cycle (Figure S3.6). We then summed these products to arrive at the total drug costs for that cycle:

$Total\ drug\ costs = \sum_{j=1}^{m} \sum_{i=1}^{n} monthly\ drug\ cost_i \times \%\ pts\ receiving\ drug_i \times \%\ pts\ on\ treatment_{ij}$, where i is the chemotherapy drug in the comparator line, j is the jth cycle, n is the total number of chemotherapy drugs, and m indicates the total number of cycles.

We retrieved the proportion of patients receiving each chemotherapy drug and the corresponding probability of being on that treatment over time for each tumor type from the TriNetx EHR database (Table S3.3).

Drug administration costs were calculated based on the 2022 Medicare physician fee schedule. The costs were calculated based on the Healthcare Common Procedure Coding System (HCPCS) codes on each IV infusion in the first hour, IV infusion for each additional hour, and IV injection were used for the cost calculations [115].

Pre-progression healthcare resource use (HRU) was estimated based on the recommended schedule of first and follow-up visits from the study protocols and NCCN guidelines (Table S3.4). HRU associated costs include one-time immunohistochemistry test, first and follow-up outpatient consultation, CT scan, and lab tests (complete blood count, hepatic function test, renal function tests) to monitor disease progression. Additionally, CA 19-9 test and CEA test were included for patients with pancreatic and colorectal cancers, respectively [116, 117]. These costs were determined using HCPCS codes and the Medicare physician fee schedule [115]. We obtained post-progression treatment costs for each tumor type from published literature and applied these costs to each cycle during the post-progression state (Table 3.1) [118].

Health State Utilities

We used utility values for each tumor type obtained from previously published health preferences studies and applied these values to pre-progression and post-progression health states, respectively. A targeted literature review was conducted in Tufts CEA Registry to assess health state utility values (HSUV) used in published CEA studies related to the treatment and screening of metastatic cancers. We assumed that HSUVs did not vary across treatments in the model. Additionally, we accounted for the negative impact of experiencing any grade 3 or 4 adverse event (AE) by applying a disutility value. The disutility values were obtained from a systematic review of utility values for chemotherapy-related AEs (Table S3.1) [119].

Adverse Events and Associated Costs

The model considered grade 3/4 treatment-related AEs that occurred in at least 5% of patients, as reported in the Keynote 158 trials, as these AEs were considered more severe and likely to impact patient health or costs. We used AE rates in the chemotherapy arm of the Keynote 177 trial to the comparator arms in our model due to the lack of comparators in the single-arm basket trials. To estimate the cost of treating AEs, we obtained the percentage of patients experiencing each AE, the duration of occurrence, and the cost of hospitalization for specific AEs with diagnosis codes from the Healthcare Cost and Utilization Project (Table 3.1) [120]. We applied the costs of treating AEs as a one-time cost in the first cycle for each arm. The model assumed the median time to resolution of grade 3/4 treatment-related AEs to be one month [35].

Value of Stratification (VoS)

Incremental net health benefit (INHB) refers to the difference between the benefits gained with the intervention and the health forgone elsewhere in the health-care system (i.e., ΔQALYs -ΔCosts/WTP), expressed in QALY terms. The incremental net monetary benefit (INMB) was calculated by multiplying the difference in effectiveness of the two interventions by a WTP threshold and then subtracting the difference in their costs (i.e., ΔQALYs*WTP - ΔCosts). The value of stratification measures the difference between the INHB of a fully stratified recommendation, i.e., recommendations for each tumor type, and a tumor-aggregated recommendation with no stratification [109, 110]. We summarized the INHB, INMB, incremental total QALYs, and incremental total costs of three policy scenarios: (1) a tumor-aggregated recommendation (TAR policy) on pembrolizumab (i.e., coverage for all tumor types regardless of their incremental benefits); (2) a tumor-specific recommendation based on undiscounted LYs gained or superior long-term comparative effectiveness (TSR-CE policy) comparing pembrolizumab to SoC; and (3) a tumor-specific recommendation based on positive INHB (TSR-NHB policy) (i.e., only cover a subset of tumor types with positive INHB comparing pembrolizumab to SoC).

To calculate the annual incidence of included metastatic cancers with at least one prior systemic treatment, we sourced epidemiology inputs from various data sources, including SEER statistics, TriNetx EHR database, and prior publications (see Table S3.5) [121-123].

Sensitivity Analyses

Deterministic sensitivity analyses (DSA) were performed across a wide range of values to test the robustness of the model and to identify the impact of parameter uncertainty and key drivers of model outcomes. Cost inputs were varied within ± 25% of baseline values and

alternative HSUVs from the targeted literature review were tested in the DSA. Different parametric estimation methods were tested to assess the uncertainty in extrapolation of efficacy inputs. We used a tornado chart to display the results of the one-way sensitivity analyses.

Furthermore, a probabilistic sensitivity analysis (PSA) was performed in which model parameters of all intervention and comparator groups were varied simultaneously according to assigned distributions, using Monte Carlo simulation for 5,000 iterations. The ranges and distributions used in the sensitivity analyses were summarized in Table 3.1. We plotted cost-effectiveness acceptability curves (CEACs) to show the uncertainty surrounding the cost-effectiveness of pembrolizumab. The CEACs present the probability that an intervention is cost-effective over a given range of WTP per QALY values.

All costs were expressed in 2022 US dollars. The cost-effectiveness model was programmed in MS Excel (Redmond, WA). All other analyses were conducted in R 4.1.2 with survival and flexsurv packages (The Core R Team).

Due to the nature of the study, which involved using de-identified secondary data and no direct contact or interaction with human participants, the ethical approval was not obtained.

3.4 RESULTS

Stratified Base Case Analyses

The base case analyses reflect the cost-effectiveness of pembrolizumab in treating cancer patients with 8 types of locally advanced or metastatic MSI-H solid tumors who failed prior systemic therapies, compared with SoC chemotherapies. At a lifetime horizon, pembrolizumab was associated with increased effectiveness compared to chemotherapies in colorectal (QALYs: +0.64, LYs: +0.64), endometrial (QALYs: +3.79, LYs: +5.47), and small intestine cancers (QALYs: +1.73, LYs: +2.48). However, pembrolizumab was not associated with better long-

term effectiveness among patients with other tumor types (gastric, cholangiocarcinoma, pancreatic, ovarian, and brain) (Table 2). The total costs and a breakdown by cost categories by tumor type are shown in Table S3.6. Compared to chemotherapies, pembrolizumab was shown to be a cost-effective strategy in treating metastatic colorectal and endometrial cancers (ICERs/QALY: $92,079 vs. $132,391, respectively). Pembrolizumab was considered less costly and less effective in treating metastatic brain cancer (QALYs: -0.83, Costs: -$170,631) and not cost-effective in treating other cancers (gastric, cholangiocarcinoma, pancreatic, small intestine, and ovarian). Combined with variable incremental total costs (range: -$170,631 to $480,563) and incremental total QALYs (range: -2.03 to 3.79), a range of ICER results were derived across tumor types (Table 3.2).

Tumor-Aggregated versus Tumor-Specific Recommendations

Table 3 summarized key outcomes from recommendations based on three policy scenarios: TAR, TSR-CE, and TSR-NHB. For models with trial-based effectiveness inputs for pembrolizumab (Policy 1a/2a/3a), TSR-CE policy preferred pembrolizumab among patients with colorectal and endometrial cancers. For models with BHM-derived effectiveness inputs for pembrolizumab (Policy 1b/2b/3b), in addition to colorectal and endometrial cancers, TSR-CE policy covered patients with small intestine cancer. In both scenarios, TSR-NHB policy preferred pembrolizumab among patients with colorectal, endometrial, and brain cancers. The INHB and INMB calculations were using a WTP threshold of $150,000 per QALY. Mean and 95% credible ranges (CR) from the 5,000 PSA simulations were summarized to characterize the uncertainty around estimates.

A substantial increase in the INHB and INMB was associated with changing decision criteria from tumor-aggregated recommendation to tumor-specific recommendations. This was represented by an average value of stratification of 13,998 NHB per QALY (95%CR: 3,759, 24,058) or $2.10 billion (b) NMB (95%CR: $0.56b, $3.61b) for a heterogeneity-adjusted model comparing TSR-CE decision to TAR decision (Policy 2b vs. 1b). A higher VoS was associated with comparing TSR-NHB to TAR (Policy 3b vs. 1b), at the mean of 15,983 NHB per QALY (95%CR: 5,673, 26,056) or $2.40 billion NMB (95%CR: $0.85b, $3.91b). Using TSR-CE based on superior comparative effectiveness provided the greatest incremental total QALYs among all decision scenarios.

Aggregated ICERs and value-based prices were obtained for pembrolizumab (Table 3.4). The mean ICER for TSR-CE and TSR-NHB decisions are $117,138 and $89,833 per QALY, respectively. However, pembrolizumab is less effective and more costly compared to chemotherapies under the TAR decision. The value-based prices for pembrolizumab at three commonly used WTP thresholds are shown in Table 3.4.

Incorporating Heterogeneity into Basket Trial Efficacy Inputs

On average, adjusting trial-based median effectiveness inputs using BHM-derived median PFS and median OS was associated with an increase in the incremental total QALYs for both TAR (Policy 1b: -0.06, 95%CI: -1.56, 1.42; vs. Policy 1a: -0.12, 95%CI: -1.60, 1.39) and TSR-NHB policy scenarios (Policy 3b: 0.83, 95%CI: -2.39, 4.03; vs. Policy 3a: 0.80, 95%CI: -2.43, 4.05). However, the small increments in effectiveness were offset by the increase in total costs, resulting in lower INHB and INMB estimates comparing heterogeneity-adjusted models to the original models. In addition, in the TSR-CE policy scenario, adjusting heterogeneity in

efficacy inputs has resulted in expanded coverage of pembrolizumab to patients with small intestine cancers, due to superior undiscounted life years compared to chemotherapies (Undiscounted LYs: Policy 2b: 7.34 vs. 4.45; Policy 2a: 4.33 vs. 4.45).

Sensitivity Analyses

The CEAC was generated to show the probability that pembrolizumab is cost-effective compared to SoC chemotherapies at different WTP thresholds per QALY gained under different policy scenarios (Figure 2). Using tumor-aggregated recommendation, pembrolizumab was less likely to be cost-effective than SoC chemotherapies under any WTP thresholds ($0-$300k per QALY). On the contrary, pembrolizumab was more likely to be cost-effective than chemotherapies over $211,100 and $203,000 per QALY thresholds using TSR-CE and TSR-NHB, respectively. Overall, this suggests there is high uncertainty for pembrolizumab being cost-effective versus chemotherapies in tumor-agnostic decisions, especially when it is recommended for all tumor types. The CEAC stratified by individual tumor types are summarized in Figure S3.7. The figure shows that pembrolizumab is 65% cost-effective than chemotherapies at a WTP threshold of $150,000 per QALY for colorectal and endometrial cancers.

The one-way DSA results for colorectal and endometrial cancers are shown in the tornado chart (Figure S3.8). The most influential parameters on ICER per QALY were consistent for both cancers: PFS estimation of chemotherapy using the generalized gamma distribution instead of gompertz, post-progression treatment costs, drug acquisition cost for pembrolizumab, and utility values. All other parameters had negligible influences on the base-case ICER.

3.5 DISCUSSION

Our study addressed multiple challenges of estimating the cost-effectiveness of pembrolizumab in MSI-H metastatic/advanced cancers and is the first decision analytic model to do so for a tumor-agnostic therapy from the U.S. payer perspective. Our main findings suggest that recommendations for using pembrolizumab in specific patient populations, based on comparative effectiveness or NHB/NMB, result in greater overall value to the healthcare system compared to a tumor-aggregated recommendation. In other words, a policy that fully approves pembrolizumab across tumors could result in an average annual loss of $2.10 billion (13,998 QALYs) to $2.40 billion (15,983 QALYs) to the U.S., compared to an optimized approval decision based on stratification approaches, assuming the same level of uptake across tumor types. We also found substantial variation in the ICER values across tumor types, with pembrolizumab being a cost-effective strategy in treating colorectal cancer and endometrial cancer at $150,000 WTP per QALY threshold, compared to SoC chemotherapies. However, pembrolizumab was not found to be cost-effective in treating other assessed cancers. Variation in patients' prognosis, different SoC treatments, health-related quality of life, and healthcare costs across tumor types contribute to substantial variation in relevant outcomes that influence the economic value of the drug.

Implications on Coverage and Reimbursement Decision Making

To date, five TAD's have been approved by the U.S. FDA and many more are in the drug development pipeline. Early approval of these innovative therapies can bring substantial health benefits to patients with later-stage or rare diseases. However, reimbursement authorities are then faced with a myriad of challenges in their assessments. A key challenge is the potential

heterogeneity in the comparative effectiveness and value across tumor types due to treatment effect heterogeneity and variable SoC options. The average cost-effectiveness across patient populations may not convey sufficient information to decision makers about the drug's value for money in different tumor types. Therefore, separate ICERs should be estimated for individual tumor types where possible. In this context, even when TAD's receive marketing authorization for all tumor types, reimbursement authorities should consider restrictions to the subset of tumor types with the positive comparative or net health benefit. In our analysis, stratified policy decisions based on comparative effectiveness would warrant reimbursement for pembrolizumab only in treating colorectal, endometrial, and small intestine cancers. Alternatively, healthcare systems valuing both effectiveness and costs could cover pembrolizumab for treating colorectal, endometrial, and brain cancers. Reimbursement agencies assessing tumor-agnostic drugs can optimize recommendations by limiting reimbursement to subpopulations with greatest benefits, increasing allocative efficiency [109, 110].

HTA bodies in different countries have applied various approaches to evaluating tumor-agnostic therapies. For example, the French agency (Haute Autorité de Santé [French National Authority for Health]) requests prospectively planned real-world comparator data if randomized trials are not feasible [124]. The English HTA agency (National Institute for Health and Care Excellence [NICE]) offers conditional coverage with re-assessment of additional evidence later. In contrast, the European Medicines Agency (EMA) considers a high response rate with a long duration of response in a basket trial sufficient to support efficacy and approval of tumor-agnostic drugs. However, reimbursement policies for these therapies may vary depending on the specific insurance provider. In the absence of compelling evidence of clinical effectiveness or value, healthcare systems may limit patient access or employ other mechanisms to deal with the

uncertainty. For example, decision-makers may consider using coverage with evidence generation approaches like those employed by CMS or the Cancer Drugs Fund in the UK. However, not all healthcare systems have established processes and infrastructure to allow for conditional reimbursement or post-authorization evidence generation [125]. Another option may be the use of outcomes-based contracts or formularies (A.K.A. value-based, managed access or patient access schemes) [126-128]. Value of information studies may be useful for quantifying the expected value of research in reducing decision uncertainty and determining whether existing evidence is sufficient for making decisions.

Adjusting Heterogeneous Treatment Effects using BHM

Tumor-agnostic therapies have thus far been approved through accelerated approval pathways by the FDA, based on multi-cohort, single-arm basket trials. These trials pose unique challenges for HTA agencies, given the considerable heterogeneity in treatment effects and uncertainty due to small sample sizes in certain tumor types. To incorporate these estimates into comparative effectives as well as cost-effectiveness models, they must be adjusted in a systematic manner. BHM addressed these challenges by borrowing information across subgroups in a basket trial to infer subgroup-specific treatment efficacy. In Chapter 1, we have demonstrated that the BHM approach shrunk the original tumor-specific estimates towards a pooled average treatment effect and reduced uncertainty around estimates overall. In this Chapter, we observed a slightly narrower 95% credible range of the long-term comparative effectiveness, measured by incremental total QALYs with the BHM-derived median PFS and OS estimates, compared to the original models (Policy 1b/3b vs. Policy 1a/3a). Using BHM to inform model inputs in economic evaluation is relatively new. Murphy et al's used BHM to

estimate the posterior probability of response for different histologies in a basket trial, to inform a response-based cost-effectiveness analysis of tumor-agnostic drug, Larotrectinib [48]. The authors used hypothetical data in their study, which limits the ability to draw meaningful real-world conclusions. NICE has not seen any firm evidence to support the adequacy of the response-based model in modeling long-term outcomes [18].

RWD-based External Controls as Comparators

Our study suggests using external controls from RWD as comparators is a viable approach. The results showed considerable variation in long-term effectiveness (in QALYs and LYs) of both pembrolizumab and SoC comparators, resulting in a wide range of incremental benefits. For instance, treatment with pembrolizumab compared to chemotherapy showed additional long-term benefits for patients with metastatic endometrial, colorectal, and small intestine cancers, but not for patients with metastatic gastric, cholangiocarcinoma, pancreatic, ovarian, and brain cancers. We applied eligibility criteria consistently to all cancers; however, the SoC chemotherapies mix and therapeutic effectiveness differ in nature in various cancers, resulting in heterogeneous comparative effectiveness. Two other studies have explored the use of external controls for assessing cost-effectiveness of TAD. Huygens et al. has used data from the Hartwig Medical Foundation, which cover 44 hospitals in The Netherlands, to estimate relative effectiveness of entrectinib compared to SoC [108]. Briggs et al. described 3 options about the estimation of the counterfactual for tumor-agnostic treatments with only single-arm trial data available: (1) historical controls reported in the literature, (2) previous line of therapy in intervention patients of basket trials, or (3) nonresponders in basket trials [112]. To reduce clinical uncertainty when using external controls from RWD, payers and HTA agencies should

plan for data collection on the clinical effectiveness of comparators through post-marketing studies or data collection in clinical practice as early as possible. This will enable more informed recommendations to be made. However, some payers may prefer to manage decision uncertainty by negotiating price discounts or outcomes-based agreements rather than using more complex arrangements that require data collection [125].

Our study fitted independent parametric models to obtain long-term effectiveness for treatment and control groups (Table S3.2). Alternatively, hazard ratios can be applied to the comparators to obtain long-term outcomes of the treatment group. For example, the average treatment effect (ATE) on overall survival of using pembrolizumab versus chemotherapy for all cancer X patients can be estimated if both individual-level covariates and outcome data are available. The ATE = $E[Y_i(1) - Y_i(0)]$ is useful to evaluate what is the expected effect on the outcome in the population. Weights can be defined as $w_i = \frac{Z_i}{e_i} + \frac{(1-Z_i)}{1-e_i}$, where Z_i is an indicator variable for treatment assignment and e_i denotes the propensity score for the ith subject. In this case, we didn't apply the hazard ratio approach as only average treatment effect on the treated (ATT) rather than ATE could be estimated in colorectal and endometrial cancers (see Chapter 2). Due to the lack of linked individual-level covariates and survival data in the treated group, weights (if not equal to 1) cannot be assigned to each individual to estimate ATE. The ATT is defined as $E[Y_i(1) - Y_i(0)|Z = 1]$ and the weighting method ($w_{i,ATT} = Z_i + \frac{(1-Z_i)e_i}{1-e_i}$) was to adjust the control group to resemble the treated group and leaving the outcomes of the treated group untouched (i.e., given weight of 1). In addition, one could estimate ATE by taking the weighted average of the ATT and the average treatment effect on the untreated (ATU), which represents the treatment effect for those who did not receive the treatment. The ATU is defined as $E[Y_i(1)-$

$Y_i(0)|Z = 0]$. The use of weights equal to $w_{i,ATU} = \frac{(1-Z_i)e_i}{1-e_i} + (1 - Z_i)$ allows one to adjust the treated group to resemble the control group [85, 129]. Similar to ATE, ATU could be estimated with linked patient-level covariates and survival outcome in the treated group.

3.6 LIMITATIONS

Our study has several limitations. Only Kaplan-Meier survival curves of colorectal and endometrial cancers are publicly available for pembrolizumab, which limits our ability to fit alternative parametric models to the Kaplan-Meier survival curves of other cancers. Sensitivity analysis can be conducted by fitting different parametric models when more survival data (e.g., KM survival curves or time-to-event data at different follow-up times, etc.) becomes available. In addition, when there are more events and longer follow-up time, more flexible survival models such as cubic spline models and piecewise models (e.g., KM curve with an extrapolated tail) could more accurately predict long-term survival to estimate cost-effectiveness [130]. For example, we fitted piecewise model for estimating PFS in colorectal cancer in the sensitivity analysis. Using a VAFSS drug cost also limits the generalizability for commercial payers or plans that do not receive the same level of rebates on drug costs as the VA. It is similar to the lowest net price offered to commercial buyers and average net price in Medicare Part D [131]. We did not adjust the prognostic effect of MSI-H status in the comparator population due to lack of MSI-H status information in the RWD. However, among patients with metastatic or advanced cancers who received chemotherapies, our targeted literature review found similar or non-significantly different prognosis of MSI-H tumors to that of MSI-low/microsatellite stable tumors [132-136]. Future studies should explore the generalizability of our approach to other

tumor-agnostic indications. For example, the feasibility of finding the right data sources to generate external controls when the molecular biomarker has a strong prognostic effect among the external control population. Finally, this study assumes that there is the same level of uptake of each policy across tumors. Future studies should examine how the results would change after accounting for the potential varying levels of uptake of a new technology post approval. Clinicians treating patients with different cancers may react differently to a novel treatment option, particularly considering the presence of heterogeneous treatment effects, varying availability of alternative treatment options, and potential differences in safety profiles across tumor types.

3.7 CONCLUSION

This paper provides a detailed analysis of the cost-effectiveness of pembrolizumab versus SoC chemotherapies, demonstrating that pembrolizumab is highly cost-effective for colorectal and endometrial cancers. By limiting reimbursement to those with the greatest benefits, optimized recommendations for using pembrolizumab in specific patient populations result in greater overall value to the healthcare system and improve allocation efficiency, compared to a tumor-aggregated recommendation. It is essential for regulatory agencies, manufacturers, payers, and HTA bodies to consider the challenges of assessing tumor-agnostic therapies associated with the inevitable uncertainty arising from basket trials.

3.8 TABLES & FIGURES

Table 3.1. Model parameters: base-case values and sources

Parameter	Base-Case Estimate	Source	PSA distribution
Monthly drug acquisition cost for pembrolizumab	$14,336.39	[137]	Gamma
Drug administration costs, $			
Cost of intravenous infusion (IV) infusion up to one hour	$140.16	[115]	Gamma
Cost of IV infusion each additional hour	$29.76	[115]	Gamma
Cost of IV injection	$102.00	[115]	Gamma
Healthcare resource use per month, $			
Pre-progression monitoring cost for pancreatic cancer during 1st year	$547.70	[115]	Gamma
Pre-progression monitoring cost for pancreatic cancer after 1st year	$485.88	[115]	Gamma
Pre-progression monitoring cost for colorectal cancer during 1st year	$805.92	[115]	Gamma
Pre-progression monitoring cost for colorectal cancer after 1st year	$744.10	[115]	Gamma
Pre-progression monitoring cost for other cancers during 1st year	$377.51	[115]	Gamma
Pre-progression monitoring cost for other cancers after 1st year	$315.69	[115]	Gamma
Post-progression treatment cost for brain cancer	$21,959.53	[118]	Gamma
Post-progression treatment cost for cholangiocarcinoma	$11,203.94	[118]	Gamma
Post-progression treatment cost for colorectal cancer	$12,909.85	[118]	Gamma
Post-progression treatment cost for endometrial cancer	$9,908.41	[118]	Gamma
Post-progression treatment cost for gastric cancer	$15,731.49	[118]	Gamma
Post-progression treatment cost for ovarian cancer	$14,489.92	[118] Assumed the same as uterus cancer	Gamma
Post-progression treatment cost for pancreatic cancer	$15,948.47	[118]	Gamma
Post-progression treatment cost for small intestine cancer	$12,208.68	[118] Assumed the same as colorectal cancer	Gamma
Health State Utility Values			
Progression-free, colorectal cancer	0.81	[138]	Beta
Progressive disease, colorectal cancer	0.65	[139]	Beta
Progression-free, endometrial cancer	0.75	KEYNOTE-158 [39]	Beta

Progressive disease, endometrial cancer	0.63	KEYNOTE-158 [39]	Beta
Progression-free, gastric cancer	0.80	[140]	Beta
Progressive disease, gastric cancer	0.58	[140]	Beta
Progression-free, cholangiocarcinoma	0.77	KEYNOTE-158 [19]	Beta
Progressive disease, cholangiocarcinoma	0.67	KEYNOTE-158 [19]	Beta
Progression-free, pancreatic cancer	0.80	[141]	Beta
Progressive disease, pancreatic cancer	0.78	[141]	Beta
Progression-free, small intestine cancer	0.81	Assumed the same as colorectal cancer	Beta
Progressive disease, small intestine cancer	0.65	Assumed the same as colorectal cancer	Beta
Progression-free, ovarian cancer	0.72	[142]	Beta
Progressive disease, ovarian cancer	0.65	[142]	Beta
Progression-free, brain cancer	0.69	[143]	Beta
Progressive disease, brain cancer	0.44	[143]	Beta
Adverse event (AE)-related disutility			
Diarrhea	-0.060	[119]	Normal
Fatigue	-0.115	[119]	Normal
Abdominal pain	-0.070	[119]	Normal
Anemia	-0.090	[119]	Normal
Hypertension	-0.010	[119]	Normal
Hypokalemia	-0.070	[119]	Normal
Neutropenia	-0.897	[119]	Normal
AE hospitalization costs, $			
Diarrhea	$10,006.13	[120]	Gamma
Fatigue	$12,605.46	[120]	Gamma
Abdominal pain	$9,133.60	[120]	Gamma
Anemia	$9,708.69	[120]	Gamma
Hypertension	$8,703.33	[120]	Gamma
Hypokalemia	$9,907.67	[120]	Gamma
Neutropenia	$16,886.45	[120]	Gamma
Other Parameters			
Baseline age, years	60	TriNetx Databases	
Mean body weight, kg	70	[144]	
Mean body surface area (m^2)	1.86	[144]	
Annual discount rate, %	3	Assumption	
Time horizon, years	40	Assumption	
Cycle length	1 month	Assumption	

Table 3.2. Stratified base case results

Tumor type	Treatment	Total Costs	Total QALYs	Total LYs	ΔCosts	ΔQALYs	ΔLYs	Incremental Cost per QALY Gained	Incremental Cost per LY Gained
Colorectal	Pembro	$401,565	4.72	6.06	$58,860	0.64	0.64	$92,079	$92,218
	Chemo	$342,705	4.08	5.42					
Endometrial	Pembro	$742,400	7.74	11.07	$501,792	3.79	5.47	$132,391	$91,721
	Chemo	$240,608	3.95	5.60					
Gastric	Pembro	$240,909	0.83	1.29	$31,209	-2.03	-2.58	Less effective, more costly	Less effective, more costly
	Chemo	$209,700	2.86	3.88					
Cholangiocarcinoma	Pembro	$311,170	1.49	2.15	$105,554	-0.06	-0.06	Less effective, more costly	Less effective, more costly
	Chemo	$205,616	1.55	2.21					
Pancreatic	Pembro	$91,987	0.39	0.49	-$138,427	-1.12	-1.43	$123,946	$96,518
	Chemo	$230,415	1.50	1.92					
Small Intestine	Pembro	$792,396	4.29	6.04	$480,563	1.73	2.48	$277,417	$193,922
	Chemo	$311,833	2.56	3.56					
Ovarian	Pembro	$602,577	2.26	3.45	$89,943	-0.18	-0.23	Less effective, more costly	Less effective, more costly
	Chemo	$512,634	2.43	3.68					
Brain	Pembro	$182,071	0.36	0.74	-$170,631	-0.83	-1.46	$204,766	$116,595
	Chemo	$352,703	1.20	2.20					

QALY, quality-adjusted life years; LY, life years; Pembro, pembrolizumab; Chemo, chemotherapy.

Table 3.3. Value of stratification: outcome estimates based on tumor-aggregated versus tumor-specific recommendations

Policy Scenario	**Incremental Total Costs**	**Incremental Total QALYs**	**INHB at $150k/QALY**	**INMB at $150k/QALY**
	Mean (95%CR)	Mean (95%CR)	Mean (95%CR)	Mean (95%CR)
Trial-based effectiveness inputs				
Policy 1a (TAR): recommending pembro for all tumor types vs. chemo for all	$61,193	-0.12	-19,881	-$2.98b
	(-$119,828, $334,953)	(-1.60, 1.39)	(-58,194, 17,541)	(-$8.73b, $2.63b)
Policy 2a (TSR-CE): recommending pembro for CRC and endometrial cancers and chemo for the rest vs. chemo for all	$168,929	0.81	-5,683	-$0.85b
	(-$182,832, $727,198)	(-2.26, 3.93)	(-42,525, 30,211)	(-$6.38b, $4.53b)
Policy 3a (TSR-NHB): recommending pembro for CRC, endometrial, and brain cancers and chemo for the rest vs. chemo for all	$162,560	0.80	-4,934	-$0.74b
	(-$203,715, $742,360)	(-2.43, 4.05)	(-41,868, 31,100)	(-$6.28b, $4.67b)
VoS (ΔPolicy 2a vs. 1a)	-	-	14,194	$2.13b
			(3,889, 24,150)	($0.58b, $3.62b)
VoS (ΔPolicy 3a vs. 1a)	-	-	14,947	$2.24b
			(4,752, 24,766)	($0.71b, $3.71b)
BHM-derived effectiveness inputs				
Policy 1b (TAR): recommending pembro for all tumor types vs. chemo for all	$73,565	-0.06	-20,886	-$3.13b
	(-$108,077, $350,492)	(-1.56, 1.42)	(-59,284, 16,069)	(-$8.89b, $2.41b)
Policy 2b (TSR-CE): recommending pembro for CRC, endometrial, and small intestine cancers and chemo for the rest vs. chemo for all	$198,629	0.94	-6,888	-$1.03b
	(-$152,422, $769,376)	(-2.15, 4.03)	(-43,061, 28,323)	(-$6.46b, $4.25b)
Policy 3b (TSR-NHB): recommending pembro for CRC, endometrial, and brain cancers and chemo for the rest vs. chemo for all	$167,485	0.83	-4,903	-$0.74b
	(-$197,578, $763,772)	(-2.39, 4.03)	(-40,852, 30,217)	(-$6.13b, $4.53b)
VoS (ΔPolicy 2b vs. 1b)	-	-	13,998	$2.10b
			(3,759, 24,058)	($0.56, $3.61b)
VoS (ΔPolicy 3b vs. 1b)	-	-	15,983	$2.40b
			(5,673, 26,056)	($0.85b, $3.91b)

Note:

1. A willingness-to-pay threshold of $150k per QALY was used.
2. TAR: Tumor-aggregated recommendation preferred pembrolizumab for all cancers.
3. TSR-CE: Tumor-specific recommendation based on comparative effectiveness.
4. TSR-NHB: Tumor-specific recommendation based on positive INHB.

CR, credible range; CRC, colorectal cancer; EC, endometrial cancer; SI, small intestine cancers; INHB, incremental net health benefit; INMB, incremental net monetary benefit; b, billion; QALY, quality-adjusted life years.

Table 3.4. Aggregated incremental cost-effectiveness ratios and value-based prices

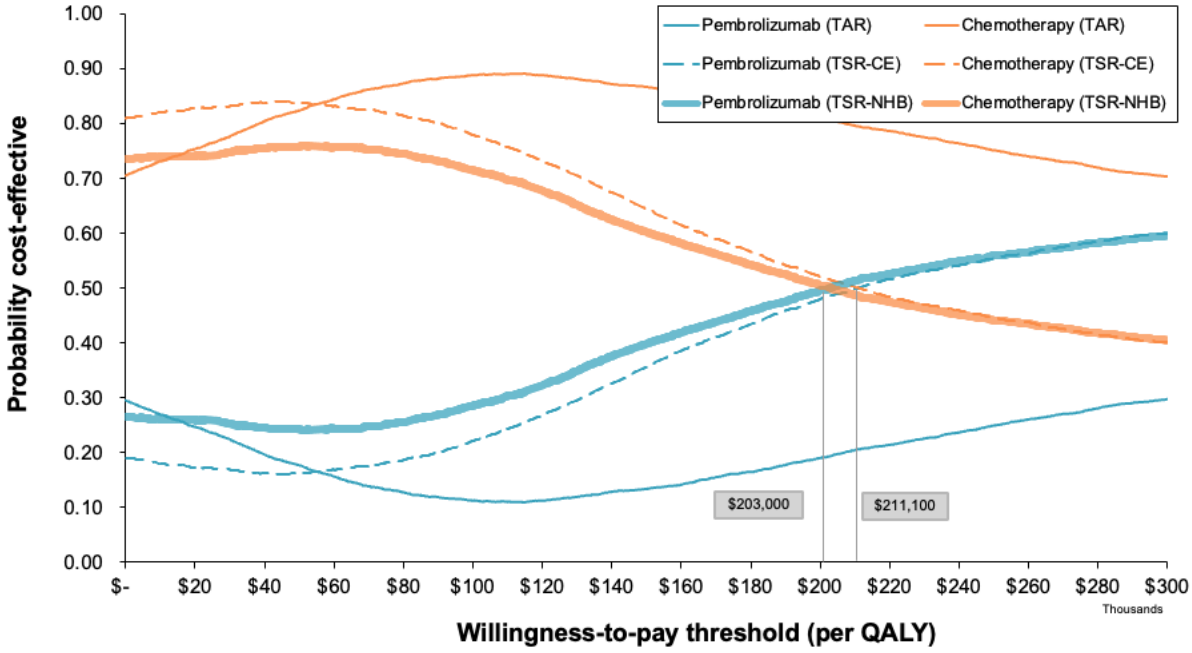
Policy Scenario	**ICER Mean** (95%CR)	**VBP at $50k/QALY** (% change from net price)	**VBP at $100k/QALY** (% change from net price)	**VBP at $150k/QALY** (% change from net price)
Policy 1b (TAR): recommending pembro for all tumor types vs. chemo for all	Less effective, more costly	-	-	-
Policy 2b (TSR-CE): recommending pembro for CRC, endometrial, and small intestine cancers and chemo for the rest vs. chemo for all	$117,138 (-$703,721, $838,017)	$8,401 (-41%)	$12,456 (-13%)	$16,512 (+15%)
Policy 3b (TSR-NHB): recommending pembro for CRC, endometrial, and brain cancers and chemo for the rest vs. chemo for all	$89,833 (-$527,641, $746,665)	$11,298 (-21%)	$15,353 (+7%)	$19,698 (+37%)

Note:

1. BHM-derived effectiveness inputs were applied to the pembrolizumab arms.

Abbreviations: CR: credible range; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years; VBP: value-based price.

Figure 3.1. Cost effectiveness acceptability curves



Notes:

1. TAR: Tumor-aggregated recommendation preferred pembrolizumab for all cancers.
2. TSR-CE: Tumor-specific recommendation based on comparative effectiveness preferred pembrolizumab for colorectal, endometrial, and small intestine cancers.
3. TSR-NHB: Tumor-specific recommendation based on positive INHB preferred pembrolizumab for colorectal, endometrial, and brain cancers.
4. BHM-derived effectiveness inputs were applied to the pembrolizumab arms.

Abbreviations: QALY, quality-adjusted life years.

3.9 SUPPLEMENT

Table S3.1. Patient demographic and clinical characteristics of pembrolizumab trials

Characteristics	**KEYNOTE 158 (N=233)**	**KEYNOTE 164 (N=63)**
Author, year	Marabelle, 2019	Le, 2019
Age, years		
Median (range)	60 (20-87)	59 (23-83)
Female sex, N (%)	137 (59%)	30 (52%)
Performance status, N (%)		
0	113 (49%)	22 (35%)
1	120 (52%)	41 (65%)
Prior lines of therapy, N (%)		
0	7 (3%)	0 (0%)
1	87 (37%)	24 (38%)
2	61 (26%)	20 (32%)
≥3	78 (34%)	19 (30%)
Prior (neo)adjuvant therapy	55 (24%)	17 (27%)
Tumor types, %		
Colorectal	0 (0%)	63 (100%)
Endometrial	49 (21%)	0 (0%)
Gastric	24 (10%)	0 (0%)
Cholangiocarcinoma	22 (9%)	0 (0%)
Pancreatic	22 (9%)	0 (0%)
Small intestine	19 (8%)	0 (0%)
Ovarian	15 (6%)	0 (0%)
Brain	13 (6%)	0 (0%)
Other	69 (30%)	0 (0%)

Table S3.2. Effectiveness inputs

	Pembrolizumab	Real-world Control
Policy 1a, 2a, 3a: parametric models plus trial-based median survival estimates	• Colorectal and endometrial: parametric models fitted to Kaplan-Meier survival curves • Other tumor types: exponential model fitted to trial-reported median OS or PFS	• Parametric models fitted to Kaplan-Meier survival curves
Policy 1b, 2b, 3b: parametric models plus BHM-derived median survival estimates	• Colorectal and endometrial: parametric models fitted to Kaplan-Meier survival curves • Other tumor types: exponential model fitted to BHM-derived median OS or PFS	• Parametric models fitted to Kaplan-Meier survival curves

OS, overall survival; PFS, progression-free survival; BHM, Bayesian hierarchical models.

Table S3.3. Estimated proportion of patients receiving chemotherapies and monthly drug costs by tumor types

Chemotherapy drug	Proportion of patients receiving each drug	Monthly drug costs based on dosing schedules	Weighted monthly drug costs
Colorectal			
fluorouracil (day 1)	73.32%	$228.20	$167.32
fluorouracil	73.32%	$3,253.61	$2,385.70
leucovorin	57.31%	$1,152.13	$660.32
bevacizumab	44.30%	$4,401.71	$1,949.86
oxaliplatin	40.95%	$478.61	$195.98
irinotecan	37.73%	$449.36	$169.52
capecitabine	22.23%	$3,118.99	$693.33
Endometrial			
carboplatin	67.40%	$229.81	$154.89
paclitaxel	60.79%	$4,439.39	$2,698.69
doxorubicin liposomal	17.08%	$1,119.00	$191.12
bevacizumab	15.15%	$8,397.40	$1,272.33
doxorubicin	14.51%	$426.00	$61.81
cisplatin	6.89%	$226.70	$15.61
docetaxel	6.89%	$363.02	$25.00
gemcitabine	3.58%	$2,579.03	$92.36
ifosfamide	1.74%	$1,122.54	$19.59
temsirolimus	1.38%	$4,448.57	$61.28
topotecan	1.19%	$7,389.68	$88.21
pemetrexed	0.46%	$10,296.72	$47.28
Gastric			
fluorouracil (day 1)	65.65%	$228.20	$149.81
fluorouracil	65.65%	$3,253.61	$2,135.96
leucovorin	46.56%	$782.95	$364.58
oxaliplatin	40.65%	$478.61	$194.55
capecitabine	18.51%	$3,118.99	$577.37
irinotecan	18.32%	$514.02	$94.17
paclitaxel	12.60%	$4,191.92	$527.99
carboplatin	10.50%	$229.81	$24.12
docetaxel	10.50%	$403.03	$42.30
cisplatin	9.16%	$367.87	$33.70
levoleucovorin	0.95%	$1,480.91	$14.13
Cholangiocarcinoma			
gemcitabine	48.69%	$3,512.57	$1,710.40
fluorouracil (day 1)	39.67%	$228.20	$90.52
fluorouracil	39.67%	$3,253.61	$1,290.63

cisplatin	36.58%	$429.71	$157.19
oxaliplatin	30.17%	$478.61	$144.38
leucovorin	26.84%	$731.23	$196.27
capecitabine	22.09%	$3,118.99	$688.99
irinotecan	14.25%	$449.36	$64.04
irinotecan liposome	0.71%	$12,677.62	$90.34
levoleucovorin	0.48%	$1,323.49	$6.29
Pancreatic			
fluorouracil (day 1)	59.58%	$228.20	$135.95
fluorouracil	59.58%	$3,253.61	$1,938.42
leucovorin	45.05%	$782.95	$352.69
oxaliplatin	37.85%	$478.61	$181.14
irinotecan	37.52%	$514.02	$192.84
gemcitabine	33.49%	$2,341.72	$784.18
capecitabine	20.28%	$3,118.99	$632.45
cisplatin	4.69%	$340.06	$15.95
levoleucovorin	2.05%	$0.00	$0.00
docetaxel	1.32%	$469.70	$6.20
Small intestine			
fluorouracil (day 1)	61.97%	$228.20	$141.42
fluorouracil	61.97%	$3,253.61	$2,016.32
leucovorin	41.78%	$782.95	$327.15
oxaliplatin	39.44%	$478.61	$188.75
irinotecan	35.21%	$514.02	$180.99
bevacizumab	21.60%	$4,401.71	$950.60
capecitabine	20.66%	$3,118.99	$644.30
paclitaxel	11.74%	$4,725.52	$554.64
gemcitabine	10.80%	$3,512.57	$379.29
carboplatin	5.16%	$229.81	$11.87
docetaxel	1.41%	$403.03	$5.68
Ovarian			
carboplatin	60.74%	$229.81	$139.60
paclitaxel	40.59%	$4,439.39	$1,801.85
doxorubicin liposomal	32.37%	$877.31	$283.98
bevacizumab	28.88%	$8,498.90	$2,454.46
doxorubicin	22.50%	$456.92	$102.80
gemcitabine	11.02%	$3,122.29	$344.05
cisplatin	5.92%	$226.70	$13.43
docetaxel	5.83%	$403.03	$23.50
topotecan	3.63%	$8,106.33	$294.03
cyclophosphamide	2.89%	$464.29	$13.43
fluorouracil (day 1)	2.02%	$152.13	$3.07
fluorouracil	2.02%	$2,063.10	$41.68

etoposide	1.74%	$1,499.20	$26.16
pemetrexed	1.56%	$8,053.67	$125.72
oxaliplatin	1.47%	$319.07	$4.69
leucovorin	1.29%	$384.04	$4.94
capecitabine	1.06%	$2,651.14	$28.00
ifosfamide	0.32%	$1,548.42	$4.98
melphalan	0.18%	$779.89	$1.43
vinorelbine	0.14%	$445.01	$0.61
vinblastine	0.09%	$377.52	$0.35
dactinomycin	0.05%	$5,137.48	$2.36
vincristine	0.05%	$501.81	$0.23
Brain			
bevacizumab	49.11%	$8,498.90	$4,173.57
temozolomide	32.59%	$4,936.15	$1,608.65
lomustine	19.20%	$399.84	$76.75
etoposide	17.41%	$1,998.93	$348.03
carboplatin	13.39%	$297.02	$39.78
carmustine	9.82%	$1,600.30	$157.17
cisplatin	5.80%	$161.14	$9.35
vincristine	4.91%	$83.64	$4.11
procarbazine	3.13%	$1,566.56	$48.95

Table S3.4. Pre-progression monitoring unit costs and frequency of resource use

Item	Unit cost (2022 USD)	Frequency per month	Source
Outpatient consultation	$ 183.07	0.5	[20, 35]
CT scan - year one	$ 512.17	0.5	[20, 35]
CT scan - year two +	$ 512.17	0.4	[20, 35]
Lab tests			
Immunohistochemistry test	$ 102.43	1.0	One time
Complete blood count	$ 8.58	1.4	[20, 35]; trial protocols
Hepatic function tests	$ 9.85	1.4	[20, 35]; trial protocols
Renal function tests	$ 10.47	1.4	[20, 35]; trial protocols
CA 19-9 test (pancreatic cancer)	$ 117.50	1.4	Assumed the same with other lab tests
CEA test (colorectal cancer)	$ 295.77	1.4	[20, 35]; trial protocols

Table S3.5. Epidemiology inputs

	Cancer incidence (2023)	**% metastatic diseases**	**% receiving ≥2L therapies**	**Metastatic cancer incidence among those progressed on ≥1L therapies**
Colorectal	151,030	21%	48%	15,379
Endometrial	62,653	4.5%	46%	1,287
Gastric	26,380	34.1%	36%	3,242
Cholangiocarcinoma	8,000	40.4%	34%	1,098
Pancreatic	62,210	48.7%	30%	9,096
Small intestine	11,790	27.1%	40%	1,280
Ovarian	19,880	50.6%	57%	5,733
Brain	88,970	2.2%	27%	540

Table S3.6. Total costs by cost categories and by tumor types

Tumor type	Treatment	Drug Acquisition and Administration Costs	Treatment Associated Adverse Event Costs	Healthcare Costs		Total Costs
				Pre-Progression Monitoring Costs	Post-Progression Treatment Costs	
Colorectal	Pembro	$173,866	$1,356	$44,146	$182,197	$401,565
	Chemo	$15,373	$7,311	$32,283	$287,738	$342,705
Endometrial	Pembro	$196,893	$1,356	$26,023	$518,127	$742,400
	Chemo	$9,111	$7,311	$14,978	$209,208	$240,608
Gastric	Pembro	$66,782	$1,356	$1,825	$170,947	$240,909
	Chemo	$8,024	$7,311	$11,448	$182,918	$209,700
Cholangiocarcinoma	Pembro	$84,955	$1,356	$2,299	$222,561	$311,170
	Chemo	$9,761	$7,311	$3,619	$184,926	$205,616
Pancreatic	Pembro	$46,282	$1,356	$1,842	$42,507	$91,987
	Chemo	$9,706	$7,311	$5,303	$208,095	$230,415
Small Intestine	Pembro	$232,639	$1,356	$9,388	$549,014	$792,396
	Chemo	$12,862	$7,311	$6,578	$285,082	$311,833
Ovarian	Pembro	$48,114	$1,356	$1,347	$551,760	$602,577
	Chemo	$12,503	$7,311	$3,682	$489,138	$512,634
Brain	Pembro	$27,537	$1,356	$817	$152,361	$182,071
	Chemo	$16,826	$7,311	$4,131	$324,436	$352,703

Pembro, pembrolizumab; Chemo, chemotherapy.

Figure S3.1. Partitioned survival model structure

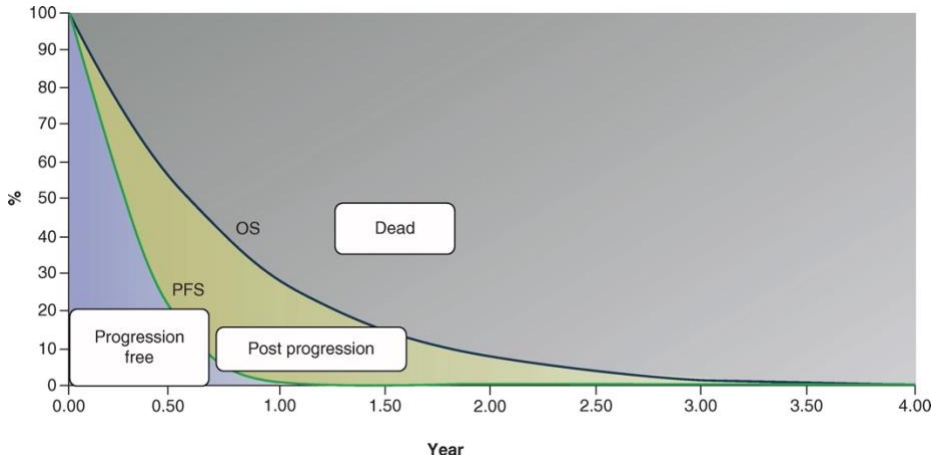
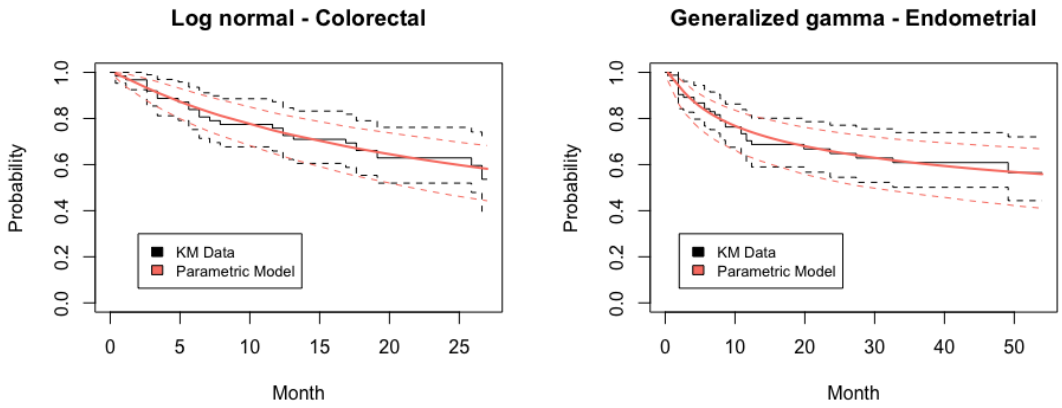


Figure S3.2. Parametric models fitted to pembrolizumab survival data in colorectal and endometrial cancers

a. Overall survival



b. Progression-free survival

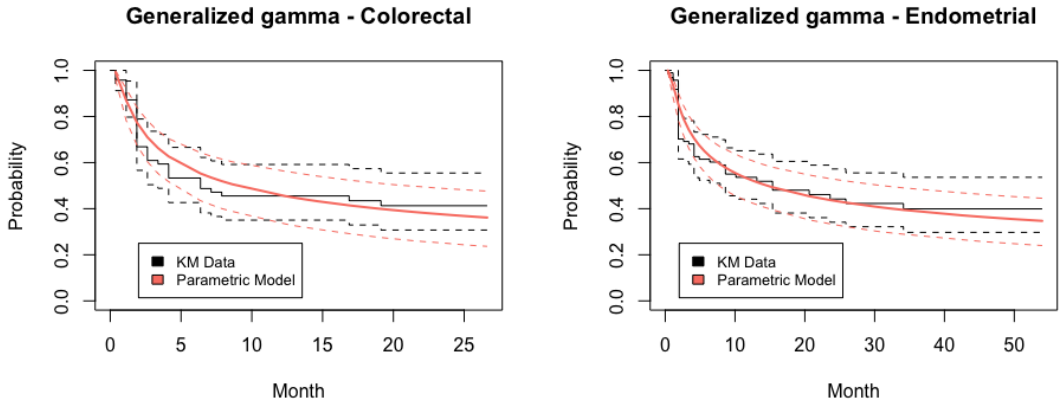
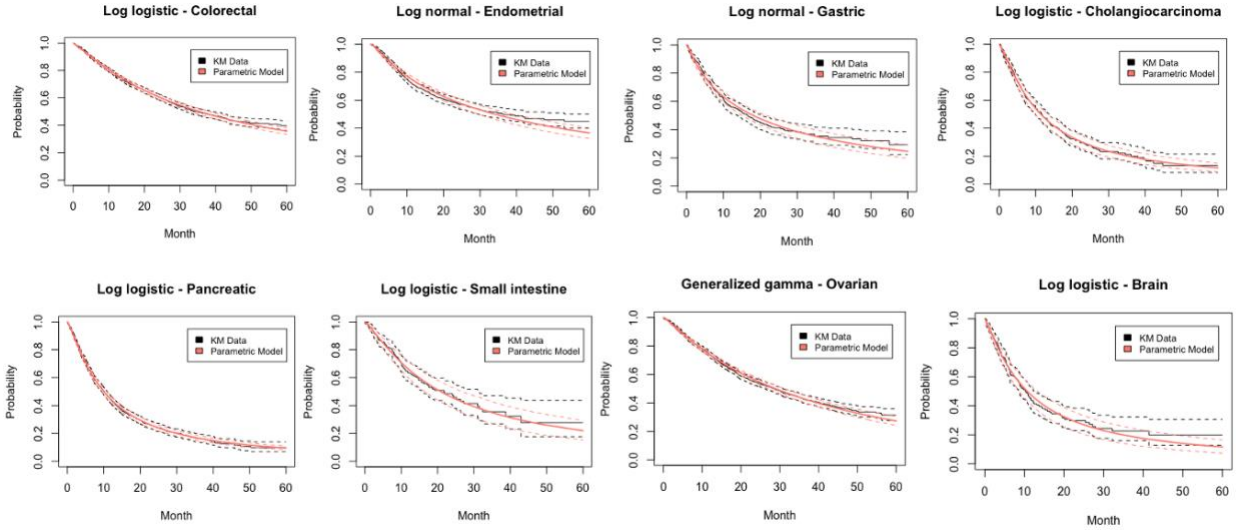


Figure S3.3. Parametric models fit to chemotherapy cohorts

a. Overall survival



b. Progression-free survival

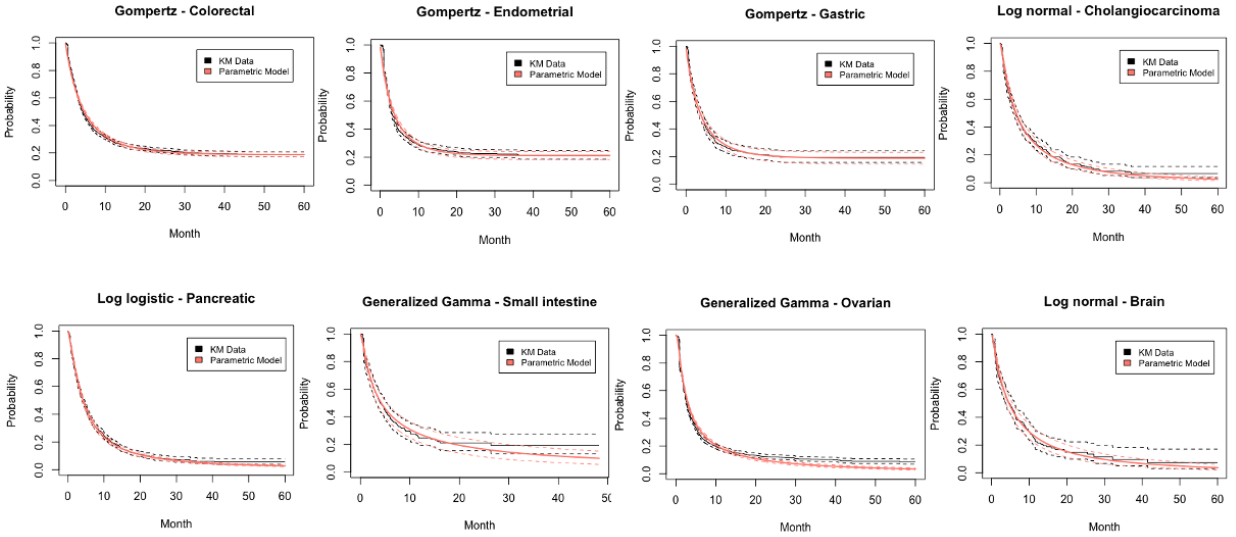


Figure S3.4. Hazard functions for overall survival among chemotherapy cohorts

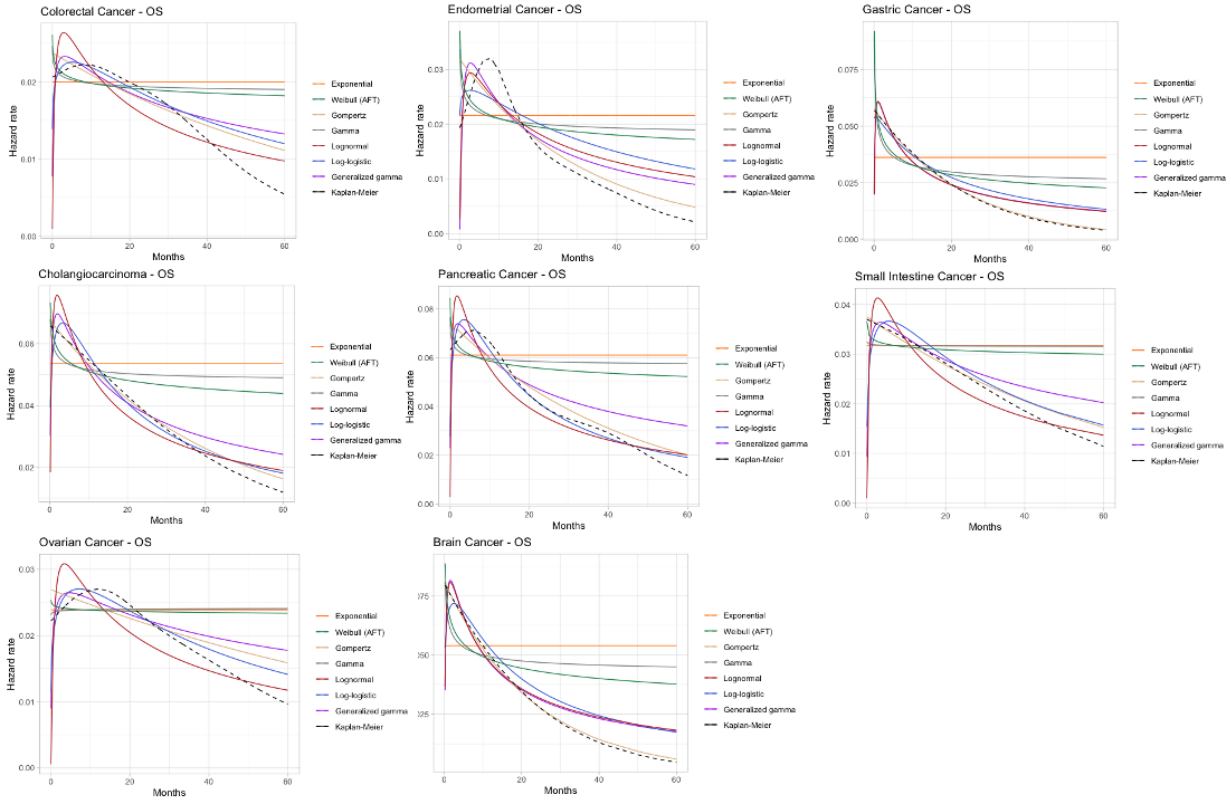


Figure S3.5. Hazard functions for progression-free survival among chemotherapy cohorts

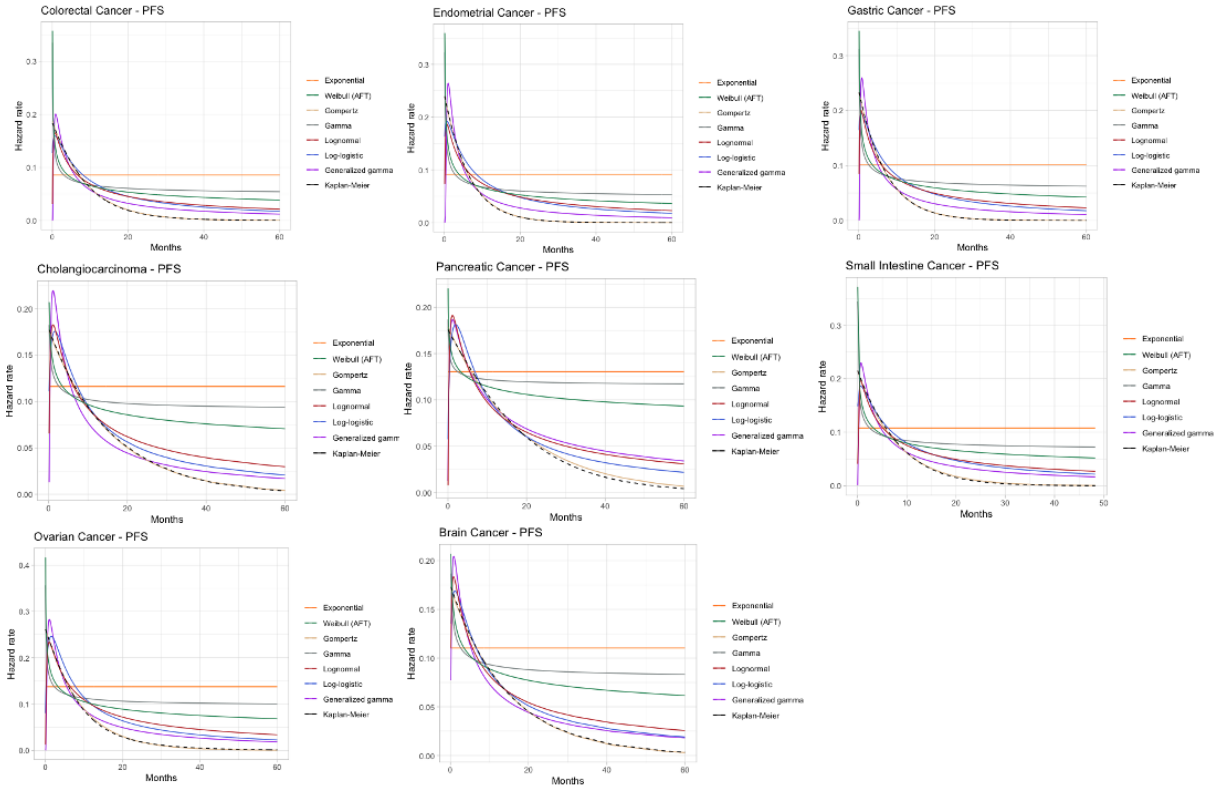


Figure S3.6. Time to treatment discontinuation among chemotherapy cohorts

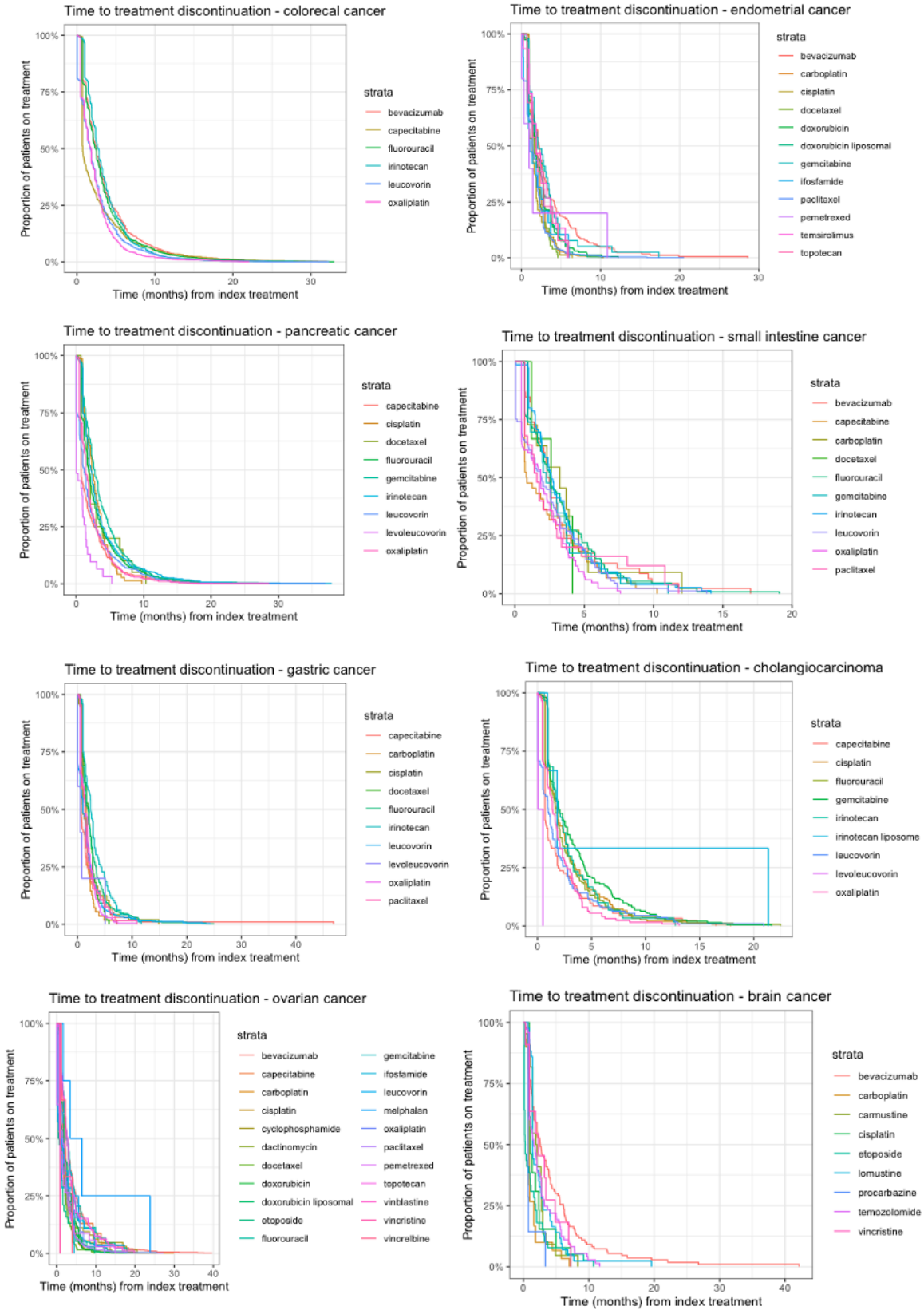
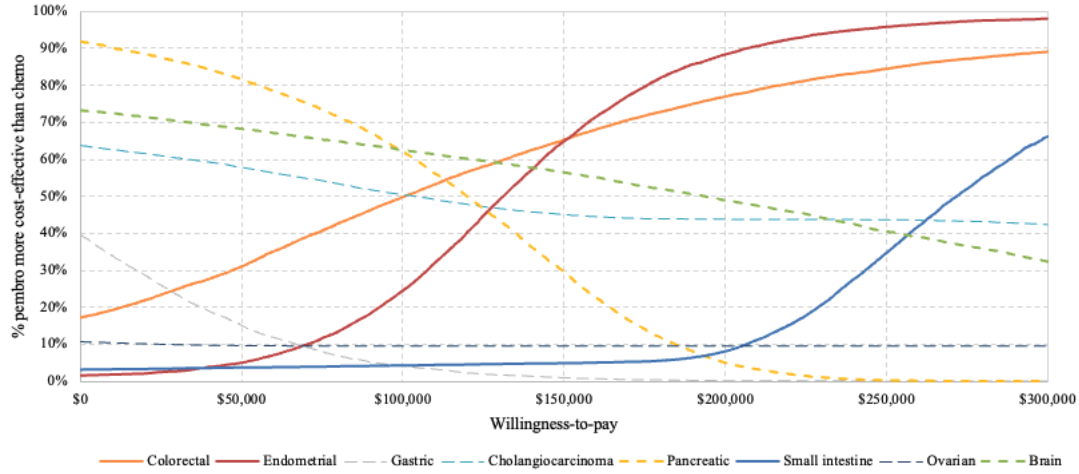
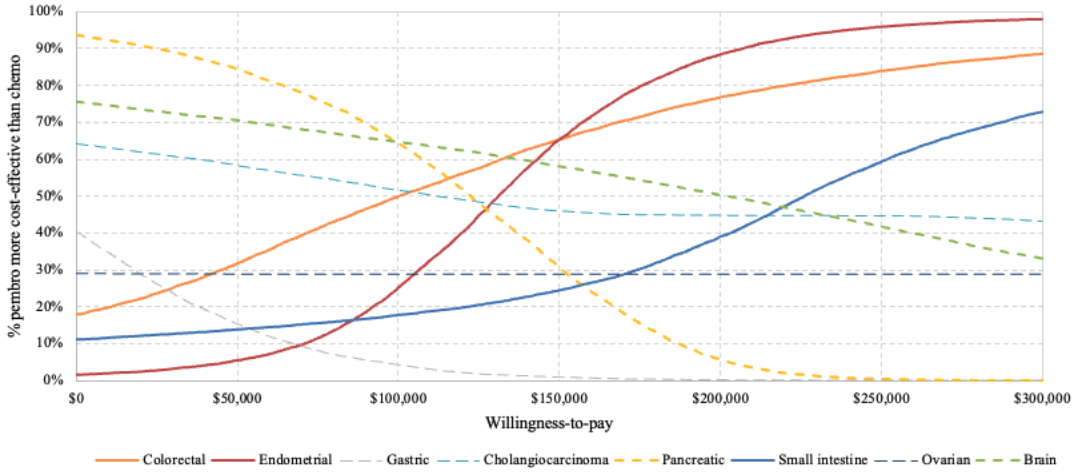


Figure S3.7. Cost effectiveness acceptability curve by tumor types

a. BHM-derived effectiveness inputs



b. Raw effectiveness inputs

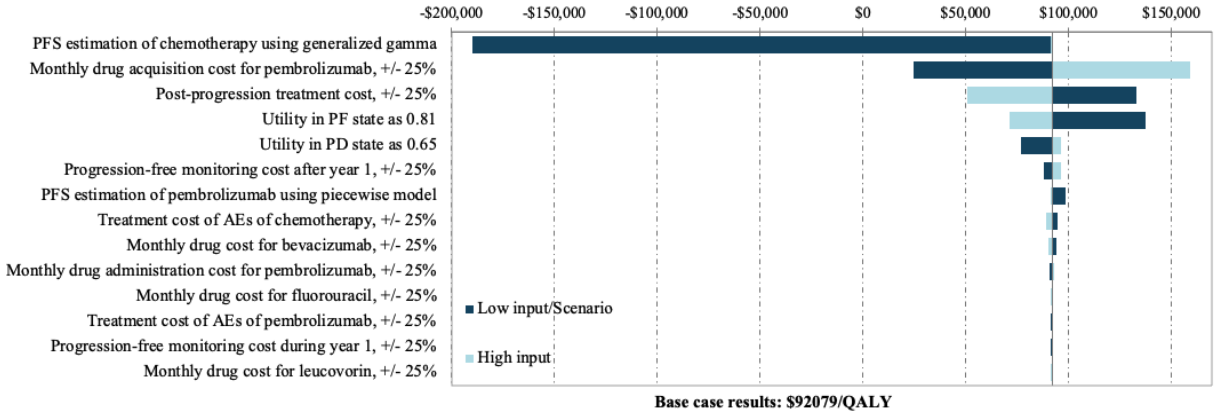


Notes:

1. Dashed lines indicate tumor types where pembrolizumab was shown to be less effective than chemotherapies in the base case.

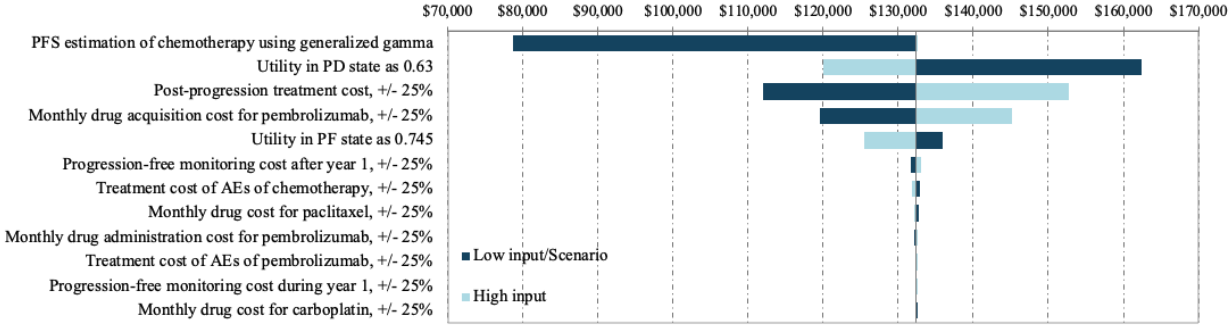
Figure S3.8. Tornado Diagram for Colorectal and Endometrial Cancers

a. Colorectal cancer



Base case results: $92079/QALY

b. Endometrial cancer



Base case results: $132391/QALY

BIBLIOGRAPHY

1. Nelsen, R.B., An introduction to copulas. 2006: Springer.
2. Westreich, D., et al., Transportability of Trial Results Using Inverse Odds of Sampling Weights. Am J Epidemiol, 2017. 186(8): p. 1010-1014.
3. Woods, B., et al., Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. 2017, Sheffield: Decision Support Unit, ScHARR, University of Sheffield.
4. Colwell, J., NCI-MATCH Trial Draws Strong Interest. Cancer Discov, 2016. 6(4): p. 334.
5. Garber, K., Tissue-agnostic cancer drug pipeline grows, despite doubts. Nat Rev Drug Discov, 2018. 17(4): p. 227-229.
6. Ramsey, S.D., V. Shankaran, and S.D. Sullivan, Basket Cases: How Real-World Testing for Drugs Approved Based on Basket Trials Might Lead to False Diagnoses, Patient Risks, and Squandered Resources. J Clin Oncol, 2019. 37(36): p. 3472-3474.
7. US Food and Drug Administration. FDA grants accelerated approval to pembrolizumab for first tissue/site agnostic indication. 2017 [cited 2021 January 17]; Available from: https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm560040.htm.
8. US Food and Drug Administration. FDA grants accelerated approval to dabrafenib in combination with trametinib for unresectable or metastatic solid tumors with BRAF V600E mutation. 2022; Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-dabrafenib-combination-trametinib-unresectable-or-metastatic-solid.
9. US Food and Drug Administration. FDA grants accelerated approval to dostarlimab-gxly for dMMR advanced solid tumors. 2021 [cited 2022 Jan 20]; Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-grants-accelerated-approval-dostarlimab-gxly-dmmr-advanced-solid-tumors.
10. US Food and Drug Administration. Approval package for application number 212725 Orig1s000, 212726Orig1s000. Trade name: ROZLYTREK capsules, 100mg and 200 mg (entrectinib) 2019; Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/212725Orig1s000,%2021272 6Orig1s000Approv.pdf.
11. US Food and Drug Administration. Drug approval package: Vitrakvi (larotrectinib). 2018 [cited 2020 January 17]; Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210861Orig1s000_21171Orig 1s000TOC.cfm.
12. Boyiadzis, M.M., et al., Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. J Immunother Cancer, 2018. 6(1): p. 35.
13. Li, M.-C. and H.-P. Hsieh, Tumor-agnostic inhibitors in oncology: A new phase for precision medicine. Journal of the Chinese Chemical Society, 2020. 67(12): p. 2216-2224.

14. Offin, M., D. Liu, and A. Drilon, Tumor-Agnostic Drug Development. Am Soc Clin Oncol Educ Book, 2018. 38: p. 184-187.

15. Redig, A.J. and P.A. Jänne, Basket trials and the evolution of clinical trial design in an era of genomic medicine. 2015. 33: p. 975-977.

16. Dubuque, C.M., P; Suarez, S; Scaife, JG; DeWoskin, V;, Oncology basket trials: An emerging paradigm shift in trial design & treatment approach. 2018.

17. Delimpaltadakis, N., et al., PCN262 Understanding the Challenges in the Evaluation of Tumour-Agnostic Therapies: Barriers Posed By the Current Trial Designs and Future Trends. Value in Health, 2020. 23: p. S469.

18. NICE, Larotrectinib for treating NTRK fusion-positive solid tumours. Technology appraisal guidance [TA630]. 2020.

19. Maio, M., et al., Pembrolizumab in microsatellite instability high or mismatch repair deficient cancers: updated analysis from the phase II KEYNOTE-158 study. Ann Oncol, 2022. 33(9): p. 929-938.

20. Marabelle, A., et al., Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair-Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. J Clin Oncol, 2020. 38(1): p. 1-10.

21. Spiegelhalter, D.A., KR; Myles, J;, Bayesian approaches to clinical trials and Health-Care Evaluation. 2004, New York: Wiley.

22. Gelman, A., et al., Bayesian Data Analysis (3rd ed.). . 2003: Chapman and Hall/CRC.

23. Thall, P.F., et al., Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. Stat Med, 2003. 22(5): p. 763-80.

24. Lipsky, A.M., et al., The importance of "shrinkage" in subgroup analyses. Ann Emerg Med, 2010. 55(6): p. 544-552 e3.

25. Thall, P.F. and J.K. Wathen, Bayesian designs to account for patient heterogeneity in phase II clinical trials. Curr Opin Oncol, 2008. 20(4): p. 407-11.

26. McGlothlin, A.E. and K. Viele, Bayesian Hierarchical Models. JAMA, 2018. 320(22): p. 2365-2366.

27. McElreath, R., Statistical rethinking : a Bayesian course with examples in R and Stan / Richard McElreath. Second edition ed. Chapman & Hall/CRC texts in statistical science series. 2020, Boca Raton: CRC Press.

28. Higgins, J.P. and S.G. Thompson, Quantifying heterogeneity in a meta-analysis. Stat Med, 2002. 21(11): p. 1539-58.

29. Bernardo, J.M. The Concept of Exchangeability and its Applications. 2001.

30. Hannigan, L.J., et al., Improving the Estimation of Subgroup Effects for Clinical Trial Participants with Multimorbidity by Incorporating Drug Class-Level Information in Bayesian Hierarchical Models: A Simulation Study. Med Decis Making, 2022. 42(2): p. 228-240.

31. Schmitz, S., R. Adams, and C. Walsh, Incorporating data from various trial designs into a mixed treatment comparison model. Statistics in medicine, 2013. 32(17): p. 2935-2949.

32. Zhang, J., et al., Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. Stat Methods Med Res, 2019. 28(5): p. 1293-1310.

33. Luo, S. and J. Wang, Bayesian hierarchical model for multiple repeated measures and survival data: an application to Parkinson's disease. Stat Med, 2014. 33(24): p. 4279-91.

34. Jiang, L., et al., Optimal Bayesian hierarchical model to accelerate the development of tissue-agnostic drugs and basket trials. Contemp Clin Trials, 2021. 107: p. 106460.

35. Le, D.T., et al., Phase II Open-Label Study of Pembrolizumab in Treatment-Refractory, Microsatellite Instability-High/Mismatch Repair-Deficient Metastatic Colorectal Cancer: KEYNOTE-164. J Clin Oncol, 2020. 38(1): p. 11-19.

36. Borenstein, M., et al., A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods, 2010. 1(2): p. 97-111.

37. Cunanan, K.M., et al., Variance prior specification for a basket trial design using Bayesian hierarchical modeling. Clin Trials, 2019. 16(2): p. 142-153.

38. Killip, S., Z. Mahfoud, and K. Pearce, What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med, 2004. 2(3): p. 204-8.

39. O'Malley, D.M., et al., Pembrolizumab in Patients With Microsatellite Instability-High Advanced Endometrial Cancer: Results From the KEYNOTE-158 Study. J Clin Oncol, 2022. 40(7): p. 752-761.

40. Stigler, S.M., Studies in the History of Probability and Statistics. XXXII: Laplace, Fisher, and the discovery of the concept of sufficiency. Biometrika, 1973. 60(3): p. 439-445.

41. Rider, P.R., Variance of the median of small samples from several special populations. Journal of the American Statistical Association, 1960. 55(289): p. 148-150.

42. Gelman, A., Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 2006. 1(3): p. 18.

43. Kwok, H. and R.J. Lewis, Bayesian hierarchical modeling and the integration of heterogeneous information on the effectiveness of cardiovascular therapies. Circ Cardiovasc Qual Outcomes, 2011. 4(6): p. 657-66.

44. Zou, G. and A. Donner, Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. Biometrics, 2004. 60(3): p. 807-11.

45. Ridout, M.S., C.G. Demetrio, and D. Firth, Estimating intraclass correlation for binary data. Biometrics, 1999. 55(1): p. 137-48.

46. Chen, E.Y., V. Raghunathan, and V. Prasad, An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate. JAMA Intern Med, 2019. 179(7): p. 915-921.

47. Chugh, R., et al., Phase II multicenter trial of imatinib in 10 histologic subtypes of sarcoma using a Bayesian hierarchical statistical model. 2009. 27: p. 3148-3153.

48. Murphy, P., et al., Exploring heterogeneity in histology-independent technologies and the implications for cost-effectiveness. 2021. 41: p. 165-178.

49. Neuenschwander, B., et al., Robust exchangeability designs for early phase clinical trials with multiple strata. Pharm Stat, 2016. 15(2): p. 123-34.

50. James, W. and C. Stein, Estimation with quadratic loss. Breakthroughs in statistics: Foundations and basic theory, 1992: p. 443-460.

51. Chu, W.M., et al., Risk of major comorbidities among workers with hemophilia: A 14-year population-based study. Medicine (Baltimore), 2018. 97(6): p. e9803.

52. Raudenbush, S.W. and A.S. Bryk, Empirical Bayes Meta-Analysis. Journal of Educational Statistics, 1985. 10(2): p. 75-98.

53. Sutton, A.J. and K.R. Abrams, Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res, 2001. 10(4): p. 277-303.

54. Food, U.S., et al., Guidance for Industry Expedited Programs for Serious Conditions – Drugs and Biologics. 2014, FDA: Silver Spring, MD.

55. Agrawal, S., et al., Use of Single-Arm Trials for US Food and Drug Administration Drug Approval in Oncology, 2002-2021. JAMA Oncol, 2023. 9(2): p. 266-272.

56. Krebs, M.G., et al., Intrapatient comparisons of efficacy in a single-arm trial of entrectinib in tumour-agnostic indications. ESMO Open, 2021. 6(2): p. 100072.

57. Usmani, S.Z., et al., Daratumumab monotherapy compared with historical control data in heavily pretreated and highly refractory patients with multiple myeloma: An adjusted treatment comparison. American Journal of Hematology, 2017. 92(8): p. E146-E152.

58. Martino, M., et al., Mobilization of hematopoietic progenitor stem cells in allogeneic setting with lenograstim by subcutaneous injection, in daily or twice-daily dosing: a single-center prospective study with historical control. Transfusion, 2015. 55(8): p. 2032-8.

59. Isayama, H., et al., Comparison of partially covered nitinol stents with partially covered stainless stents as a historical control in a multicenter study of distal malignant biliary obstruction: the WATCH study. Gastrointest Endosc, 2012. 76(1): p. 84-92.

60. Stefanelli, A., et al., Can we decrease the acute proctitis in prostate cancer patients using hyaluronic acid during radiation therapy: a prospective historically controlled clinical study. Eur Rev Med Pharmacol Sci, 2012. 16(5): p. 639-45.

61. Kawaguchi, T., et al., A multi-institutional phase II trial of consolidation S-1 after concurrent chemoradiotherapy with cisplatin and vinorelbine for locally advanced non-small cell lung cancer. Eur J Cancer, 2012. 48(5): p. 672-7.

62. Thorlund, K., et al., Synthetic and External Controls in Clinical Trials - A Primer for Researchers. Clin Epidemiol, 2020. 12: p. 457-467.

63. Mishra-Kalyani, P.S., et al., External control arms in oncology: current use and future directions. Ann Oncol, 2022. 33(4): p. 376-383.

64. Franklin, J.M., et al., Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. Clin Pharmacol Ther, 2019. 105(4): p. 867-877.

65. Khozin, S., G.M. Blumenthal, and R. Pazdur, Real-world Data for Clinical Evidence Generation in Oncology. J Natl Cancer Inst, 2017. 109(11).

66. Arondekar, B., et al., Real-World Evidence in Support of Oncology Product Registration: A Systematic Review of New Drug Application and Biologics License Application Approvals from 2015-2020. Clin Cancer Res, 2022. 28(1): p. 27-35.

67. Hatswell, A.B., G; Berlin, JA; Irs, A; Freemantle, N;, Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. BMJ Open, 2016. 6: p. e011666.

68. Patel, D., et al., Use of External Comparators for Health Technology Assessment Submissions Based on Single-Arm Trials. Value Health, 2021. 24(8): p. 1118-1125.

69. United States Congress. 21st Century Cures Act. 2016; Available from: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf.

70. Mack, C., et al., When Context Is Hard to Come By: External Comparators and How to Use Them. Ther Innov Regul Sci, 2020. 54(4): p. 932-938.

71. Food, U.S., et al., Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics. Guidance for Industry. 2018, FDA: Silver Spring, MD.

72. US Food and Drug Administration. Framework for FDA's Real-World Evidence Program. 2018; Available from: https://www.fda.gov/media/120060/download.

73. Friends of Cancer Research. CHARACTERIZING THE USE OF EXTERNAL CONTROLS FOR AUGMENTING RANDOMIZED CONTROL ARMS AND CONFIRMING BENEFIT. 2019 [cited 2022 December 1]; Available from: https://friendsofcancerresearch.org/wp-content/uploads/Panel-1_External_Control_Arms2019AM_2.pdf.

74. Burcu, M., et al., Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. Pharmacoepidemiol Drug Saf, 2020. 29(10): p. 1228-1235.

75. Seeger, J.D., et al., Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. Pharmacoepidemiol Drug Saf, 2020. 29(11): p. 1382-1392.

76. Hess, L.M., et al., Defining treatment regimens and lines of therapy using real-world data in oncology. Future Oncol, 2021. 17(15): p. 1865-1877.

77. Liu, J., et al., Real-world utilization and outcomes of systemic therapy among patients with advanced or recurrent endometrial cancer in the United States. Curr Med Res Opin, 2022. 38(11): p. 1935-1945.

78. Barzi, A., et al., Real-World Dosing Patterns and Outcomes of Patients With Metastatic Pancreatic Cancer Treated With a Liposomal Irinotecan Regimen in the United States. Pancreas, 2020. 49(2): p. 193-200.

79. George, D.J., et al., Treatment Patterns and Outcomes in Patients With Metastatic Castration-resistant Prostate Cancer in a Real-world Clinical Practice Setting in the United States. Clin Genitourin Cancer, 2020. 18(4): p. 284-294.

80. Bikov, K.A., et al., Algorithm for identifying chemotherapy/biological regimens for metastatic colon cancer in SEER-Medicare. Med Care, 2015. 53(8): p. e58-64.

81. Hatswell, A.J., et al., Approaches to Selecting "Time Zero" in External Control Arms with Multiple Potential Entry Points: A Simulation Study of 8 Approaches. Med Decis Making, 2022. 42(7): p. 893-905.

82. Rizal, J., et al., The Application of Copula Continuous Extension Technique for Bivariate Discrete Data: A Case Study on Dependence Modeling of Seismicity Data. Mathematical Modelling of Engineering Problems, 2021. 8(5).

83. Nikoloulopoulos, A., On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. Journal of Statistical Planning and Inference, 2013. 143(11): p. 14.

84. Brookhart, M.A., et al., Variable selection for propensity score models. Am J Epidemiol, 2006. 163(12): p. 1149-56.

85. Austin, P.C., An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res, 2011. 46(3): p. 399-424.

86. Guyot, P., et al., Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol, 2012. 12: p. 9.

87. Jemielita, T., et al., Overall survival with second-line pembrolizumab in patients with non–small-cell lung cancer: Randomized phase III clinical trial versus propensity-adjusted real-world data. JCO clinical cancer informatics, 2021. 5: p. 56-65.

88. Coleman, R.L., et al., Survival outcomes for dostarlimab and real-world (RW) treatment (tx) paradigms in post-platinum patients (pts) with advanced/recurrent (A/R) endometrial cancer (EC): The GARNET trial versus an external control arm from the Flatiron Health database. Journal of Clinical Oncology, 2022. 40(16_suppl): p. 5593-5593.

89. Carrigan, G., et al., Using Electronic Health Records to Derive Control Arms for Early Phase Single-Arm Lung Cancer Trials: Proof-of-Concept in Randomized Controlled Trials. Clin Pharmacol Ther, 2020. 107(2): p. 369-377.

90. Davies, J., et al., Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. J Comp Eff Res, 2018. 7(9): p. 855-865.

91. Wilkinson, S., et al., Assessment of Alectinib vs Ceritinib in ALK-Positive Non-Small Cell Lung Cancer in Phase 2 Trials and in Real-world Data. JAMA Netw Open, 2021. 4(10): p. e2126306.

92. Torres, A.Z., et al., Analysis of a Real-World Progression Variable and Related Endpoints for Patients with Five Different Cancer Types. Adv Ther, 2022. 39(6): p. 2831-2849.

93. Walker, B., et al., Comparisons of Real-World Time-to-Event End Points in Oncology Research. JCO Clin Cancer Inform, 2021. 5: p. 45-46.

94. Dron, L., et al., Minimizing control group allocation in randomized trials using dynamic borrowing of external control data - An application to second line therapy for non-small cell lung cancer. Contemp Clin Trials Commun, 2019. 16: p. 100446.

95. Lim, J., et al., Minimizing Patient Burden Through the Use of Historical Subject-Level Data in Innovative Confirmatory Clinical Trials: Review of Methods and Opportunities. Ther Innov Regul Sci, 2018. 52(5): p. 546-559.

96. Ventz, S., et al., The design and evaluation of hybrid controlled trials that leverage external data and randomization. Nat Commun, 2022. 13(1): p. 5783.

97. Ventz, S., et al., Design and Evaluation of an External Control Arm Using Prior Clinical Trials and Real-World Data. Clin Cancer Res, 2019. 25(16): p. 4993-5001.

98. Italiano, A., et al., Larotrectinib versus Prior Therapies in Tropomyosin Receptor Kinase Fusion Cancer: An Intra-Patient Comparative Analysis. Cancers (Basel), 2020. 12(11).

99. Jin, Z. and F.A. Sinicrope, Prognostic and Predictive Values of Mismatch Repair Deficiency in Non-Metastatic Colorectal Cancer. Cancers (Basel), 2021. 13(2).

100. Guan, W.-L., et al., The Impact of Mismatch Repair Status on Prognosis of Patients With Gastric Cancer: A Multicenter Analysis. Frontiers in Oncology, 2021. 11(4885).

101. Trivedi, P.K. and D.M. Zimmer, Copula modeling: an introduction for practitioners. Foundations and Trends® in Econometrics, 2007. 1(1): p. 1-111.

102. Denuit, M. and P. Lambert, Constraints on concordance measures in bivariate discrete data. Journal of Multivariate Analysis, 2005. 93(1): p. 40-57.

103. Machado, J.A.F. and J.M.C.S. Silva, Quantiles for Counts. Journal of the American Statistical Association, 2005. 100(472): p. 1226-1237.

104. Benali, F., et al. MTCopula: Synthetic Complex Data Generation Using Copula. in 23rd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP). 2021. Nicosia, Cyprus.

105. ATKearney. Preparing health systems for tumor-agnostic treatment. 2019 [cited 2022 December 1]; Available from: https://www.kearney.com/industry/health/article/-/insights/preparing-health-systems-for-tumour-agnostic-treatment.

106. Horgan, D., et al., Bringing Onco-Innovation to Europe's Healthcare Systems: The Potential of Biomarker Testing, Real World Evidence, Tumour Agnostic Therapies to Empower Personalised Medicine. Cancers (Basel), 2021. 13(3).

107. Merola, D., et al., Oncology Drug Effectiveness from Electronic Health Record Data Calibrated Against RCT Evidence: The PARSIFAL Trial Emulation. Clinical Epidemiology, 2022: p. 1135-1144.

108. Huygens, S., et al., Cost-Effectiveness Analysis of Treating Patients With NTRK-Positive Cancer With the Histology-Independent Therapy Entrectinib. Value Health, 2023. 26(2): p. 193-203.

109. Espinoza, M.A., et al., The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. 2014. 34: p. 951-964.

110. Coyle, D., M.J. Buxton, and B.J. O'Brien, Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. Health Econ, 2003. 12(5): p. 421-7.

111. Michels, R.E., et al., Economic Evaluation of a Tumour-Agnostic Therapy: Dutch Economic Value of Larotrectinib in TRK Fusion-Positive Cancers. Appl Health Econ Health Policy, 2022. 20(5): p. 717-729.

112. Briggs, A., et al., Comparison of Alternative Methods to Assess the Cost-Effectiveness of Tumor-Agnostic Therapies: A Triangulation Approach Using Larotrectinib as a Case Study. Value Health, 2022. 25(6): p. 1002-1009.

113. Haines, P., G. Tremblay, and A. Briggs, A criterion-based approach for the systematic and transparent extrapolation of clinical trial survival data. Journal of Health Economics and Outcomes Research, 2015. 2(2): p. 147-160.

114. National vital statistics system. 2022; Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr71/nvsr71-01.pdf.

115. CMS. Physician Fee Schedule. Available from: https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeesched.

116. Overman, M.J., et al., Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. Lancet Oncol, 2017. 18(9): p. 1182-1191.

117. NCCN guideline. Colon Cancer. Available from: https://www.nccn.org/patients/guidelines/colon/files/assets/common/downloads/files/colon.pdf.

118. Mariotto, A.B., et al., Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst, 2011. 103(2): p. 117-28.

119. Shabaruddin, F.H., et al., A systematic review of utility values for chemotherapy-related adverse events. Pharmacoeconomics, 2013. 31(4): p. 277-88.

120. HCUPnet. AHRQ Data Tools. Available from: https://datatools.ahrq.gov/hcupnet.

121. SEER. SEER statistics. Available from: https://seer.cancer.gov/statistics/.

122. Siegel, R.L., et al., Cancer statistics, 2022. CA Cancer J Clin, 2022. 72(1): p. 7-33.

123. American Cancer Society. Key Statistics for Bile Duct Cancer. 2022 [cited 2022 December 1]; Available from: https://www.cancer.org/cancer/bile-duct-cancer/about/key-statistics.html.

124. Lengline, E., et al., Basket clinical trial design for targeted therapies for cancer: a French National Authority for Health statement for health technology assessment. Lancet Oncol, 2021. 22(10): p. e430-e434.

125. Cooper, S., et al., How should we assess the clinical and cost effectiveness of histology independent cancer drugs? BMJ, 2020. 368: p. l6435.

126. Carlson, J.J., S. Chen, and L.P. Garrison, Jr., Performance-Based Risk-Sharing Arrangements: An Updated International Review. Pharmacoeconomics, 2017. 35(10): p. 1063-1072.

127. Yeung, K., M. Li, and J. Carlson, Using Performance-Based Risk-Sharing Arrangements to Address Uncertainty in Indication-Based Pricing. J Manag Care Spec Pharm 2017. 23(10): p. 5.

128. Chen, Y., et al., Designing a Value-Based Formulary for a Commercial Health Plan: A Simulated Case Study of Diabetes Medications. Value Health, 2023. 26(7): p. 1022-1031.

129. Rosenbaum, P.R., D, Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association, 1984. 79(387): p. 8.

130. Palmer, S., et al., A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies. Value Health, 2023. 26(2): p. 185-192.

131. Congressional Budget Office. Prescription Drugs: Spending, Use, and Prices. 2022 [cited 2022 Aug 18]; Available from: https://www.cbo.gov/publication/57772#footnote-073.

132. Dell'Aquila, E., et al., Prognostic and predictive factors in pancreatic cancer. Oncotarget, 2020. 11(10): p. 924-941.

133. Nopel-Dunnebacke, S., et al., Prognostic value of microsatellite instability and p53 expression in metastatic colorectal cancer treated with oxaliplatin and fluoropyrimidine-based chemotherapy. Z Gastroenterol, 2014. 52(12): p. 1394-401.

134. Des Guetz, G., et al., Microsatellite instability does not predict the efficacy of chemotherapy in metastatic colorectal cancer. A systematic review and meta-analysis. Anticancer Res, 2009. 29(5): p. 1615-20.

135. Kim, S.Y., et al., The benefit of microsatellite instability is attenuated by chemotherapy in stage II and stage III gastric cancer: Results from a large cohort with subgroup analyses. Int J Cancer, 2015. 137(4): p. 819-25.

136. Webber, E.M., et al., Systematic review of the predictive effect of MSI status in colorectal cancer patients undergoing 5FU-based chemotherapy. BMC Cancer, 2015. 15: p. 156.

137. U.S. Department of Veterans Affairs. VA Federal Supply Schedule Service. Available from: https://www.fss.va.gov/.

138. Hoyle, M., et al., Cost-effectiveness of cetuximab, cetuximab plus irinotecan, and panitumumab for third and further lines of treatment for KRAS wild-type patients with metastatic colorectal cancer. Value Health, 2013. 16(2): p. 288-96.

139. Ramsey, S.D., et al., Quality of life in survivors of colorectal carcinoma. Cancer, 2000. 88(6): p. 1294-303.

140. Shiroiwa, T., T. Fukuda, and K. Shimozuma, Cost-effectiveness analysis of trastuzumab to treat HER2-positive advanced gastric cancer based on the randomised ToGA trial. Br J Cancer, 2011. 105(9): p. 1273-8.

141. Picozzi, V., et al., Health-Related Quality of Life in Patients with Metastatic Pancreatic Cancer. J Gastrointest Cancer, 2017. 48(1): p. 103-109.

142. Fisher, M. and M. Gore, Cost-effectiveness of trabectedin plus pegylated liposomal doxorubicin for the treatment of women with relapsed platinum-sensitive ovarian cancer

in the UK: analysis based on the final survival data of the OVA-301 trial. Value Health, 2013. 16(4): p. 507-16.

143. Lester-Coll, N.H., et al., Health State Utilities for Patients with Brain Metastases. Cureus, 2016. 8(7): p. e667.

144. Chu, J.N., et al., Cost-effectiveness of immune checkpoint inhibitors for microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer. Cancer, 2019. 125(2): p. 278-289.

VITA

Yilin Chen earned her Bachelor of Science in Global Health from Wuhan University in China. She received her Master of public Health from University of Washington. She also completed her Doctor of Philosophy in Health Economics and Outcomes Research (HEOR) from the Comparative Health Outcomes, Policy and Economics (CHOICE) Institute at the University of Washington. During her PhD, she completed her dissertation titled Evaluating Heterogeneity in Treatment Effects and Economic Value of Tumor-Agnostic Drugs. She has gained formal training in biostatistics, econometrics, epidemiology, economic evaluation, and health policy. Her research interests include real-world evidence for health technologies, the assessment of clinical and economic value of pharmaceuticals and diagnostics, evaluation of health care policies, and global health.