The Oncologist®

Health Outcomes and Economics of Cancer Care

Median Survival or Mean Survival: Which Measure Is the Most Appropriate for Patients, Physicians, and Policymakers?

Omer Ben-Aharon,[a] Racheli Magnezi,[a] Moshe Leshno,[b] Daniel A. Goldstein[c,d,e]

[a]Department of Management, Health System Management Program, Bar Ilan University, Ramat Gan, Israel; [b]Coller School of Management, Tel Aviv University, Tel Aviv, Israel; [c]Davidoff Cancer Center, Rabin Medical Center, Petah Tiqva, Israel; [d]Department of Oncology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [e]Department of Health Policy and Management, University of North Carolina, Chapel Hill, North Carolina, USA

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Immunotherapies • Median survival • Mean survival • Weibull

Abstract

Introduction. Understanding the efficacy of treatments is crucial for patients, physicians, and policymakers. Median survival, the most common measure used in the outcome reporting of oncology clinical trials, is easy to understand; however, it describes only a single time point. The interpretation of the hazard ratio is difficult, and its underlying statistical assumptions are not always met. The objective of this study was to evaluate alternative measures based on the mean benefit of novel oncology treatments.

Materials and Methods. We reviewed all U.S. Food and Drug Administration (FDA) approvals for oncology agents between 2013 and 2017. We digitized survival curves as reported in the clinical trials used for the FDA approvals and implemented statistical transformations to calculate for each trial the restricted mean survival time (RMST), as well as the mean survival using Weibull distribution. We compared the mean survival with the median survival benefit in each clinical trial.

Results. The FDA approved 83 solid tumor indications for oncology agents between 2013 and 2017, of which 27 approvals based on response rates, whereas 49 approvals were based on survival endpoints (progression-free survival and overall survival). The average improvement in median overall survival or progression-free survival was 4.6 months versus 3.6 months improvement in the average RMST and 6.1 months improvement in mean survival using Weibull distribution.

Conclusion. Mean survival may supply valuable information for different stakeholders. Its inclusion should be considered in the reporting of prospective clinical trials. ***The Oncologist*** 2019;24:1469–1478

Implications for Practice: Mean survival may supply valuable information for different stakeholders. Its inclusion should be considered in the reporting of clinical trials.

Introduction

Understanding the efficacy of oncology treatments is crucial for patients, physicians, and policymakers. The most traditional measure is the hazard ratio (HR); however, the underlying assumptions regarding proportionality of the hazards (PH) are not met in all clinical trials. Additionally, its actual meaning is difficult for patients, physicians, and policymakers to understand. It is often difficult to appreciate the magnitude of a treatment effect using the HR. Therefore, the improvement in median survival is often used to describe the magnitude of clinical benefit provided by a new treatment. As the median survival time is insensitive to outliers, it is expected to be much shorter than the mean survival time in the presence of many long-term survivors, who will skew the mean survival time distribution. However, although the median is easy to understand, it describes only the outcome at a single time point. An important policy paper from the American Society of Clinical Oncology discussed the importance of raising the bar for clinically meaningful outcomes of clinical trials. The recommendations incorporated the need for higher median survival benefits and HR [1]. There is a current lack of a tool that is both clear and scientifically valid to understand the clinical benefit of a new intervention.

In survival analyses, the HR is the ratio of the hazard rates corresponding to the conditions described by two levels of an explanatory variable. For example, in a drug study, the treated population may die at half the rate per unit time as the control population. The HR would be 0.5, indicating lower hazard of death with the treatment. However, in an era of value-based cancer care, it is important to better understand the

Correspondence: Daniel A. Goldstein, M.D., Department of Oncology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. Telephone: 972-58-795-0421; e-mail: danielagoldstein@gmail.com Received March 4, 2019; accepted for publication May 20, 2019; published Online First on July 18, 2019. http://dx.doi.org/10.1634/theoncologist.2019-0175

magnitude of benefit of an intervention. For example, using the HR does not clearly describe how much longer these patients will live for. The HR can be used as a valid measure only if the PH assumption is met. Although in practice this assumption is not necessarily kept, the HR value is being used as a major criteria for approvals of novel oncology treatments.

The challenges in outcome reporting can be demonstrated using the following example: in the clinical trials used for the approval of two agents by the U.S. Food and Drug Administration (FDA) in 2015, similar HRs were reached—0.71 in the case of trametinib and dabrafenib for melanoma [2] and 0.67 in the case of liposomal irinotecan for pancreas cancer [3]. Both trials used overall survival (OS) as their primary endpoint. However, although the improvement in median survival of patients treated with trametinib and dabrafenib (25.1 months) versus the control arm (18.7 months) was 6.4 months, the improvement in median survival of irinotecan (6.1 months) versus the control arm (4.2 months) was more moderate, at 1.9 months.

Kaplan-Meier curves provide survival probability information throughout the study follow-up for a group of patients. Visually, the more space there is between the curves, the more effective the treatment is. Therefore, the area between the curves within a specific time window is a reasonable summary to quantify the survival benefit. This alternative measure is the restricted mean survival time (RMST), where the measurement is restricted only to the follow-up period of the clinical study [4].

The HR and RMST are complementary techniques that provide alternative methods of summarizing the same information [5]. In a comparative clinical study with progression-free survival (PFS) or OS as the endpoint, the HR is routinely used to report the survival results of the study. The clinical interpretation of the HR may not be straightforward, whereas a quantitative method that is based on the RMST may be used as a primary tool to better understand the clinical interpretation of the HR [4].

Although the median improvement in survival is easy to understand, it does not capture the long-term survival profile well, especially in cases in which a minority of patients have the potential to gain a durable survival. Simple calculation of RMST alone may underestimate the treatment benefit. In these cases, it may be appropriate to apply an extrapolation beyond the follow-up period of the clinical trial. In this study, we demonstrate the implementation of mean calculation using a suggested statistical distribution. This extrapolation may be relevant for modern immuno-oncology agents. A good example of the need for such an extrapolation is the pivotal clinical trial analyzing the safety and efficacy of ipilimumab for metastatic melanoma. This trial demonstrated that after 5 years of follow-up, 16% of patients were alive, compared with 8% of patients who received cytotoxic chemotherapy [6].

The primary objective of our study was to demonstrate the use of RMST and mean survival in trial reporting, in an effort to provide evidence to use such metrics in the reporting of future clinical trials. The secondary objective was to analyze all FDA drug approvals between 2013 and 2017 to understand more accurately the clinical benefit of these drugs.

Materials and Methods

Research Population

We reviewed all FDA approvals for oncology agents between 2013 and 2017. The information was extracted from the FDA Hematology/Oncology (Cancer) Approvals Notifications (http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm [7]. We included drugs used in the metastatic setting of treatment of solid tumors. We collected data regarding study endpoints, specifically the survival parameters: HR and median OS or PFS in each of the clinical trial arms.

We cross-checked the clinical outcomes in FDA approvals with the results reported in the respective registration studies; we compared the reported outcome in each published study with endpoints used for approval of the drug by the FDA. As raw data of the outcomes of clinical trials are not available, we used the GetData Graph Digitizer tool to extract the data points from the OS or PFS plots, and these data points were then used to fit parametric survival models.

Measure Calculations

The overall mortality or progression-free rate for each trial arm was derived from the corresponding OS or PFS curve from the clinical trial. We used RMST and Weibull distribution as parametric survival models:

RMST—the difference in RMST is the area between the curves within a specific time window [4]. We calculated this measure based on the follow-up period of the clinical trials used for the FDA approvals.

Weibull distribution—this distribution is a continuous probability distribution named after the Swedish mathematician Waloddi Weibull. The probability density function of a Weibull random variable is

$$f(t \mid a,b) = \frac{b}{a}\left(\frac{t}{a}\right)^{b-1} e^{-1(t/a)^b}$$

where $b > 0$ is the shape parameter and $a > 0$ is the scale parameter of the distribution. Its complementary cumulative distribution function is a stretched exponential function. The Weibull distribution is related to a number of other probability distributions; in particular, it interpolates between the exponential distribution ($b = 1$) and the Rayleigh distribution ($b = 2$ and $a = \sqrt{2\sigma}$). Given the parameters of Weibull distribution, the mean of the distribution is given by

$$a\Gamma\left(1+\frac{1}{b}\right)$$

where the gamma function is defined by

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx.$$

We considered the use of additional parametric survival models, such as log-normal and log-logistic distributions. However, as outcomes of the statistical processing were weak (correlation of digitized with original data was low), we omitted them from our analysis.

Curve Digitization

We incorporated the Nelder-Mead algorithm for fitting of survival curves [8, 9]: nonlinear optimization problems for which derivatives may not be known. In this situation, the defined problem was to create the Weibull distribution based on survival curve data. We evaluated the estimation by the "evaluate" function of the "distribution fitting tool" of MATLAB (detailed MATLAB codes are enclosed in supplemental online Appendix 1).

Model Appropriateness

To assess the fitness and appropriateness of Weibull as a parametric model, for each clinical trial, we calculated the median survival measures in each arm and compared them with the values reported in the original studies.

In addition, we calculated the HR based on the digitized curves. We improved the digitization in an iterative process so that the average deviation between the calculated HR and the HR as reported in the original studies would be minimized to less than 5%. In addition, we calculated the correlation between the HR as reported in the clinical trials with the HR as calculated based on the digitized survival curves. Furthermore, we ran a linear regression between the reported and the calculated HR.

We calculated the correlation between HR of the registration studies and the median OS and PFS as reported in these studies, as well as correlation with the RMST and mean survival using Weibull distribution as calculated by our analyses.

Subgroup Analyses

In addition to comparing median, RMST, and mean survival values in the entire research population, we conducted subgroup analyses of these measures specifically for immuno-oncology agents, biomarker-driven targeted therapies, and vascular endothelial growth factor (VEGF)-targeted therapies.

Although most graphs in Figures 2 and 3 demonstrate a visual cutoff at 35 months, the Weibull extrapolation is implemented to infinity, essentially until no patient survives.

Results

Research Population

Between 2013 and 2017, 147 indications were approved for oncology agents for marketing by the FDA [7], of which 83 were indicated for solid tumors, 42 for hematologic malignancies, and 22 for other indications (four tests, three biosimilars, two updates in dosages, two bleeding prevention, two imaging, and an additional nine various approvals). A full list of approvals is available in supplemental online Table 1.

Of the 83 solid tumor approvals, 49 were based on survival endpoints (PFS, 27; OS, 22), 27 on objective response rate, and 7 on other endpoints (recurrence-free survival – 2; invasive disease-free survival – 2; each of pathological complete response / event-free survival / disease-free survival – 1). Two approvals were excluded from our analysis as they referred to label updates (limitation of indication and replacement of capsules with tablets). We omitted three additional indications because of missing data (median or median range were not reached). A flow chart of the 44 included and remaining excluded approvals is shown in supplemental online Figure 1.

Model Appropriateness

To validate the appropriateness of the Weibull model, we compared the median values received by our digitized curves to the original outcome as reported. The outcomes were quite similar—a difference of 0.1 month, with an average median survival for comparator arm of 7.55 months in the original study versus 7.68 months in the Weibull extrapolation and an average median survival for the test arm of 12.20 months and 12.27 months, respectively (full data can be reviewed at Table 1 in the appendix).

The correlation between the HR value as reported in the registration trials used for the drug approvals by FDA with the HR calculated based on the digitized survival curves was 0.86. This high correlation reflects evidence of the digitizing accuracy as well as probable robustness of the Weibull distribution as a solid statistical estimate of the published survival curve. The respective difference between the averages of these two sets of values was 3.7% (average HR 0.60 in the original studies vs. average HR 0.64 in the digitized curves). This relatively low difference demonstrates accurate digitization of survival curves.

As a complementary validation, we ran a linear regression (with intercept defined as zero) between the reported and the digitized HR. The coefficient was 0.934, further demonstrating the similarity between the original and the digitized curves (the full regression output is available in supplemental online Fig. 2).

We found a negative correlation between the HR and the improvement in median survival as reported in the registration studies: −0.68. As expected, this relationship reflects the fact that with a lower HR, the improvement in median OS and PFS is higher. The correlation between reported HR and improvement in calculated RMST was found to be similar, at a level of −0.72. The correlation between reported HR and improvement in calculated mean OS and PFS using Weibull distribution remained negative, at −0.60.

Overall Results and Subgroups Analyses

The average improvement in median OS and PFS as reported in the registration studies was 4.6 months. The average improvement in RMST was lower, at 3.6 months, whereas the average improvement in mean OS or PFS using Weibull distribution was higher, at 6.1 months (see Table 1). The same trend was observed in the subgroup analyses: the improvement in median OS or PFS, RMST, and mean OS or PFS using Weibull extrapolation was 3.2 months, 2.5 months, and 6.9 months, respectively, with the immuno-oncology agents [10–17]; For non-immuno-oncology agents, the improvement was 5.0 months, 3.9 months, and 6.0 months, respectively [18–50] (Fig. 1A).

In additional subgroup analyses, we found the following results: the improvement in median OS or PFS, RMST, and mean OS or PFS using Weibull distribution was 5.9 months, 4.6 months, and 7.4 months, respectively, in the biomarker-driven targeted therapies [19, 21, 22, 24, 25, 33, 38, 43, 46, 48, 49]; For non-biomarker-driven targeted therapies, the improvement was 4.1 months, 3.2 months, and 5.6 months, respectively [10–18, 20, 23, 26–32, 34–37, 39–42, 44, 45, 47, 50] (Fig. 1B).

In the third subgroup analysis, we found the following results: the improvement in median OS or PFS, RMST, and mean OS or PFS using Weibull distribution was 2.1 months, 2.0 months, and 2.2 months, respectively, in the VEGF-driven targeted therapies [26–31, 36]; for non-VEGF-targeted therapies, the improvement was 5.1 months, 3.9 months, and

Table 1. Survival measures of FDA 2013–2017 oncology drug approvals

Drug									Approval			Improvement in survival / PFS (months)				
Brand	Drug	Indication	Line of therapy	Comparator	Primary endpoint	Immunotherapy?	Targeted?	VEGF?	FDA year	FDA date	Registration study	HR (registration study)	Median (registration study)	RMST	HR (Weibull)	Mean (Weibull)
Nexavar	Sorafenib	Thyroid	2nd line	Placebo	PFS	-	-	-	2013	22-11-2013	2014	0.59	5.0	3.8	0.65	5.8
Xalkori	Crizotinib	NSCLC	-	Chemotherapy	PFS	-	Yes	-	2013	20-11-2013	2013	0.49	4.7	4.0	0.60	4.1
Abraxane	Paclitaxel	Pancreas	1st line	Gemcitabine	OS	-	-	-	2013	06-09-2013	2013	0.72	1.8	2.3	0.76	2.4
Gilotrif	Afatinib	NSCLC	1st line	Pemetrexed/ Cisplatin	PFS	-	Yes	-	2013	12-07-2013	2013	0.58	4.2	3.8	0.59	4.5
Tafinlar	Dabrafenib	Melanoma	-	Dacarbazine	PFS	-	Yes	-	2013	29-05-2013	2012	0.30	2.4	2.6	0.52	2.7
Xofigo	Radium Ra 223 dichloride	Prostate	-	Placebo	OS	-	-	-	2013	15-05-2013	2014	0.70	3.1	2.8	0.74	3.6
Tarceva	Erlotinib	NSCLC	1st line	Chemotherapy	PFS	-	Yes	-	2013	14-05-2013	2012	0.37	4.5	6.7	0.40	7.5
Kadcyla	Ado-trastuzumab emtansine	Breast	2nd line	Lapatinib +Capecitabine	PFS	-	Yes	-	2013	22-02-2013	2013	0.65	3.2	3.5	0.63	5.2
Avastin	Bevacizumab	Colorectal	2nd line	Chemotherapy	OS	-	-	Yes	2013	23-01-2013	2013	0.81	1.4	2.0	0.81	2.0
Cyramza	Ramucirumab	NSCLC	Advanced	Docetaxel	OS	-	-	Yes	2014	12-12-2014	2014	0.86	1.4	1.7	0.87	2.1
Avastin	Bevacizumab	Ovarian	2nd line	Chemotherapy	PFS	-	-	Yes	2014	14-11-2014	2014	0.48	3.3	3.4	0.51	3.5
Cyramza	Ramucirumab	Gastro	Advanced	Placebo + Paclitaxel	OS	-	-	Yes	2014	05-11-2014	2014	0.81	2.2	1.4	0.91	1.1
Avastin	Bevacizumab	Cervix	-	Chemotherapy	OS	-	-	Yes	2014	14-08-2014	2014	0.71	3.7	2.5	0.87	2.5
Cyramza	Ramucirumab	Gastro	2nd line	Placebo	OS	-	-	Yes	2014	21-04-2014	2014	0.78	1.4	1.7	0.78	1.8
Keytruda	Pembrolizumab	Melanoma	1st line	Ipilimumab	OS (PFS data)	Yes	-	-	2015	18-12-2015	2015	0.58	1.3	2.5	0.56	3.5
Portrazza	Necitumumab	NSCLC	1st line	Gemcitabine + Cisplatin	OS	-	-	-	2015	24-11-2015	2015	0.84	1.6	1.7	0.86	1.9
Opdivo	Nivolumab	Renal	2nd line	Everolimus	OS	Yes	-	-	2015	23-11-2015	2015	0.73	5.4	2.1	0.90	2.9
Mekinist + Tafinlar	Trametinib + Dabrafenib	Melanoma	-	Dabrafenib + Placebo	OS	-	Yes	-	2015	20-11-2015	2015	0.71	6.4	2.0	0.82	5.3
Cotellic	Cobimetinib	Melanoma	-	Placebo + Vemurafenib	PFS	-	Yes	-	2015	10-11-2015	2016	0.58	5.1	3.2	0.66	4.4
Yondelis	Trabectedin	Liposarcoma	2nd line	Dacarbazine	PFS	-	-	-	2015	23-10-2015	2016	0.55	2.7	2.3	0.53	2.5
Onivyde	Irinotecan	Pancreas	2nd line	5-FU/LV	OS	-	-	-	2015	22-10-2015	2016	0.67	1.9	1.5	0.81	1.5
Opdivo	Nivolumab	NSCLC	2nd line	Docetaxel	OS	Yes	-	-	2015	09-10-2015	2015	0.73	2.8	1.8	0.68	5.5
Lonsurf	Trifluridine/ Tipiracil	Colorectal	2nd line	Placebo	OS	-	-	-	2015	22-09-2015	2015	0.68	1.8	1.8	0.73	2.2
Cyramza	Ramucirumab	Colorectal	2nd line	FOLFIRI	OS	-	-	Yes	2015	24-04-2015	2015	0.84	1.6	1.6	0.84	2.2
Opdivo	Nivolumab	NSCLC	2nd line	Docetaxel	OS	Yes	-	-	2015	04-03-2015	2015	0.59	3.2	3.5	0.56	7.5
Lenvima	Lenvatinib	Thyroid	-	Placebo	PFS	-	-	-	2015	13-02-2015	2015	0.21	14.7	11.3	0.20	23.2
Ibrance	Palbociclib	Breast	Advanced	Letrozole	PFS	-	Yes	-	2015	03-02-2015	2015	0.49	10.0	10.1	0.48	14.7
Opdivo	Nivolumab	HNSCC	2nd line	Investigator's Choice	OS	Yes	-	-	2016	10-11-2016	2016	0.70	2.4	2.3	0.55	5.9
Keytruda	Pembrolizumab	NSCLC	1st line	Chemotherapy	PFS	Yes	-	-	2016	24-10-2016	2016	0.50	4.3	3.6	0.40	14.4
Lartruvo	Olaratumab	STS	-	Doxorubicin	OS	-	-	-	2016	19-10-2016	2016	0.46	11.8	9.6	0.54	17.0
Tecentriq	Atezolizumab	NSCLC	2nd line	Docetaxel	OS	Yes	-	-	2016	18-10-2016	2016	0.73	2.9	1.8	0.62	6.3
Lenvima	Lenvatinib	Renal	2nd line	Everolimus	PFS	-	-	-	2016	13-05-2016	2015	0.40	9.1	5.7	0.40	8.8
Cabometyx	Cabozantinib	Renal	2nd line	Everolimus	PFS	-	-	-	2016	25-04-2016	2015	0.58	3.6	2.9	0.62	3.6
Afinitor	Everolimus	NET	-	Placebo	PFS	-	-	-	2016	26-02-2016	2016	0.48	7.1	3.4	0.75	3.7

(continued)

Table 1. (continued)

Drug								Approval			Improvement in survival / PFS (months)				
Brand	Drug	Indication	Line of therapy	Comparator	Primary endpoint	Immunotherapy?	Targeted?	VEGF?	FDA year	FDA date	Registration study	HR (registration study)	Median (registration study)	RMST	HR (Weibull)	Mean (Weibull)
Ibrance	Palbociclib	Breast	2nd line	Placebo plus Fulvestrant	PFS	-	Yes	-	2016	19-02-2016	2016	0.46	4.9	2.6	0.68	3.5
Halaven	Eribulin	Liposarcoma	2nd line	Dacarbazine	OS	-	-	-	2016	28-01-2016	2016	0.77	2.2	2.3	0.84	2.5
Cabometyx	Cabozantinib	Renal	1st line	Sunitinib	PFS	-	-	-	2017	19-12-2017	2017	0.66	2.6	2.7	0.70	2.9
Verzenio	Abemaciclib	Breast	2nd line	Placebo	PFS	-	Yes	-	2017	28-09-2017	None	0.55	7.1	4.8	0.52	11.4
Zykadia	Ceritinib	NSCLC	-	Chemotherapy	PFS	-	Yes	-	2017	26-05-2017	2017	0.55	8.5	6.4	0.50	16.5
Keytruda	Pembrolizumab	Urothelial carcinoma	2nd line	Chemotherapy	OS	Yes	-	-	2017	18-05-2017	2017	0.73	2.9	2.3	0.62	9.6
Stivarga	Regorafenib	Liver	2nd line	Placebo	OS	-	-	-	2017	27-04-2017	2017	0.63	2.8	3.6	0.68	4.7
Ibrance	Palbociclib	Breast	1st line	Placebo plus Letrozole	PFS	-	Yes	-	2017	31-03-2017	2016	0.58	10.3	5.0	0.63	9.5
Tagrisso	Osimertinib	NSCLC	2nd line	Chemotherapy	PFS	-	Yes	-	2017	30-03-2017	2017	0.30	5.7	4.8	0.35	7.0
Zejula	Niraparib	Ovarian	2nd line	Placebo	PFS	-	-	-	2017	27-03-2017	2016	0.26	15.5	7.8	0.34	16.7
Average													**4.6**	**3.6**		**6.1**

Abbreviations: 5-FU, Fluorouracil; FDA, U.S. Food and Drug Administration; HNSCC, head and neck squamous cell carcinoma; HR, hazard ratio; LV, leucovorin; NET, neuroendocrine tumor; NSCLS, non-small-cell lung cancer; OS, overall survival; PFS, progression-free survival; RMST, restricted mean survival time; STS, soft tissue sarcoma; VEGF, vascular endothelial growth factor.

6.9 months, respectively [10–25, 32–35, 37–50]. In the case of VEGF-targeted targeted drugs, the decrease in RMST and mean survival improvement was the highest compared with other subgroups examined (Fig. 1C).

Generally, the improvement in the mean OS or PFS using Weibull distribution compared with the median survival as reported in the original registration studies was relatively high in the immuno-oncology subgroup. This can be demonstrated in the case of pembrolizumab for first-line non-small cell lung cancer (NSCLC) [15] (Fig. 2A); the improvement in mean survival using Weibull distribution was 14.4 months versus improvement of 4.3 months in the median. However, in some cases, the results of immuno-oncology agents were reversed; in the case of nivolumab for second-line kidney cancer [11] (Fig. 2B), the 2.9-month improvement in mean survival using Weibull distribution was lower than the 5.4-month improvement in median survival.

Other agents demonstrated similar results; in some cases, the mean survival using Weibull distribution presented superior survival compared with the median (lenvatinib for thyroid cancer [37], Fig. 2C), whereas in other cases, the mean survival using Weibull distribution presented inferior survival compared with the median (everolimus for neuroendocrine tumors [42], Fig. 2D).

There were drugs in which median, RMST, and mean survival measures all demonstrated similar outcomes: dabrafenib for *BRAF* positive melanoma [22] (Fig. 3A), bevacizumab for ovarian cancer [28] (Fig. 3B), necitumumab for NSCLC [32] (Fig. 3C), eribulin for liposarcoma [44] (Fig. 3D), and cabozantinib for kidney cancer [45] (Fig. 3E).

Discussion

When analyzing the clinical trials of modern oncology agents approved by the FDA during 2013 to 2017, we found that the average RMST (3.6 months) is lower than the average median OS or PFS (4.6 months). However, when assuming Weibull distribution for the survival after the trial follow-up period, the average mean OS or PFS is higher, at 6.1 months.

Our preplanned subgroup analyses demonstrated significant improvements of mean versus median survival in the immuno-oncology agents, with an average mean survival using Weibull distribution of 6.9 months (compared with an average median of 3.2 months). In an era of new immuno-oncology agents, understanding efficacy purely by the median survival is inappropriate [51]. When some therapies are expected to provide durable responses to a small percentage of patients, a high level of attention to the tail of survival curves is needed. This is possible with modeling techniques in which the survival curves can be digitized to incorporate differences in the distribution of outcomes beyond simply the point estimate of the median OS [52]. In this study, we implemented a statistical model that assists in estimating the potential survival for a population of patients. When measuring the benefit in terms of median and RMST survival, outcomes of the immuno-oncology agents are inferior to other drugs analyzed in this study. However, when implementing the mean OS or PFS using Weibull distribution, the immuno-oncology agents present an improved survival profile.

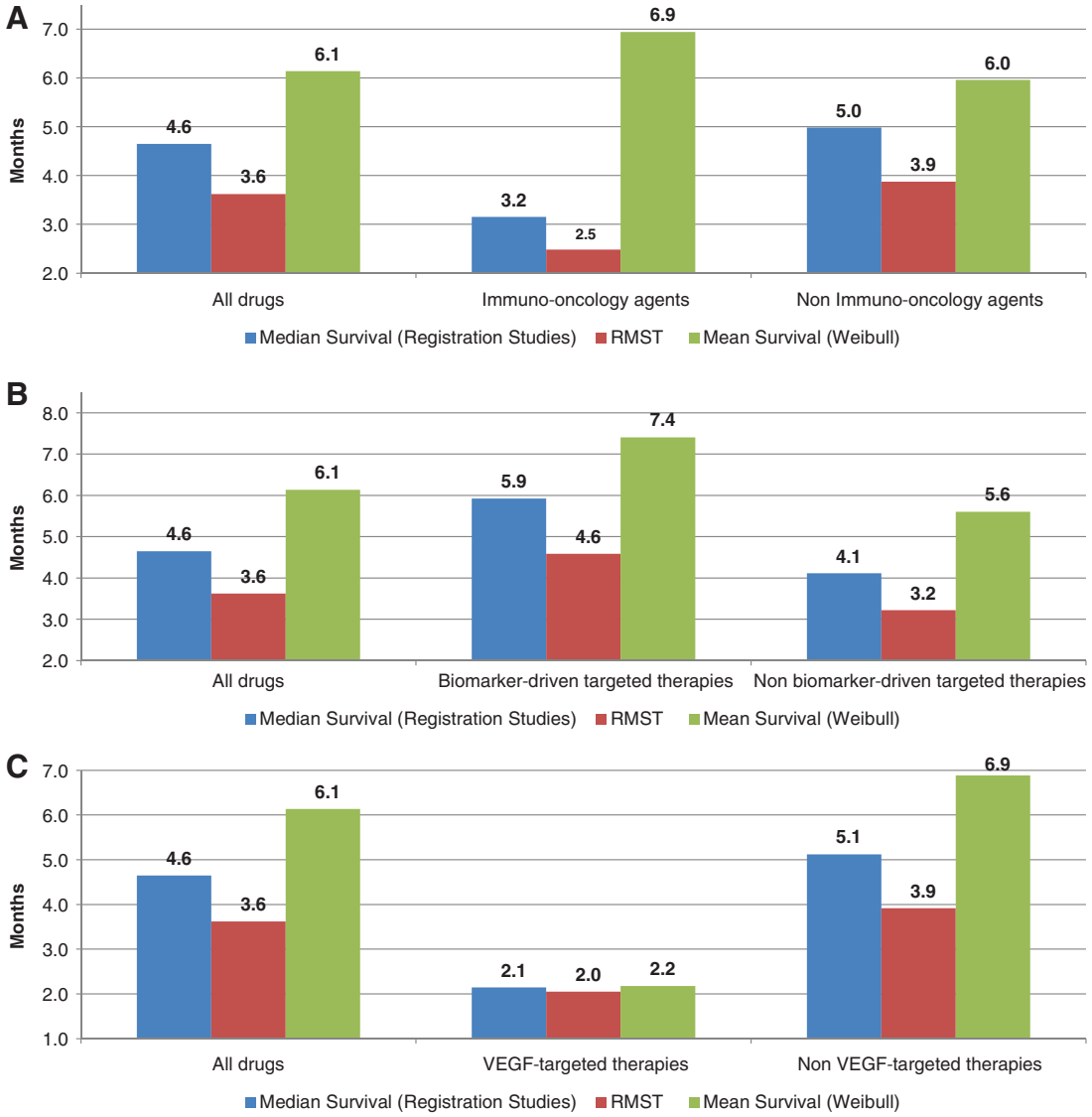


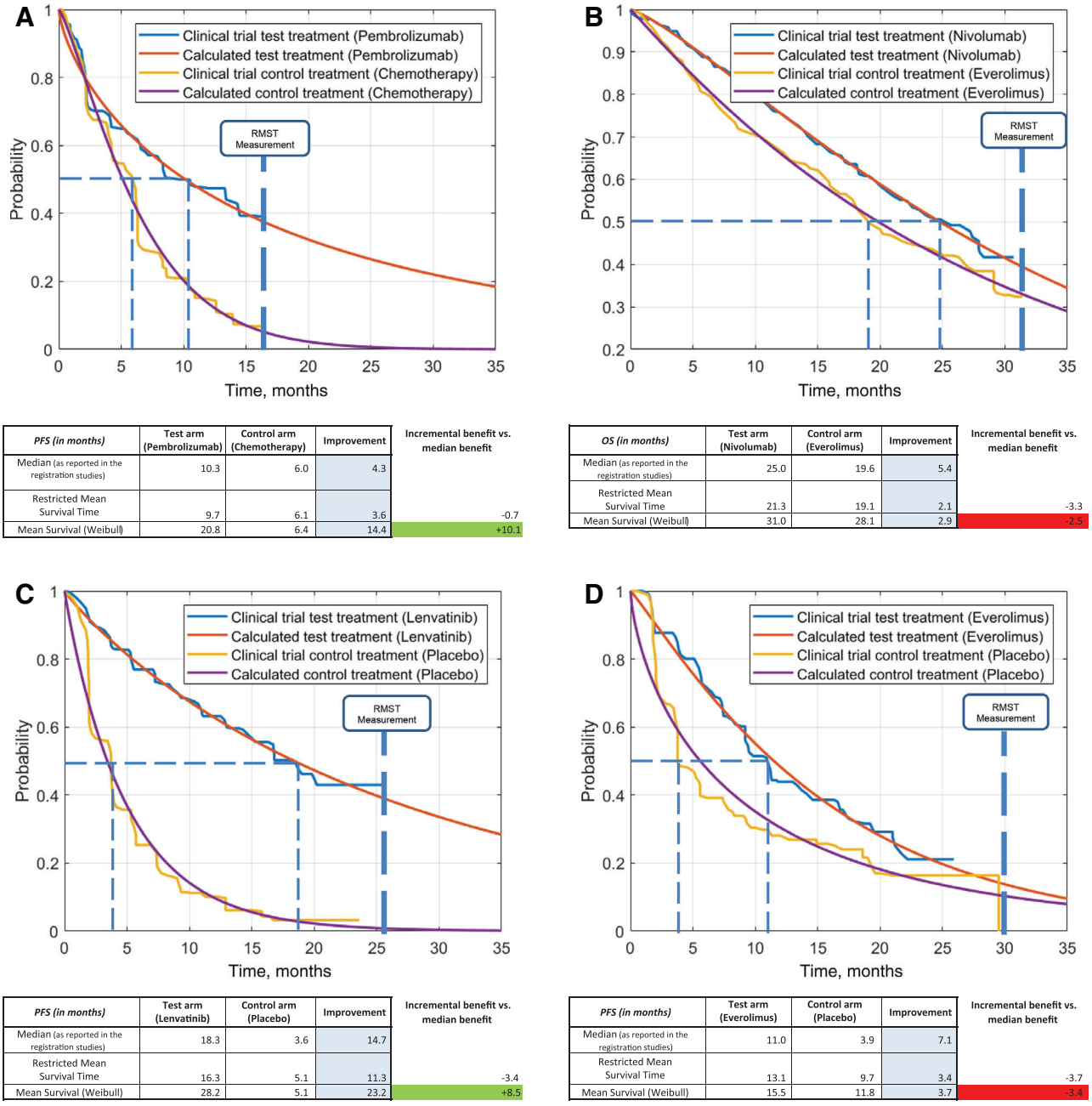
Figure 1. Average improvement in survival measures, mean vs. median: in FDA 2013–2017 Oncology drug approvals. Abbreviations: RMST, restricted mean survival time; VEGF, vascular endothelial growth factor.

An additional subgroup analysis examined biomarker-driven targeted therapies. This group of agents had better results in the three survival measures: median (5.9 months vs. 4.6 months in the entire group), RMST (4.6 months vs. 3.6 months, respectively), and mean (7.4 months vs. 6.1 months, respectively).

The survival measures of the VEGF-targeted therapies were similar: median survival of 2.1 months, RMST of 2.0 months, and mean survival using Weibull distribution of 2.2 months. These outcomes are inferior to the results of other therapies with less potential for durable survival [53].

The subgroup analysis highlights the differences between mean and median survival measures in the immuno-oncology agents. In this subgroup, the discordances between mean and median values were the highest, stemming from the potential for durable survival for a minority of patients. In these cases, the proportional hazards assumption is very questionable and there is space for alternative measures. The potential survival is underestimated using the median survival but better captured when incorporating the long-term mean survival. Therefore, we propose that the use of the mean as a complementary measure in reporting of prospective trials will be most valuable in the case of immuno-oncology agents. Adding the mean as an additional reported measure may supply meaningful value to clinicians when trying to understand the long-term impact of new treatments, as well as payers who are required to prioritize drugs covered under budget constraints of reimbursement plans.

Various value-based frameworks were recently presented as a means to address the trend of increasing health care expenditure as a whole, and specifically oncology pharmaceuticals [54–58]. In light of these frameworks, understanding the true benefit of innovative treatments is essential. The tremendous interest and excitement associated with immuno-oncology is focused on the tail of the survival curves and the potential for durable survival in some patients. The HR and RMST difference are complementary techniques that provide alternative methods of summarizing treatment effects [5]. To identify more precisely the perceived value, the design and analysis of a conventional cancer clinical trial can be improved by adopting a robust statistical procedure that enables clinically



PFS (in months)	Test arm (Pembrolizumab)	Control arm (Chemotherapy)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	10.3	6.0	4.3	
Restricted Mean Survival Time	9.7	6.1	3.6	-0.7
Mean Survival (Weibull)	20.8	6.4	14.4	+10.1

OS (in months)	Test arm (Nivolumab)	Control arm (Everolimus)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	25.0	19.6	5.4	
Restricted Mean Survival Time	21.3	19.1	2.1	-3.3
Mean Survival (Weibull)	31.0	28.1	2.9	-2.5

PFS (in months)	Test arm (Lenvatinib)	Control arm (Placebo)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	18.3	3.6	14.7	
Restricted Mean Survival Time	16.3	5.1	11.3	-3.4
Mean Survival (Weibull)	28.2	5.1	23.2	+8.5

PFS (in months)	Test arm (Everolimus)	Control arm (Placebo)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	11.0	3.9	7.1	
Restricted Mean Survival Time	13.1	9.7	3.4	-3.7
Mean Survival (Weibull)	15.5	11.8	3.7	-3.4

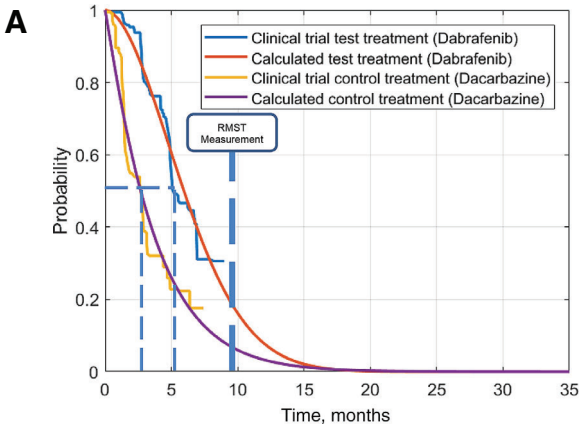
Figure 2. Model demonstration: major differences.
Abbreviations: OS, overall survival; PFS, progression-free survival; RMST, restricted mean survival time.

meaningful interpretations of the treatment effect. The RMST-based quantitative method may be used as a primary tool for future cancer trials or to help us to better understand the clinical interpretation of the HR [4].

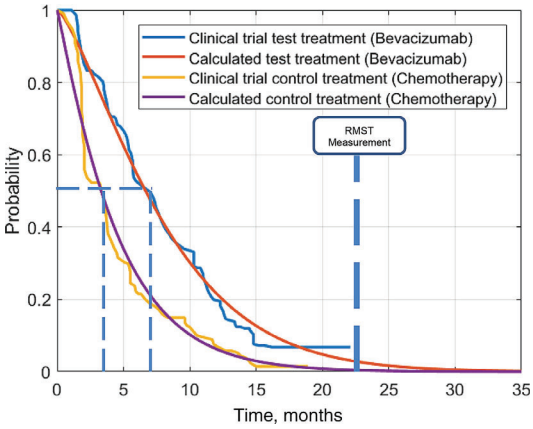
A major limitation of this study is the interpretation of RMST within the time frame chosen for the cutoff of the clinical trials (t). The comparisons between gains with RMST and medians or parametric means (which are unrestricted) need to be seen in the context of t and may well explain disparate results within trials. In other words, the difference between RMSTs and the other two metrics depends on the choice of t. To moderate this effect, we used only the therapeutic outcomes, including follow-up periods, as reported in the clinical studies used for the FDA drug approvals.

An additional limitation is that the processing of digitized survival curves rather than using the original data generated in the clinical trials may decrease statistical certainty. However, we indeed reached a high correlation between reported clinical results and digitized data.

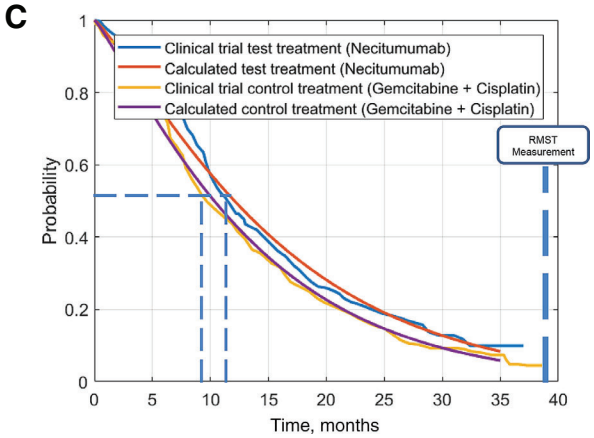
Another limitation is assuming Weibull distribution for the mean survival calculation. As we used extrapolation commencing at the end of the clinical trial follow-up period—where the highest amount of uncertainty lies—by definition, this extrapolation has a high level of uncertainty. This limitation is relevant for the entire population analyzed in this work as a whole, and specifically to the immuno-oncology agents where the Weibull distribution may not be an appropriate estimation for the tail of the survival curve. However, because of the lack of available data, there is no clear alternative



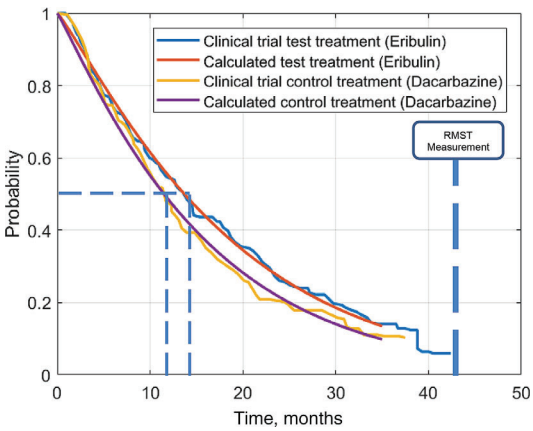
PFS (in months)	Test arm (Dabrafenib)	Control arm (Dacarbazine)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	5.1	2.7	2.4	
Restricted Mean Survival Time	5.9	3.3	2.6	0.2
Mean Survival (Weibull)	6.4	3.7	2.7	0.3



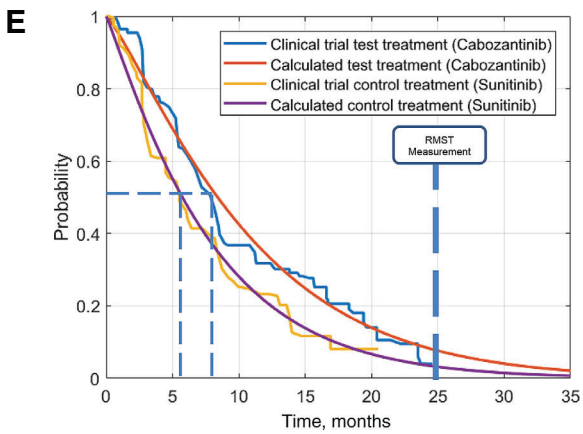
PFS (in months)	Test arm (Bevacizumab)	Control arm (Chemotherapy)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	6.7	3.4	3.3	
Restricted Mean Survival Time	7.9	4.5	3.4	0.1
Mean Survival (Weibull)	8.0	4.5	3.5	0.2



OS (in months)	Test arm (Necitumumab)	Control arm (Gemcitabine + Cisplatin)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	11.5	9.9	1.6	
Restricted Mean Survival Time	14.8	13.1	1.7	0.1
Mean Survival (Weibull)	15.4	13.5	1.9	0.3



OS (in months)	Test arm (Eribulin)	Control arm (Dacarbazine)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	13.5	11.3	2.2	
Restricted Mean Survival Time	16.9	14.7	2.3	0.1
Mean Survival (Weibull)	18.1	15.6	2.5	0.3



PFS (in months)	Test arm (Cabozantinib)	Control arm (Sunitinib)	Improvement	Incremental benefit vs. median benefit
Median (as reported in the registration studies)	8.2	5.6	2.6	
Restricted Mean Survival Time	10.0	7.4	2.7	0.1
Mean Survival (Weibull)	10.7	7.8	2.9	0.3

Figure 3. Model demonstration: minor differences.
Abbreviations: OS, overall survival; PFS, progression-free survival; RMST, restricted mean survival time.

option [59]. We chose to use Weibull as an exclusive model to fit all survival curves as alternative parametric models were found to be inferior. However, we recognize that even with the Weibull extrapolation, a degree of uncertainty remains.

The presence of censoring may potentially contaminate the calculation of mean survival measures, although this phenomenon may also affect the median value as reported in the results of the clinical trials.

Incorporating the RMST in addition to the median survival in the standard reporting of oncology clinical trials will potentially supply more meaningful information to patients, physicians, and policymakers. The suggested measures may be more appropriate for chemotherapy and biomarker-driven targeted therapies. In the case of immuno-oncology agents, measures that are based on the mean of the entire population may be less important than milestone reporting [51]. However, measures of the mean are likely still more important than the median results.

Conclusion

Understanding the efficacy of treatments is crucial for patients, physicians, and policymakers. The most common and straightforward measure used in the outcome reporting of oncology clinical trials is the median survival. Although the median is easy to understand, it describes only a single time point. The most scientifically common measure is the HR; however, there is frequently a potential violation of the PH assumption underlying this measure, and its actual interpretation is difficult for patients, physicians, and policymakers. In cases in which a small proportion of patients are expected to achieve long-term survival and parametric extrapolation presents a good fit to the original data set, it should be considered to add mean as a supplementary measure to the trial outcomes. Mean survival may supply valuable and accurate information for different stakeholders—physicians in their communication with peers and with patients; regulatory bodies in interpretation of approved drugs efficacy; and policymakers in reimbursement decisions.

Author Contributions

Conception/design: Omer Ben-Aharon, Daniel A. Goldstein, Moshe Leshno
Provision of study materials or patients: Omer Ben-Aharon, Daniel A. Goldstein, Moshe Leshno
Collection and/or assembly of data: Omer Ben-Aharon, Daniel A. Goldstein, Moshe Leshno
Data analysis and interpretation: Omer Ben-Aharon, Daniel A. Goldstein, Racheli Magnezi, Moshe Leshno
Manuscript writing: Omer Ben-Aharon, Daniel A. Goldstein
Final approval of manuscript: Omer Ben-Aharon, Daniel A. Goldstein, Racheli Magnezi, Moshe Leshno

Disclosures

The authors indicated no financial relationships.

References

1. Ellis LM, Bernstein DS, Voest EE et al. American Society of Clinical Oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277–1280.

2. Long GV, Stroyakovskiy D, Gogas H et al. Dabrafenib and trametinib versus dabrafenib and placebo for Val600 BRAF-mutant melanoma: A multicentre, double-blind, phase 3 randomised controlled trial. Lancet 2015;386:444–451.

3. Wang-Gillam A, Li CP, Bodoky G et al. Nanoliposomal irinotecan with fluorouracil and folinic acid in metastatic pancreatic cancer after previous gemcitabine-based therapy (NAPOLI-1): A global, randomised, open-label, phase 3 trial. Lancet 2016;387:545–557.

4. Pak K, Uno H, Kim DH et al. Interpretability of cancer clinical trial results using restricted mean survival time as an alternative to the hazard ratio. Jama Oncol 2017;3:1692–1696.

5. A'Hern RP. An obligatory end point for time-to-event analysis in cancer trials? J Clin Oncol 2016;34:3474–3476.

6. Maio M, Grob JJ, Aamdal S et al. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. J Clin Oncol 2015;33:1191–1196.

7. U.S. Food & Drug Administration. Hematology/Oncology (Cancer) Approvals & Safety Notifications. 2018. Available at https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm. Accessed January 1, 2018.

8. Nelder JA, Mead R. A simplex method for function minimization. Comput J 1965;7:308–313.

9. Lagarias JC, Reeds JA, Wright MH et al. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim 1998;9:112–147.

10. Robert C, Schachter J, Long GV et al. Pembrolizumab versus ipilimumab in advanced melanoma. N Engl J Med 2015;372:2521–2532.

11. Motzer RJ, Escudier B, McDermott DF et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. N Engl J Med 2015;373:1803–1813.

12. Borghaei H, Paz-Ares L, Horn L et al. Nivolumab versus docetaxel in advanced non-squamous non–small-cell lung cancer. N Engl J Med 2015;373:1627–1639.

13. Brahmer J, Reckamp KL, Baas P et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl J Med 2015;373:123–135.

14. Ferris RL, Blumenschein G, Fayette J et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med 2016;375:1856–1867.

15. Reck M, Rodríguez-Abreu D, Robinson AG et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med 2016;375:1823–1833.

16. Fehrenbacher L, Spira A, Ballinger M et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): A multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837–1846.

17. Bellmunt J, de Wit R, Vaughn DJ et al. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. N Engl J Med 2017;376:1015–1026.

18. Brose MS, Nutting CM, Jarzab B et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic diff erentiated thyroid cancer: A randomised, double-blind, phase 3 trial. Lancet 2014;384:319–328.

19. Shaw AT, Kim DW, Nakagawa K et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med 2013;368:2385–2394.

20. Von Hoff DD, Ervin T, Arena FP et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med 2013;369:1691–1703.

21. Sequist LV, Yang JC, Yamamoto N et al. Phase III study of afatinib or cisplatin plus pemetrexed in phase III study of afatinib or cisplatin plus pemetrexed in EGFR mutations. J Clin Oncol 2013;31:3327–3334.

22. Hauschild A, Grob JJ, Demidov LV et al. Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial. Lancet 2012;380:358–365.

23. Hoskin P, Sartor O, O'Sullivan JM et al. Efficacy and safety of radium-223 dichloride in patients with castration-resistant prostate cancer and symptomatic bone metastases, with or without previous docetaxel use: A prespecified subgroup analysis from the randomised, double-blind, phase 3 ALSYMPC. Lancet Oncol 2014;15:1397–1406.

24. Rosell R, Carcereny E, Gervais R et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial. Lancet Oncol 2012;13:239–246.

25. Verma S, Miles D, Gianni L et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012;367:1783–1791.

26. Bennouna J, Sastre J, Arnold D et al. Continuation of bevacizumab after first progression in metastatic colorectal cancer (ML18147): A randomised phase 3 trial. Lancet Oncol 2013;14:29–37.

27. Garon EB, Ciuleanu TE, Arrieta O et al. Ramucirumab plus docetaxel versus placebo plus docetaxel for second-line treatment of stage IV non-small-cell lung cancer after disease progression on platinum-based therapy (REVEL): A multicentre, double-blind, randomised phase 3 trial. Lancet 2014;384:665–673.

28. Pujade-Lauraine E, Hilpert F, Weber B et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. J Clin Oncol 2014;32:1302–1308.

29. Wilke H, Muro K, Van Cutsem E et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): A double-blind, randomised phase 3 trial. Lancet Oncol 2014;15:1224–1235.

30. Tewari KS, Sill MW, Long HJ 3rd et al. Improved survival with bevacizumab in advanced cervical cancer. N Engl J Med 2014;370:734–743.

31. Fuchs CS, Tomasek J, Yong CJ et al. Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): An international, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet 2014;383:31–39.

32. Thatcher N, Hirsch FR, Luft AV et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): An open-label, randomised, controlled phase 3 trial. Lancet Oncol 2015;16:763–774.

33. Ascierto PA, McArthur GA, Dréno B et al. Cobimetinib combined with vemurafenib in advanced BRAFV600-mutant melanoma (coBRIM): Updated efficacy results from a randomised, double-blind, phase 3 trial. Lancet Oncol 2016;17:1248–1260.

34. Demetri GD, von Mehren M, Jones RL et al. Efficacy and safety of trabectedin or dacarbazine for metastatic liposarcoma or leiomyosarcoma after failure of conventional chemotherapy: Results of a phase III randomized multicenter clinical trial. J Clin Oncol 2016;34:786–793.

35. Mayer RJ, Van Cutsem E, Falcone A et al. Randomized trial of TAS-102 for refractory metastatic colorectal cancer. N Engl J Med 2015;372:1909–1919.

36. Tabernero J, Yoshino T, Cohn AL et al. Ramucirumab versus placebo in combination with secondline FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): A randomised, double-blind, multicentre, phase 3 study. Lancet Oncol 2015;16:499–508.

37. Schlumberger M, Tahara M, Wirth LJ et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. N Engl J Med 2015;372:621–630.

38. Finn RS, Crown JP, Lang I et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): A randomised phase 2 study. Lancet Oncol 2015;16:25–35.

39. Tap WD, Jones RL, Van Tine BA et al. Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: An open-label phase 1b and randomised phase 2 trial. Lancet 2016;388:488–497.

40. Motzer RJ, Hutson TE, Glen H et al. Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: A randomised, phase 2, open-label, multicentre trial. Lancet Oncol 2015;16:1473–1482.

41. Choueiri TK, Escudier B, Powles T et al. Cabozantinib versus everolimus in advanced renal-cell carcinoma. N Engl J Med 2015;373:1814–1823.

42. Yao JC, Fazio N, Singh S et al. Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): A randomised, placebo-controlled, phase 3 study. Lancet 2016;387:968–977.

43. Cristofanilli M, Turner NC, Bondarenko I et al. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine. Lancet Oncol 2016;17:425–439.

44. Schöffski P, Chawla S, G Maki RG et al. Eribulin versus dacarbazine in previously treated patients with advanced liposarcoma or leiomyosarcoma: A randomised, open-label, multicentre, phase 3 trial. Lancet 2016;387:1629–1637.

45. Choueiri TK, Halabi S, Sanford BL et al. Cabozantinib versus sunitinib as initial targeted therapy for patients with metastatic renal cell carcinoma of poor or intermediate risk: The Alliance A031203 CABOSUN trial. J Clin Oncol 2017;35:591–597.

46. Soria JC, Tan DSW, Chiari R et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): A randomised, open-label, phase 3 study. Lancet. 2017;389:917–929.

47. Bruix J, Qin S, Merle P et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): A randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;389:56–66.

48. Finn RS, Martin M, Rugo HS et al. Palbociclib and letrozole in advanced breast cancer. N Engl J Med 2016;375:1925–1936.

49. Mok TS, Wu YL, Ahn MJ et al. Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. N Engl J Med 2017;376:629–640.

50. Mirza MR, Monk BJ, Herrstedt J et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. N Engl J Med 2016;375:2154–2164.

51. Hellmann MD, Kris MG, Rudin CM. Medians and milestones in describing the path to cancer cures: Telling "tails". JAMA Oncol 2016;2:167–168.

52. Goldstein DA. Understanding the value of cancer drugs–the devil is in the detail. Cancer 2016;122:2292–2295.

53. Ellis LM, Hicklin DJ. VEGF-targeted therapy: Mechanisms of anti-tumour activity. Nat Rev Cancer 2008;8:579–591.

54. Schnipper LE, Davidson NE, Wollins DS et al. Updating the American Society of Clinical Oncology Value Framework: revisions and reflections in response to comments received. J Clin Oncol 2016;34:2925–2934.

55. Cherny NI, Dafni U, Bogaerts J et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. Ann Oncol 2017;28:2340–2366.

56. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines) with NCCN Evidence Blocks. 2018. Available at https://www.nccn.org/evidenceblocks/. Accessed January 1, 2018.

57. ICER Institute for Clinical and Economic Review. ICER Value Assessment Framework. 2018. Available at https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/. Accessed January 1, 2018.

58. Memorial Sloan Kettering. Drug Abacus. 2018. Available at https://drugpricinglab.org/tools/drug-abacus/. Accessed Jan 1, 2018.

59. Ben-Aharon O, Magnezi R, Leshno M et al. Association of immunotherapy with durable survival as defined by value frameworks for cancer care. JAMA Oncol 2018;4:326–332.

See http://www.TheOncologist.com for supplemental material available online.