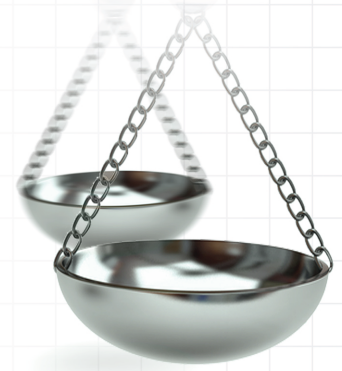




Comparative effectiveness of nivolumab versus clinical practice for advanced gastric or gastroesophageal junction cancer

Journal of **Comparative Effectiveness Research**

Ian Chau¹, Dieter Ayers^{*,2}, Sarah Goring², Shannon Cope², Beata Korytowsky³ & Pranav Abraham³

¹Royal Marsden Hospital, London & Surrey, SM2 5PT, UK

²Precision Xtract, Vancouver, B.C., V6H 3Y4, Canada

³Bristol-Myers Squibb, Princeton, NJ 08648, USA

*Author for correspondence: dieter.ayers@precisionxtract.com

Aim: To determine the effectiveness of nivolumab compared with routine clinical practice (RCP) for patients with gastric or gastroesophageal cancer refractory to, or intolerant of, two or more previous regimens, using real-world electronic patient records from a US population, a single-arm trial (CheckMate 032) and a randomized controlled trial in an Asian setting (ATTRACTION-2). **Materials & methods:** A simulated treatment comparison was conducted to predict overall survival for patients treated with nivolumab compared with RCP in the USA. **Results:** Results of the indirect simulated treatment comparison suggest that nivolumab is associated with a 50% reduction in the risk of all-cause mortality relative to RCP (Hazard ratio: 0.50; 95% CI: 0.36–0.68). **Conclusion:** The survival benefit of nivolumab may extend more generally to the USA.

First draft submitted: 27 September 2019; Accepted for publication: 27 November 2019; Published online: 24 December 2019

Keywords: gastric cancer • nivolumab • propensity scores • real-world evidence • simulated treatment comparison

Globally, gastric cancer is the third-leading cause of cancer-related mortality [1]. Prevalence varies by geography, with the highest prevalence reported in Eastern Asia [2]. Prognosis is generally poor, although there are marked differences in survival outcomes across geographies and ethnicities: 5-year overall survival has been estimated at 10–15% in the USA compared with 45–50% in Eastern Asia [3–5].

Nivolumab is one of several immunotherapies being investigated for the treatment of advanced gastric cancer. Among heavily pretreated (i.e., two or more prior lines of therapy) patients with locally advanced or metastatic gastric or gastroesophageal junction (GEJ) cancer in North America and Europe, the Phase I/II CheckMate 032 trial (NCT01928394) demonstrated that nivolumab monotherapy has antitumor activity with durable responses and a favorable tolerability profile [6]. Overall survival was measured in this study as an exploratory end point; however, the trial had no control arm, and therefore treatment effect estimates comparing nivolumab with routine clinical practice (RCP) were not assessed. Overall survival was the primary end point of the randomized, double-blind, placebo-controlled ATTRACTION-2 trial (NCT02267343) conducted among patients in Japan, South Korea and Taiwan, and who had failed two or more lines of therapy [7]. The results of this study suggest a survival benefit with nivolumab compared with placebo (hazard ratio [HR]: 0.63; 95% CI: 0.51–0.78). However, given the differences in overall survival between Asian and non-Asian populations, it is unclear whether the treatment effect observed in ATTRACTION-2 is generalizable to the US population.

In a previous study, the median overall survival observed in the CheckMate 032 trial was found to be longer than the overall survival observed in a frequency-matched cohort of patients with gastric or GEJ cancer treated in a real-world, US-based clinical setting [8]. Given that nonrandomized comparisons between clinical trials and real-world data sources are associated with considerable risks of bias, we sought to test the robustness of these findings by employing alternate methodological approaches.

Table 1. Summary of study methodology.

Comparison		Primary	Sensitivity		Exploratory
		CheckMate 032 (N = 42) vs Flatiron Random (N = 34)	CheckMate 032 (N = 42) vs Flatiron Prevalence (N = 36)	CheckMate 032 (N = 42) vs Flatiron Last line (N = 118)	ATTRACTION-2 Nivolumab (N = 330) vs ATTRACTION-2 Placebo (N = 163)
Primary analysis	STC: Nivolumab adjusted to Flatiron population versus RCP (PH models)	Main	Sensitivity	Sensitivity	NA
	STC: Nivolumab adjusted to Flatiron population versus RCP (FP models)	Main	Sensitivity	Sensitivity	NA
Sensitivity analysis	Crude: Unadjusted nivolumab versus RCP (PH models)	Sensitivity	Sensitivity	Sensitivity	NA
	Joint: Adjust for covariates simultaneously in CheckMate 032 and Flatiron cohort [†] (PH models)	Sensitivity	Sensitivity	Sensitivity	NA
	STC: Nivolumab versus RCP adjusted to CheckMate 032 population (PH models)	Sensitivity	Sensitivity	Sensitivity	NA
Exploratory analysis	Population-adjusted: Nivolumab adjusted to Flatiron population versus Placebo adjusted to Flatiron population (Random cohort) (PH models)	NA	NA	NA	Exploratory

[†] Assumes treatment effect size is not modified by target population.
 FP: Fractional polynomial; NA: Not applicable; PH: Proportional hazard; RCP: Routine clinical practice; STC: Simulated treatment comparison.

The primary objective of this study was to estimate the comparative effectiveness of nivolumab relative to RCP with respect to overall survival for the treatment of US-based patients with advanced or metastatic gastric or GEJ cancer refractory to, or intolerant of, at least two previous chemotherapy regimens.

Methodology

This study was conducted by Precision Xtract (Vancouver, Canada) and was funded by Bristol-Myers Squibb. It has been reported in accordance with the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) and STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statements [9,10].

Two different methodological approaches were used to achieve the study objective (Table 1). In the main approach, simulated treatment comparison (STC) methodology was used to estimate the comparative effectiveness between nivolumab, as evaluated in the CheckMate 032 trial, and RCP, as available in the Flatiron Health database, with respect to overall survival. In an STC, outcomes are modeled for a treatment of interest as a function of relevant patient-related factors developed from the individual patient data (IPD) of an index trial. This model is then used to predict outcomes that would be expected in another population [11]. The second approach was exploratory and involved an adjustment of the overall survival outcomes, as reported in the ATTRACTION-2 trial, by reweighting the trial's patient characteristics to match those of a US-based cohort.

To support variable selection in the development of statistical models, literature reviews were conducted to identify potential prognostic factors and treatment effect modifiers. Studies of any design, that aimed to identify prognostic factors, were eligible for the systematic literature review of prognostic factors, while a targeted review of effect modifiers was restricted to randomized controlled trials (RCTs).

Main approach: indirect comparisons using CheckMate 032 & the Flatiron Health database

Data sources & cohort definitions

CheckMate 032

We retrieved IPD for a subgroup of patients enrolled in the CheckMate 032 trial who received nivolumab as a third or later line of therapy for locally advanced or metastatic gastric or GEJ carcinoma [12].

Flatiron Health's electronic health record database

Flatiron Health's electronic health record database maintains anonymized retrospective data on cancer-related diagnoses and staging, laboratory data and medications for US patients who have had active cancer. Using eligibility criteria similar to that of the subgroup selected from CheckMate 032 [12], we selected patients from the Flatiron

Health database who were diagnosed between 1 January 2011 and 6 November 2017 and who received a third or later line of treatment with any RCP regimen.

Patients were eligible to enter the cohort at the time they started third, fourth, fifth or later lines of therapy, meaning each patient may have had multiple time points at which they were eligible to enter the cohort. In an attempt to mimic the CheckMate 032 trial, we initially constructed the cohort using a period-prevalence approach, allowing patients to enter the cohort based on the first eligible line of therapy initiated during the enrollment period of the CheckMate 032 trial (November 2013 to July 2015). However, the resulting cohort (henceforth called the ‘prevalence cohort’) included primarily patients entering at the start of third line, with later lines underrepresented. Therefore, the main analytic cohort was constructed by randomly selecting cohort entry time among eligible patients such that the distribution of lines of therapy at cohort entry matched the frequencies observed in the CheckMate 032 trial (the ‘random cohort’). A third cohort, the last line of therapy cohort, consisted of patients entering the cohort at the time of initiating their final line of treatment. In previous research, this approach has demonstrated a similar distribution of line-of-therapy at cohort entry to CheckMate 032; however, this was only considered for exploratory purposes as it uses knowledge of future events to determine cohort eligibility [13].

Analysis

Overall survival was measured as time-to-event from the time of treatment initiation to death. Observations were censored if a death had not occurred by the end of the study period, and, where available, at the last known date alive as captured in the Flatiron Health dataset.

Main analyses

An STC was conducted to compare the overall survival observed for RCP in the Flatiron Health cohorts with the overall survival predicted for these patients if they had been treated with nivolumab. Prognostic and treatment effect-modifying factors identified through the literature review were evaluated as both main effects and interactions in multiple competing models, where data were available from both the CheckMate 032 trial and the Flatiron Health database. The final models were selected based on data availability, clinical relevance and model fit statistics. Factors that led to substantial improvements in fit were considered more important than those that did not appreciably change the fit of the model.

The primary STC analysis specified a Cox proportional hazards model framework. However, standard Cox models typically do not estimate a baseline hazard as this is considered to be a nuisance parameter. As this value is necessary for analyses of relative efficacy, we used a regression that models covariates proportionally using a B-spline smoothing method for the underlying hazard [14]. The resulting prediction of effects of prognostic factors is similar to the proportional hazards model, but the underlying hazard function is used as a common reference when groups are compared. Secondary models were conducted using fractional polynomials to relax the assumption of proportionality and permit time-varying hazard ratios [15]. Full details of the model development and selection are presented in the Supplementary Materials.

Sensitivity analyses

A series of sensitivity analyses were conducted to test the assumptions and robustness of the main analysis (Table 1). First, crude hazard ratio estimates were generated to compare the observed survival data for nivolumab and RCP without adjustments for covariates. A second sensitivity analysis was conducted to extend the crude model through a joint model, in which a regression was conducted on the combined CheckMate 032 and Flatiron Health IPD datasets simultaneously. This approach assumes that the prognostic effect of each explored covariate is the same in both exposure groups. Interaction terms were explored to test that assumption. In a third sensitivity analysis, we reversed the direction of the STC and constructed models based on the outcomes observed in the Flatiron Health cohorts, with RCP as the index exposure, to predict survival outcomes for patients enrolled in CheckMate 032. A final set of sensitivity analyses consisted of reconducting all analyses using the prevalence and last-line cohorts derived from the Flatiron Health database.

Exploratory approach: population-adjusted analysis of the ATTRACTION-2 RCT

To evaluate the consistency of results observed in the ATTRACTION-2 trial relative to the effects estimated for the US context, propensity score weighting was used to account for differences in covariate distributions across the populations. In this case, we estimated the effect that would have been observed had the ATTRACTION-2 patients

Table 2. Distribution of baseline characteristics and prognostic factors in the CheckMate 032 and Flatiron Health database.

	Covariate	CheckMate 032 (n = 42)	Flatiron Health cohorts		
			Random line cohort (n = 34)	Prevalence cohort (n = 36)	Last-line cohort (n = 118)
Prognostic factors	Median age (range)	58.5 (29, 80)	64.5 (33, 79)	65.0 (34, 82)	65.0 (25, 84)
	Age >65 years	9 (21.4%)	17 (50.0%)	16 (44.4%)	54 (45.8%)
	ECOG 1	22 (52.4%)	23 (67.6%)	25 (69.4%)	87 (73.7%)
	Hemoglobin >10.5 g/dl	29 (69.0%)	20 (58.8%)	17 (47.2%)	77 (65.3%)
	Prior gastrectomy	7 (16.7%)	9 (26.5%)	6 (16.7%)	29 (24.6%)
	LDH >250 IU/l	25 (59.5%)	6/15 (40.0%)	9/17 (52.9%)	20/56 (35.7%)
	ALP >100 U/l	22 (52.4%)	28 (82.4%)	22 (61.1%)	74 (62.7%)
	GEJ	26 (61.9%)	14 (41.2%)	16 (44.4%)	59 (50.0%)
	Male	31 (73.8%)	27 (79.4%)	26 (72.2%)	85 (72.0%)
Stage	I & II	2 (4.8%)	3 (8.8%)	3 (8.3%)	15 (12.7%)
	III	11 (26.2%)	9 (26.5%)	7 (19.4%)	19 (16.1%)
	IV	28 (66.7%)	21 (61.8%)	25 (69.4%)	77 (65.3%)
	Unknown	1 (2.4%)	1 (2.9%)	1 (2.8%)	7 (5.9%)
Prior lines	2	18 (42.9%)	14 (41.2%)	34 (94.4%)	80 (67.8%)
	3	17 (40.5%)	12 (35.3%)	1 (2.8%)	27 (22.9%)
	≥4	7 (16.7%)	8 (23.5%)	1 (2.8%)	11 (9.3%)
Most commonly used agents in routine clinical practice [†]	Paclitaxel	–	13 (37.1%)	9 (25.0%)	36 (30.8%)
	Ramucirumab	–	12 (34.3%)	12 (33.3%)	33 (28.2%)
	Irinotecan	–	7 (20.0%)	7 (19.4%)	28 (23.9%)

[†] Agents may have been used alone, or in combination with others.

ECOG: Eastern Cooperative Oncology Group; GEJ: Gastroesophageal junction.

been observed in a US setting, which is analogous to the ‘average treatment effect in the controls,’ commonly found in the propensity score literature [16]. Consequently, weights were defined as:

$$w_i = \frac{Z_i (1 - e_i)}{e_i} + (1 - Z_i),$$

where Z_i represents group membership of patient i (and equals 1 if patients are from ATTRACTION-2, and 0 if they are from Flatiron Health), and e_i is the propensity score estimated with the logistic regression (i.e., the probability of being in ATTRACTION-2 conditional on the variables in the model). A proportional hazards model was conducted on the reweighted IPD, thereby adjusting the trial sample to reflect the patient characteristics, and hence the treatment effect that would have been observed had the trial been conducted among patients with characteristics similar to the US-based Flatiron Health cohort.

Software

Analyses, including covariate adjustments, were conducted in R (version 3.5.1; www.r-project.org) using the survival and bshazard packages [17,18]. Bayesian analyses were conducted with JAGS (version 4.2.0) [19].

Results

Prognostic factors & treatment effect modifiers

The systematic literature review of prognostic factors returned 22 studies of interest, from which nine prognostic factors were identified and available in the data from both CheckMate 032 and the Flatiron Health database (Table 2). The targeted literature review identified only two RCTs, which compared outcomes from immunotherapies; subgroup analyses did not reveal significant difference, as assessed by z-tests, though neither study was designed or powered to detect subgroup differences [20].

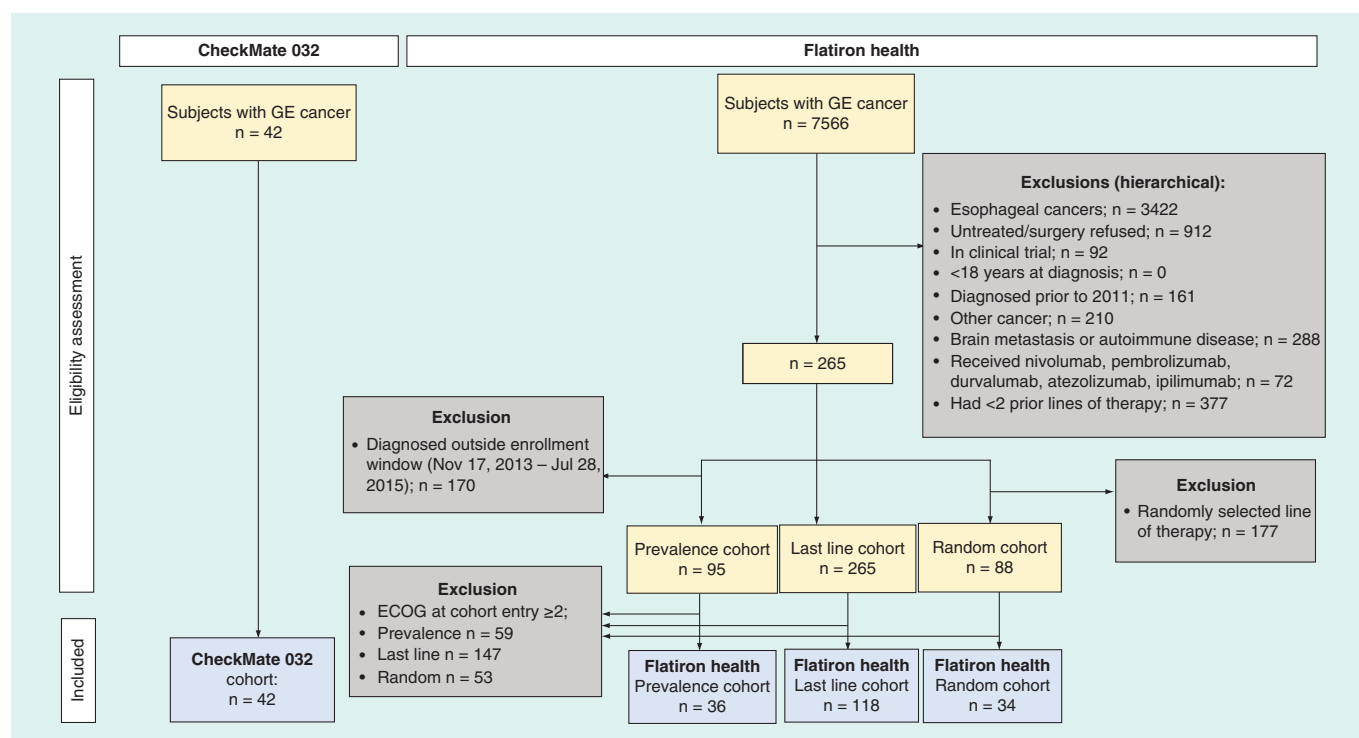


Figure 1. Selection of patient cohorts from the CheckMate 032 and the Flatiron Health database.
ECOG: Eastern Cooperative Oncology Group; GE: Gastroesophageal.

Main approach: indirect comparisons using CheckMate 032 & the Flatiron Health database

Patient characteristics

The CheckMate 032 cohort consisted of 42 patients with advanced gastric or GEJ cancer who had failed two or more previous lines of therapy. The three Flatiron Health cohorts, constructed from an initial set of 7566 patients, consisted of 34 patients in the random cohort, 36 patients in the prevalence cohort and 118 patients in the last line of therapy cohort (Figure 1). Across the trial and Flatiron Health cohorts, patient characteristics were broadly similar (Table 2). However, patients in the Flatiron Health cohorts tended to be older with worse performance status and lower hemoglobin levels. As no standardized definition of care exists in the USA for this patient population, there was considerable variability in regimen experience, with 22 unique regimens recorded.

Unadjusted outcomes

During the observation period, the median overall survival was 8.97 months (95% CI: 5.03–15.51) in the CheckMate 032 cohort, compared with 4.69 (95% CI: 3.21–8.13), 4.66 (95% CI: 3.61–10.30) and 3.84 months (95% CI: 3.08–4.69), in the random, prevalence and last line of therapy Flatiron Health cohorts, respectively.

STCs: main analysis

The final overall survival regression model constructed based on the patient data from CheckMate 032 included six covariates (Table 3). Based on coefficient size, the most important prognostic factors were Eastern Cooperative Oncology Group (ECOG) performance score, hemoglobin level and whether the patient had a prior gastrectomy.

The predicted overall survival for nivolumab, had it been used by patients in the Flatiron Health random cohort, was similar to the actual observed survival for patients treated with nivolumab in the CheckMate 032 trial (Figure 2).

The primary STC analysis compared the smoothed log-hazards from the predicted nivolumab arm to the smoothed log-hazards from the observed RCP arm to estimate a statistically significant improvement in overall survival (HR: 0.50; 95% CI: 0.36–0.68). The first-order Weibull fractional polynomial model predicted outcomes that were broadly consistent with the proportional hazards model beyond 3 months of observation, though no difference was predicted in the first 3 months of modeling (Figure 3).

Table 3. Exponentiated regression coefficients from the simulated treatment comparison in the CheckMate 032 population.

Prognostic factor	Reference level	HR (95% CI)	
		Multivariate model	Univariate model
ECOG PS 1	PS 0	1.34 (0.46–3.87)	1.36 (0.72–2.59)
Hemoglobin >10.5 g/dl	≤10.5 g/dl	0.61 (0.16–2.27)	0.73 (0.29–1.82)
LDH >250 IU/l	≤250 IU/l	0.99 (0.27–3.65)	0.99 (0.54–1.83)
ALP >100 U/l	≤100 U/l	0.92 (0.27–3.13)	0.91 (0.47–1.77)
Prior gastrectomy	No prior gastrectomy	0.66 (0.1–4.59)	0.73 (0.18–2.94)
Location (GEJ)	Gastric	0.82 (0.33–2.02)	0.99 (0.47–2.09)

ECOG: Eastern Cooperative Oncology Group; GEJ: Gastroesophageal junction; HR: Hazard ratio; PS: Performance score.

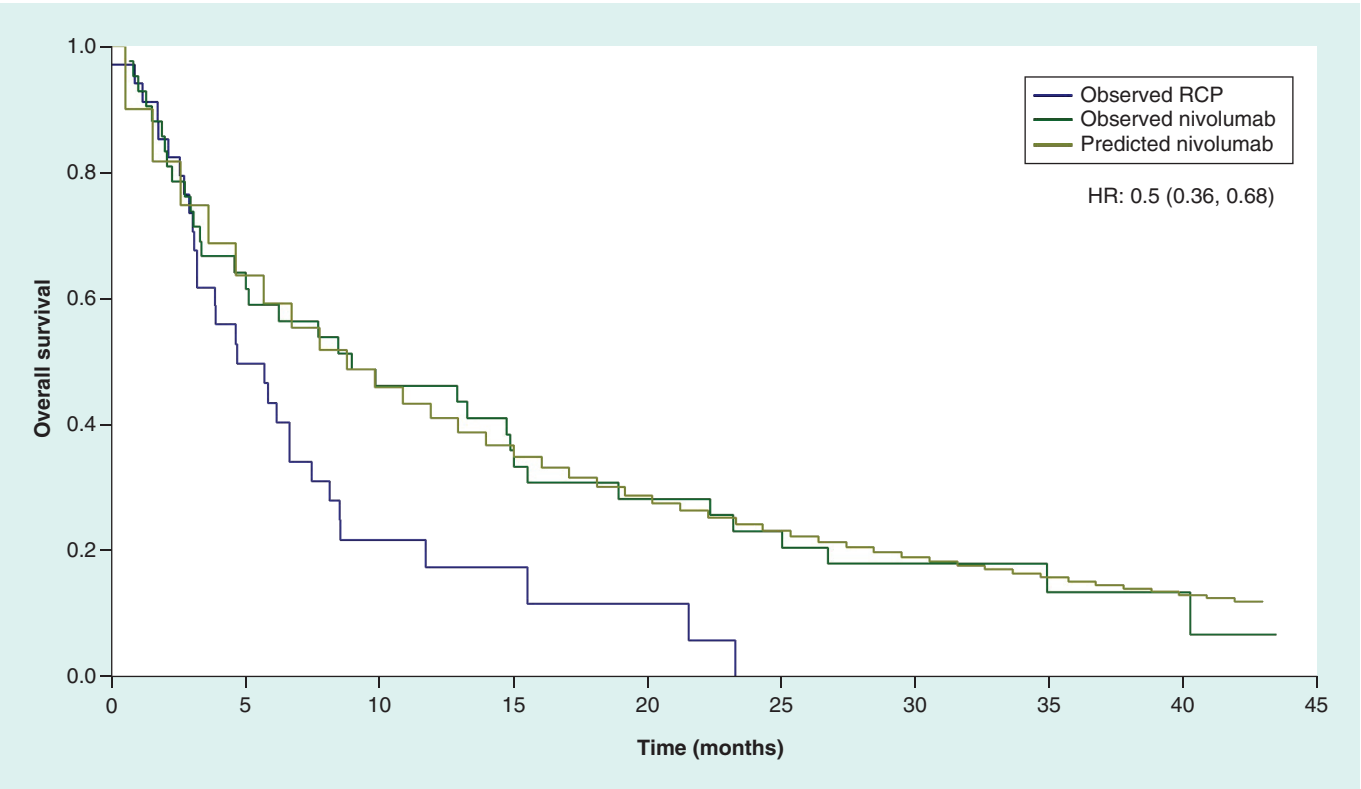


Figure 2. Cox proportional hazards model for observed overall survival with routine clinical practice (Flatiron Health random cohort) compared with observed (CheckMate 032 cohort) and predicted nivolumab (Flatiron Health random cohort). RCP: Routine clinical practice.

STCs: sensitivity analyses

The direction of association estimated in the primary analysis was consistent across modeling approaches (Table 4). Crude comparisons generated the most favorable results for nivolumab, while every other modeling approach, which adjusted data to correct for population differences, brought the hazard ratio closer to the null value of one. Predicting the RCP arm, based on the random cohort, generated a hazard ratio close to one for nivolumab compared with RCP (HR: 0.93; 95% CI: 0.67–1.30). In this case, the predicted RCP arm demonstrated better survival in the initial period of follow-up. However, as the two survival curves modeled in this scenario crossed, the assumption of proportional hazards was invalidated. As the majority of 95% credible (Bayesian framework) or confidence (frequentist framework) intervals did not span the null value, the survival benefit associated with nivolumab compared with RCP was considered statistically significant. The analyses based on fractional polynomials

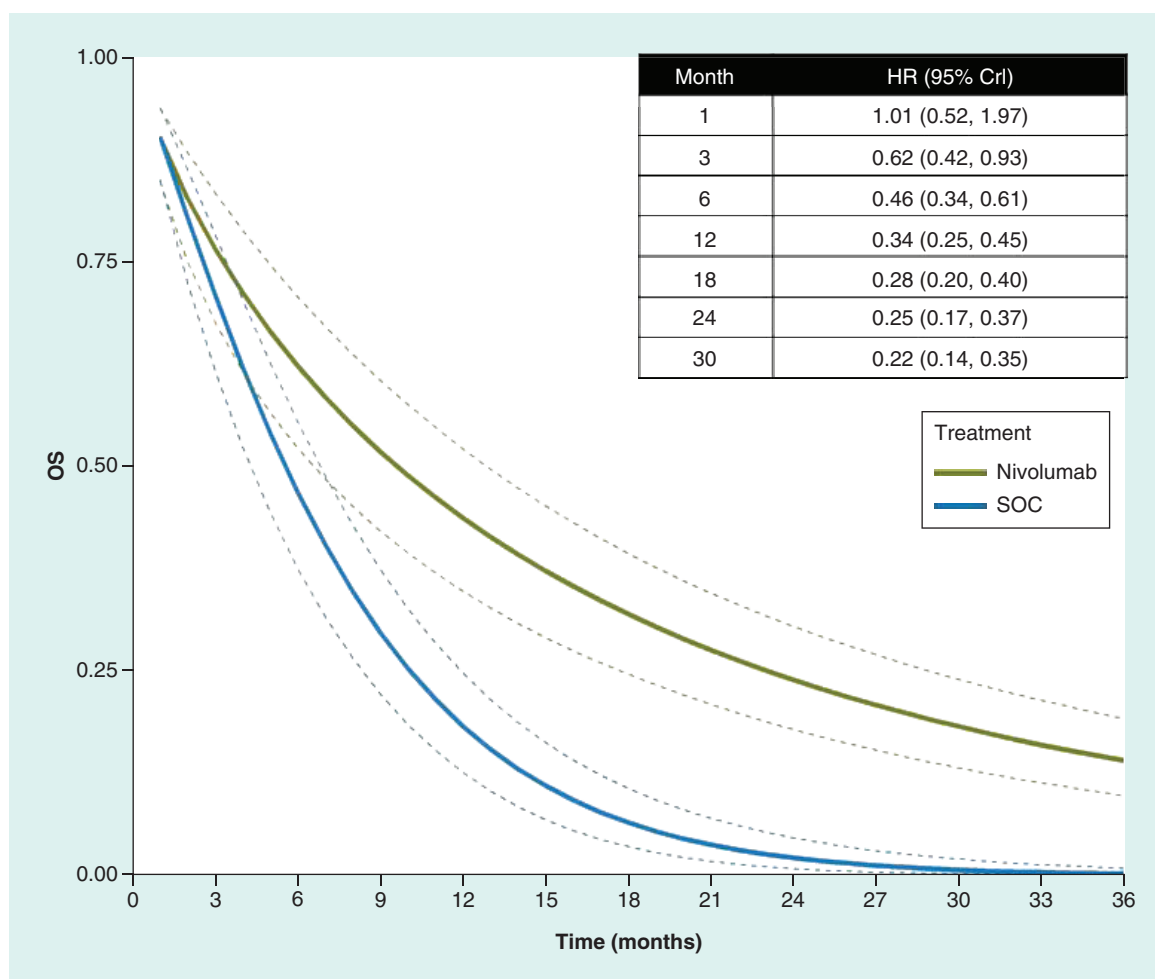


Figure 3. Weibull survival model of overall survival outcomes for predicted nivolumab and routine clinical practice based on the patient characteristics of the Flatiron Health random cohort.
CrI: Credible interval; HR: Hazard ratio; OS: Overall survival.

Table 4. Hazard ratios of nivolumab relative to routine clinical practice estimated through sensitivity analyses.

Cohort	Model	HR (95% CrI/CI)
Prevalence	Crude	0.56 (0.45–0.69)
	Joint (Frequentist)	0.66 (0.38–1.14)
	Predicted nivolumab arm	0.65 (0.52–0.82)
	Predicted routine clinical practice arm	0.59 (0.47–0.73)
Random	Crude	0.46 (0.36–0.59)
	Joint (Frequentist)	0.56 (0.32–0.99)
	Predicted nivolumab arm	0.50 (0.36–0.68)
	Predicted routine clinical practice arm	0.93 (0.67–1.30)
Last line	Crude	0.50 (0.41–0.62)
	Joint (Frequentist)	0.52 (0.34–0.80)
	Predicted nivolumab arm	0.61 (0.46–0.80)
	Predicted routine clinical practice arm	0.54 (0.43–0.67)

CrI: Credible interval; HR: Hazard ratio.

Table 5. Distribution of patient characteristics in the ATTRACTION-2 trial reweighted by propensity scores.

Variable	Flatiron Health random cohort N = 34	ATTRACTION-2			
		Observed N = 493	Overall N = 493	Propensity weighted	
				Nivolumab arm N = 330	Placebo arm N = 163
ECOG (% PS 1)	68.57	70.99	67.1	68.61	63.79
Hemoglobin (% >10.5 g/dl)	57.14	55.17	59.17	60.5	56.25
ALP (% >100 U/l)	82.86	72.01	84.07	81.67	89.33
Tumor location (% GEJ)	42.86	9.35	40.63	41.9	37.83
Age (% >65 years)	51.43	37.53	47.46	47.01	48.46
Sex (% male)	77.14	70.59	78.14	75.71	83.45

ECOG: Eastern Cooperative Oncology Group; GEJ: Gastroesophageal junction; PS: Performance score.

for the alternate cohorts, which allowed for time-varying hazards, generally supported the finding that nivolumab is favored to RCP.

Additional results from sensitivity analyses, including analyses based on the other Flatiron Health cohorts, are presented in the Supplementary Materials.

Exploratory approach: population-adjusted analysis of ATTRACTION-2 data

The propensity score reweighting of the ATTRACTION-2 patients matched patient characteristics to those of the US population (Table 5). An important difference was observed between the initial percentages of patients with the tumor located in the GEJ, which was reweighted to 40.63% in the overall population. Compared with the unadjusted hazard ratio reported in ATTRACTION-2 of 0.63 (95% CI: 0.51–0.78), the analysis based on this reweighted patient sample returned an estimated hazard ratio of 0.57 (95% CI: 0.26–1.23).

Discussion

Using indirect and population reweighting approaches to estimate treatment effect in a US population, our study showed a beneficial effect of nivolumab over RCP for the treatment of individuals with gastric or GEJ cancer refractory to or intolerant of two or more previous chemotherapy regimens. The relative effect size from our main STC model, HR: 0.50 (95% CI: 0.36–0.68), was of similar magnitude to the reweighted analysis of the ATTRACTION-2 RCT (HR: 0.57; 95% CI: 0.26–1.23) and was robust to alternate modeling assumptions, as explored in the sensitivity analyses.

There are limited effective treatment options for patients with advanced gastric or GEJ cancer who have failed multiple lines of therapy. Until recently, there were no US FDA-approved or NCCN-recommended therapies for third or later lines of therapy [21]. Pembrolizumab is the only FDA-approved therapy, having been granted accelerated approval in late 2017 based on a favorable benefit–risk profile. The FDA considered that the modest yet durable response observed among a subset of PD-L1-expressing patients in the single-arm KeyNote 059 trial was predictive of a meaningful benefit to patients in the form of improved survival [22,23]. The findings of our study demonstrating that nivolumab is associated with a doubling in median overall survival relative to a US-based cohort of patients receiving RCP provides further support for the use of PD-L1 checkpoint inhibitors in heavily pretreated patients with advanced gastric or GEJ cancer. Although safety and health-related quality-of-life were not in the scope of this research, previous interpretation of the ATTRACTION-2 RCT suggests that the safety profile of nivolumab is manageable and similar to that reported in patients with other advanced solid tumors [24,25]. In that RCT, treatment-related adverse events occurred with a frequency that was similar to placebo. If this relationship holds across populations, then the benefit of nivolumab could be obtained without the additional cost of adverse events.

The beneficial effect of nivolumab over RCP observed in the Asia-based ATTRACTION-2 RCT has been used to support regulatory licensing in Japan, where nivolumab is indicated for the treatment of unresectable advanced or recurrent gastric cancer, which has progressed after chemotherapy. Although our reweighted analysis could not adjust for all immune and inflammation signatures that may differ between Asian and non-Asian populations, this study provides insights into the relative effect size that might have been observed had the RCT been conducted in a non-Asian setting because of the adjustments for other key prognostic and potentially effect modifying differences

between the populations such as age and tumor location [5,7]. The trend toward a more protective effect in the US population compared with the Asian setting aligns with subgroup analyses from KeyNote 061 and ATTRACTION-2, which demonstrated trends toward larger treatment effect for immunotherapy among patients with GEJ tumors, which tend to be more common in the USA than in Asia [7,26].

The use of real-world evidence confers a unique advantage in that the dataset can be specified to any subset of gastrointestinal cancer patient. Because of this, we were able to generate a cohort representative of the CheckMate 032 trial to facilitate our STC approach. Our approach serves as a framework for estimating the comparative effectiveness between treatments by merging trial data and real-world evidence. Data source and cohort definition are important aspects of any pharmacoepidemiological study design and constitute a key element in the FDA's recently released framework for evaluating real world evidence [27]. The Flatiron Health database was selected for this study as it is comprehensive and research-oriented, representing over 200 cancer practices and practice groups providing care for over 750,000 patients with cancer in the USA. Several approaches to deriving patient cohorts were considered and three were evaluated for the purposes of this study to explore potential threats to internal validity due to selection bias. While analyses based on the three cohorts evaluated in this study generated consistent conclusions, there were some variations in the patient characteristics which may have impacted survival outcome estimates and the generalizability of findings. The main analysis was based on the random cohort, which selected patients weighted to match the frequencies of lines of therapies observed in the CheckMate 032 trial. The alternative cohorts, evaluated in sensitivity analyses, were considered to carry some bias. In the prevalence cohort, patients were more likely to have been initiating a third line of therapy at cohort entry, compared with the line of therapy at the baseline of the CheckMate 032 trial. This may have distorted comparability with the patients enrolled in CheckMate 032 toward longer survival times in the RCP cohort and, consequently, estimated a smaller magnitude of treatment effect. Conversely, selection of patients for the last-line cohort was based on knowledge that patients received no further lines of therapy. This may have biased the cohort toward a shorter time-to-death and estimated a larger magnitude of treatment effect. It is essential that real-world evidence-based analyses are thoroughly vetted through sensitivity analyses and that potential limitations and the direction and magnitude of biases are explicitly evaluated.

There are several strengths to this study. A comprehensive and systematic approach was taken to identify prognostic factors, coupled with a targeted review to identify potential treatment effect modifiers, thereby leveraging the literature to better inform our analyses. Given the novelty of the methods used in our study and the assumptions that were made, such as selection of cohorts from real-world evidence, we conducted a series of alternate and sensitivity analyses. Conclusions were broadly consistent irrespective of the cohort definition and supported the findings of our primary analyses. Moreover, our study was intended to complement the findings of studies that have used other methodological approaches, such as a recent study conducted using a frequency-matched approach, as well as the adjusted Cox model and population-adjusted analysis of ATTRACTION-2 presented herein [13]. Again, conclusions were consistent across methodological approaches, and all demonstrated a similar trend toward longer overall survival associated with nivolumab. However, the methodological literature for STCs is sparse, particularly in the context of time-to-event survival data [28,29]. Therefore, our methodological approach featured several new innovations and multiple different analyses to explore the consistency of our findings. First, we used smooth splines in the STC methodology, which led to more stable estimations that were less affected by variability in hazards over time, especially given small sample sizes. Second, we applied both proportional hazard and time-varying hazard approaches, and used both frequentist and Bayesian frameworks. Third, we estimated relative effect size using both the Flatiron and the CheckMate 032 patients as target populations.

Despite the rigorous approach taken in model development and the evaluations of robustness, there are additional limitations to consider. First, the comparisons made in our study were outside of a randomized trial and, thus subject to bias. For example, while we took a rigorous approach to identify prognostic factors and effect modifiers, we were limited by the quality of the evidence identified. Second, the STC represented an 'unanchored' comparison, for which the major assumption is that all effect modifiers or prognostic variables are accounted for. We were only able to account for measured factors and could not account for unmeasured covariates or capture the complex interactions that may exist. Therein lies the major assumption of the unanchored STC that all effect modifiers or prognostic variables are accounted for. That assumption could not be formally tested; however, given the large treatment effect size and the degree of prognostic significance of other variables identified in our systematic review but not included in our analysis, we expect that imbalances in prognostic factors between the nivolumab and RCP cohorts would not be sufficiently large to reverse the direction of association [30,31]. However, this approach was

necessary as there were no common comparators. While the RCP comparator group included patients who received active treatment, heavily pretreated patients may decline further active therapies and may instead opt for palliative care. While these patients were not included in the comparator group of our STC, their inclusion may have led to even larger estimated treatment effect sizes. Finally, while the evidence applied to our models was robust, analyses were ultimately based on small-sample sizes.

Conclusion

Despite the absence of head-to-head data and the inherent limitations of using nonrandomized comparisons of clinical trial data and real-world evidence, this study suggests that using nivolumab for third or later lines of therapy may confer an overall survival benefit relative to RCP for patients with advanced gastric or GEJ cancer in the US setting. The methodological approach taken in this study serves as a framework for estimating the comparative effectiveness between treatments using single-arm trial data and real-world evidence.

Summary points

- Gastric cancer is the third-leading cause of cancer-related mortality worldwide and is particularly common in Eastern Asia.
- There are marked differences in survival outcomes by geography and ethnicity, with 5-year overall survival varying from 10 to 15% in the USA compared with 45 to 50% in Eastern Asian.
- The recent ATTRACTION-2 randomized controlled trial reported a survival benefit for patients treated with nivolumab over routine clinical practice (RCP) in an Asian setting, yet it is unclear whether these findings generalize to the US setting.
- In the absence of a head-to-head trial in the US setting, three cohorts of patients were selected from Flatiron Health's electronic health record retrospective, real-world evidence database to match the characteristics of patients who were enrolled in CheckMate 032 and treated with nivolumab monotherapy.
- A simulated treatment comparison, a regression model that uses data for an index intervention and then models the outcome as a function of patient-related factors, was conducted to compare the overall survival observed for patients treated in RCP in the Flatiron Health cohorts with the predicted overall survival that would have been expected had these patients been treated with nivolumab.
- In the primary analysis, an improvement in overall survival was predicted with nivolumab compared with RCP (HR: 0.50; 95% CI: 0.36–0.68) and this direction of effect was consistent across sensitivity analyses.
- Consistency was evaluated and confirmed by using propensity scores to reweight the individual patient data from the ATTRACTION-2 trial to reflect the patient characteristics, and hence the treatment effect, of the US population, suggesting minimal or no effect modification associated with geographic region.
- Despite the absence of head-to-head data and the inherent limitations of using nonrandomized comparisons of clinical trial data and real-world evidence, this study suggests that nivolumab, used as a third or greater line of therapy, may confer an overall survival benefit compared with RCP in patients with advanced or metastatic gastric or gastroesophageal junction cancer in the US setting.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/ce-2019-0145

Author contributions

Each author listed above participated sufficiently in the work to take responsibility for the content, and that all those who qualify are listed. Each author provided substantial contributions to the conception or design of the work; or the acquisition, analysis or interpretation of data for the work; and drafting the work or revising it critically for important intellectual content; and final approval of the version to be published; and agreement to be accountable for all aspects of the work are appropriately investigated and resolved.

Financial & competing interests disclosure

The authors acknowledge National Health Service funding to the National Institute for Health Research Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and The Institute of Cancer Research. D Ayers, S Goring and S Cope are employees of or subcontractors to Precision Xtract; Precision Xtract was paid by Bristol-Myers Squibb for conducting this work. B Korytowsky and P Abraham are employees of Bristol-Myers Squibb. The authors have no other relevant affiliations or financial involvement

with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Medical writing assistance was provided by Michael Zoratti.

Ethical conduct of research

The authors state that they have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations.

Data sharing statement

The authors certify that this manuscript reports the secondary analysis of clinical trial data that have been shared with them, and that the use of this shared data is in accordance with the terms agreed upon their receipt. The source of this data is: Bristol-Myers Squibb and Flatiron Health.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

- Qi X, Liu Y, Wang W. Management of advanced gastric cancer: an overview of major findings from meta-analysis. *Oncotarget* 7(47), 78180–78205 (2016).
- Qiu M-Z, Cai M-Y, Zhang D-S *et al.* Clinicopathological characteristics and prognostic analysis of Lauren classification in gastric adenocarcinoma in China. *J. Transl. Med.* 11(1), 58 (2013).
- Gill S, Shah A, Le N, Cook EF, Yoshida EM. Asian ethnicity-related differences in gastric cancer presentation and outcome among patients treated at a canadian cancer center. *J. Clin. Oncol.* 21(11), 2070–2076 (2003).
- Hsu C, Shen Y-C, Cheng C-C, Cheng A-L, Hu F-C, Yeh K-H. Geographic difference in safety and efficacy of systemic chemotherapy for advanced gastric or gastroesophageal carcinoma: a meta-analysis and meta-regression. *Gastric Cancer* 15(3), 265–280 (2012).
- Lin SJ, Gagnon-Bartsch JA, Tan IB *et al.* Signatures of tumour immunity distinguish Asian and non-Asian gastric adenocarcinomas. *Gut* 64(11), 1721–1731 (2015).
- Janjigian YY, Bendell J, Calvo E *et al.* CheckMate-032 study: efficacy and safety of nivolumab and nivolumab plus ipilimumab in patients with metastatic esophagogastric cancer. *J. Clin. Oncol.* 36(28), 2836–2844 (2018).
- Kang YK, Boku N, Satoh T *et al.* Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, Phase III trial. *Lancet* 390(10111), 2461–2471 (2017).
- **Summarizes the results of a randomized controlled trial of nivolumab in patients with gastric cancer.**
- Janjigian YY, Le DT, Ott PA *et al.* Survival of chemotherapy (chemo) refractory gastric or gastroesophageal junction cancer (GC/GEJC) patients from Flatiron Health (FH): matched clinical characteristics to ATTRACTION-2 and CHECKMATE-032. *J. Clin. Oncol.* 36(Suppl. 4), 51–51 (2018).
- Benchimol EI, Smeeth L, Guttmann A *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* 12(10), e1001885 (2015).
- Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 370(9596), 1453–1457 (2007).
- Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons in submission to NICE (2016). <http://nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf>
- Ott P, Calvo E, Sharma P *et al.* Nivolumab monotherapy in patients with advanced gastric or gastroesophageal junction (GEJ) cancer and 2 or more prior treatment regimens: sub-analysis of the CheckMate 032 study. *Ann. Oncol.* 28(Suppl. 3), mdx262.006-mdx262.006 (2017).
- Janjigian YY, Le DT, Alexander P *et al.* Survival of chemotherapy (chemo) refractory gastric or gastroesophageal junction cancer (GC/GEJC) patients from Flatiron Health (FH): matched clinical characteristics to ATTRACTION-2 and CHECKMATE-032. Presented at: *Gastrointestinal Cancers Symposium*. CA, USA 18–20 Jan (2018).
- Rebola P, Salim A, Reilly M. bshazard: a flexible tool for nonparametric smoothing of the hazard function. *R Journal* 6(2), 114–122 (2014).
- Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med. Res. Methodol.* 11, 61 (2011).

- 16 Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.* 46(3), 399–424 (2011).
- 17 Rebora P, Salim A, Reilly M. bshazard: nonparametric smoothing of the hazard function. R package version 1.1 (2018). <https://CRAN.R-project.org/package=bshazard>
- 18 Therneau T. A package for survival analysis in S (2015). <https://CRAN.R-project.org/package=survival>
- 19 Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling (2003). <https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>
- 20 Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 326(7382), 219 (2003).
- 21 NCCN. Guidelines V1.2018 gastric cancer (2018). https://www.nccn.org/professionals/physician_gls/default.aspx
- 22 Fuchs CS, Doi T, Jang RW *et al.* Safety and efficacy of pembrolizumab monotherapy in patients with previously treated advanced gastric and gastroesophageal junction cancer: Phase II Clinical KEYNOTE-059 Trial. *JAMA Oncol.* 4(5), e180013 DOI:10.1001/jamaoncol.2018.0013 (2018).
- 23 FDA. Summary review; application number 125514Orig1s024 (2017). https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/125514Orig1s024Appov.pdf
- 24 Chau I. Checkpoint inhibition: an ATTRACTION in advanced gastric cancer? *Lancet* 390(10111), 2418–2419 (2017).
- 25 Kang Y-K, Boku N, Satoh T *et al.* Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, Phase III trial. *Lancet* 390(10111), 2461–2471 (2017).
- 26 Shitara K, Özgüroğlu M, Bang Y-J *et al.* Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, Phase III trial. *Lancet* 392(10142), 123–133 (2018).
- 27 FDA. Framework for FDA's real-world evidence program (2018). <https://www.fda.gov/media/120060/download>
- **Overview of the US FDA's views on the use of real-world data and evidence.**
- 28 Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics* 33(6), 537–549 (2015).
- **Overview of techniques for indirect treatment comparisons.**
- 29 Faria R, Hernandez Alava M, Manca A, Wailoo A. NICE DSU technical support document 17: the use of observational data to inform estimates of treatment effectiveness in technology appraisal: methods for comparative individual patient data (2015). <http://nicesdsu.org.uk/technical-support-documents/observational-data-tsd/>
- 30 Choi J. Prognostic factor for third-line chemotherapy in patients with advanced gastric cancer. *J. Clin. Oncol.* 35(Suppl. 15), e15702–e15702 (2017).
- **Examination of prognostic factors in the population of gastric cancer patients.**
- 31 Shim HJ, Yun JY, Hwang JE, Bae WK, Cho SH, Chung IJ. Prognostic factor analysis of third-line chemotherapy in patients with advanced gastric cancer. *Gastric Cancer* 14(3), 249–256 (2011).