

# Quantifying Treatment Benefit in Molecular Subgroups to Assess a Predictive Biomarker <sup>CME</sup>

Alexia Iasonos<sup>1</sup>, Paul B. Chapman<sup>2</sup>, and Jaya M. Satagopan<sup>1</sup>

## Abstract

An increased interest has been expressed in finding predictive biomarkers that can guide treatment options for both mutation carriers and noncarriers. The statistical assessment of variation in treatment benefit (TB) according to the biomarker carrier status plays an important role in evaluating predictive biomarkers. For time-to-event endpoints, the hazard ratio (HR) for interaction between treatment and a biomarker from a proportional hazards regression model is commonly used as a measure of variation in TB. Although this can be easily obtained using available statistical software packages, the interpretation of HR is not straightforward. In this article,

we propose different summary measures of variation in TB on the scale of survival probabilities for evaluating a predictive biomarker. The proposed summary measures can be easily interpreted as quantifying differential in TB in terms of relative risk or excess absolute risk due to treatment in carriers versus noncarriers. We illustrate the use and interpretation of the proposed measures with data from completed clinical trials. We encourage clinical practitioners to interpret variation in TB in terms of measures based on survival probabilities, particularly in terms of excess absolute risk, as opposed to HR. *Clin Cancer Res*; 22(9); 2114–20. ©2016 AACR.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed by the other authors.

## Editor's Disclosures

The following editor(s) reported relevant financial relationships: W.E. Barlow—None.

## CME Staff Planners' Disclosures

The members of the planning committee have no real or apparent conflict of interest to disclose.

## Learning Objectives

Upon completion of this activity, the participant should have a better understanding of statistical approaches for quantifying differential treatment benefit in biomarker-specific subgroups. The participant should focus on comparing treatment benefit with different time-to-event endpoints such as overall survival or progression-free survival.

## Acknowledgment of Financial or Other Support

This activity does not receive commercial support.

## Introduction

The development of new targeted therapies and immunotherapy has improved the prognosis of several cancer types.

Nonetheless, curative therapy remains a challenge and benefit is generally seen only in a subset of patients. This challenge has accelerated the need to identify novel biomarkers such as somatic mutations in tumors or overexpression of certain proteins that can predict treatment benefit (TB). For example, panitumumab and cetuximab (both *EGFR*-specific mAbs) are effective in treating colorectal cancer patients without *KRAS* mutations (1, 2). Thus, *KRAS* carrier status is a predictive marker in colorectal cancer. In these studies of colon cancer, it is clear why one molecularly defined subgroup is expected to benefit from a targeted treatment more than the other group. However, there are cases in which the biologic underpinnings of variation in TB are still not fully understood. Therefore, novel biomarker studies are currently being pursued to improve our understanding of the molecular basis of TB. For example, the relationship of anti-PD-1 antibody therapy to PD-L1 tumor

<sup>1</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. <sup>2</sup>Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

**Corresponding Author:** Alexia Iasonos, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, 2nd Floor, New York, NY 10017. Phone: 646-888-8255; Fax: 929-321-1517; E-mail: [iasonosa@mskcc.org](mailto:iasonosa@mskcc.org)

**doi:** 10.1158/1078-0432.CCR-15-2517

©2016 American Association for Cancer Research.

expression is currently being examined in various disease types (3–5). In this article, we refer to a genetic factor as a "predictive biomarker" when carriers of a specific genetic mutation have significantly higher TB than noncarriers, or vice versa.

Statistical models play an important role in identifying predictive biomarkers. When an interaction takes place between a biomarker and treatment in a statistical model, it means that the effect of treatment (or TB) differs in carriers versus noncarriers of a biomarker, i.e., there is differential TB. Thus, considerable interest has arisen in evaluating biomarker–treatment interactions using statistical models to identify novel predictive biomarkers. A pivotal question is how to measure interaction—the differential TB between carriers versus noncarriers in a clinically interpretable manner. In this article, we examine this issue in the context of time-to-event endpoints such as progression-free survival (PFS) or overall survival (OS).

In studies of PFS/OS, we typically demonstrate interaction using two sets of Kaplan–Meier curves for treated and untreated individuals according to the biomarker status. The magnitude of differential TB is given by a summary measure of biomarker–treatment interaction based on a single hazard ratio (HR). However, the significance of HR is often difficult to translate into meaningful clinical terms (6). Therefore, in this article, we propose alternative approaches based on survival probabilities at a specific time point that can help evaluate TB within molecular subgroups and identify predictive biomarkers in a clinically interpretable manner. We illustrate our proposed approaches using examples from published trials.

## Materials and Methods

### Conventional practice for measuring and comparing treatment benefit

For time-to-event endpoints, the conventional approach to determine whether a biomarker is predictive involves fitting a Cox proportional hazards regression model. This is typically reported by an estimated HR for treatment (referred to as the treatment main effect, which is commonly taken as a measure of TB among noncarriers); the HR for biomarker status (referred to as the biomarker main effect) and the HR for treatment–biomarker interaction effect. Assuming there are two treatment groups and two marker groups, the interaction effect is the ratio of treatment HR in carriers to noncarriers (Table 1). When there is no inter-

action effect, the TB (measured in terms of HR) is the same for carriers and noncarriers.

The conventional approach has been used widely; clinicians are familiar with HR from such models as a summary measure of the TB. An HR = 0.5 means the control arm has twice the hazard of dying relative to the treatment arm, which can be explained as follows: If we treat patient A with Drug A, then Patient A's hazard of dying is reduced 2-fold by this treatment as opposed to the patient being left untreated. If HR = 0.75 then we can say that being left untreated will increase Patient A's hazard by 33% ( $1/0.75 = 1.33$ ). An HR of 2 means the treatment arm has twice the hazard of the control arm, i.e., the treatment is harmful.

Hazard, by definition, is the rate of an event occurring in a time interval. The hazard of death in a certain time interval can be interpreted as the rate of dying in that interval, given that the person has survived up to that time point. In technical terms, the HR is the ratio of hazard of death for treated relative to control groups. The conventional approach uses these definitions of hazard and HR, with the additional assumption that the hazard for treated patients at any time interval is proportional to the hazard for untreated patients. This approach compares the ratios of the hazards of treated versus untreated individuals across all time points and determines whether the overall ratio, aggregated across the time points, is significantly different from 1.

This approach has some limitations. First, this method relies heavily on the assumption of proportional hazards, which stipulates that the effects of treatment and biomarker do not depend upon time (7). Under the proportional hazards model, the interaction effect, or the predictive value, is taken to be the same across all time points. However, treatment is likely to have a temporal effect; for example, TB may be best realized only after a certain time point. The second limitation is interpretability. Although the parameters of the proportional hazards model are easy to obtain and can be interpreted as HR, they are not easily interpretable in the clinic. Specifically, TB on the hazard scale is very difficult to convey in a dialog between clinical practitioners and patients. The explanation to Mrs. Jones about HR ignores time and gives a single summary measure, thereby making an implicit assumption that the ratio remains constant over time. However, the potential for temporal effect, mentioned above, can affect the magnitude of TB. HR is generally interpreted as the "chance" (probability), as opposed to the "hazard" of dying, although the two terms are mathematically distinct (7).

**Table 1.** Measures for comparing treatment benefit (TB) in marker-specific subgroups: quantitative value of predictive marker

Measure/approach	Marker's effect on treatment	Scale of TB	Statistical interpretation	No differential TB benefit (null setting)
Marker–treatment interaction effect	$\log\{h(d; B = 1, T = 1)\} - \log\{h(d; B = 1, T = 0)\} - \log\{h(d; B = 0, T = 1)\} + \log\{h(d; B = 0, T = 0)\}$	Log hazard scale	Interaction on the hazard scale; model based	Value of 0 indicates no difference in TB
Relative treatment benefit	$\frac{S(d; B = 1, T = 1)}{S(d; B = 1, T = 0)} \times \frac{S(d; B = 0, T = 0)}{S(d; B = 0, T = 1)}$	Ratio of survival probabilities (log of survival scale)	Marker-dependent survival ratio; non-model based (similar interpretation as relative risk <sup>a</sup> )	Value of 1 indicates no difference in TB
Absolute treatment benefit	$S(d; B = 1, T = 1) - S(d; B = 1, T = 0) - S(d; B = 0, T = 1) + S(d; B = 0, T = 0)$	Survival probability scale	Marker-dependent survival differences (on probability scale); non-model based <sup>a</sup>	Value of 0 indicates no difference in TB

NOTE:  $h(d; B = 1, T = 1)$  is the hazard of death or progression occurring at time  $d$ , and  $S(d; B = 1, T = 1)$  is the survival or progression-free probability at time  $d$  for someone in the biomarker-positive group receiving treatment.

Abbreviations: B = 1, +biomarker; B = 0, -biomarker; T = 1, treated group; T = 0, nontreated group;  $d$ , duration from start of treatment.

<sup>a</sup>non-model based when there is no adjustment for covariates; could be model based if  $S$  is conditional on known covariates.

### Proposed measures based on survival time: clinical interpretability

Unlike the HR, survival probabilities have a straightforward interpretation as the probability of remaining progression free, or free of death, beyond a certain specified time. For example, patients or clinicians might be interested in knowing the likelihood of remaining progression free beyond 6 or 12 weeks or months with and without treatment using a particular regimen. These probabilities can be influenced by mutation carrier status, especially if TB varies according to that status. When communicating TB to patients, it is important to tailor estimates of TB according to their individual biomarker status. The challenge is how to measure and compare TB between carriers and noncarriers from available data.

Consider the subset of mutation carriers, in which some individuals receive the treatment and others do not. We can calculate survival probabilities for treated and untreated carriers separately. Similarly, we can obtain these probabilities for noncarriers. When there are no additional covariates under consideration, the survival probabilities can be calculated using the Kaplan–Meier approach. Under this approach, the survival probabilities' estimates are not prone to errors introduced by deviations from the proportional hazards assumption. The methods below can be applied to any time-to-event endpoint (PFS or OS). We describe our proposed measures below and refer the reader to Supplementary Data for details about distributional properties and hypothesis tests.

#### Relative treatment benefit

Suppose we want to measure differential TB at a certain time point. We can calculate relative treatment benefit (RTB) in terms of the ratio of survival probabilities in the subset of patients with mutations, given by:

TB among carriers

$$= \frac{\text{Survival probability for treated patients with mutation}}{\text{Survival probability for untreated patients with mutation}}$$

Similarly, we can calculate TB among noncarriers as:

TB among noncarriers

$$= \frac{\text{Survival probability for treated patients without mutation}}{\text{Survival probability for untreated patients without mutation}}$$

We propose to measure predictive value as the ratio of these marker-specific TB, which is the RTB, and can be interpreted as a relative risk ratio. It is equivalent to differences between survival probabilities on the logarithm scale (Table 1). The RTB can be calculated for any time point of interest. There is no differential TB on the survival probability scale at a given time point when the corresponding RTB = 1. A hypothesis test for RTB is provided (Supplementary Data).

#### Absolute treatment benefit

The RTB is a ratio measure that quantifies the proportional change in TB for carriers relative to noncarriers. In practice, the increase or decrease in survival probabilities calculated as a difference might be more informative (and easier to understand) than a proportional change which is given by:

TB among carriers

$$= \text{Survival probability for treated patients with mutation} \\ - \text{Survival probability for untreated patients with mutation}$$

TB among noncarriers

$$= \text{Survival probability for treated patients without mutation} \\ - \text{Survival probability for untreated patients without mutation}$$

Here, we define a marker's predictive value as the difference between these differences in survival probabilities, or absolute treatment benefit (ATB). As above, this measure allows the predictive value of a biomarker to vary with time. Among carriers, the difference in survival probabilities represents the excess (or reduction) in the proportion of treated carriers who remain progression free, for example, at a particular time compared with untreated patients. The same interpretation holds among noncarriers. This measure of predictive value has the advantage of retaining the probability scale, enabling a straightforward interpretation. There is no predictive value at a given time point when the corresponding ATB = 0. A hypothesis test for ATB is provided (Supplementary Data).

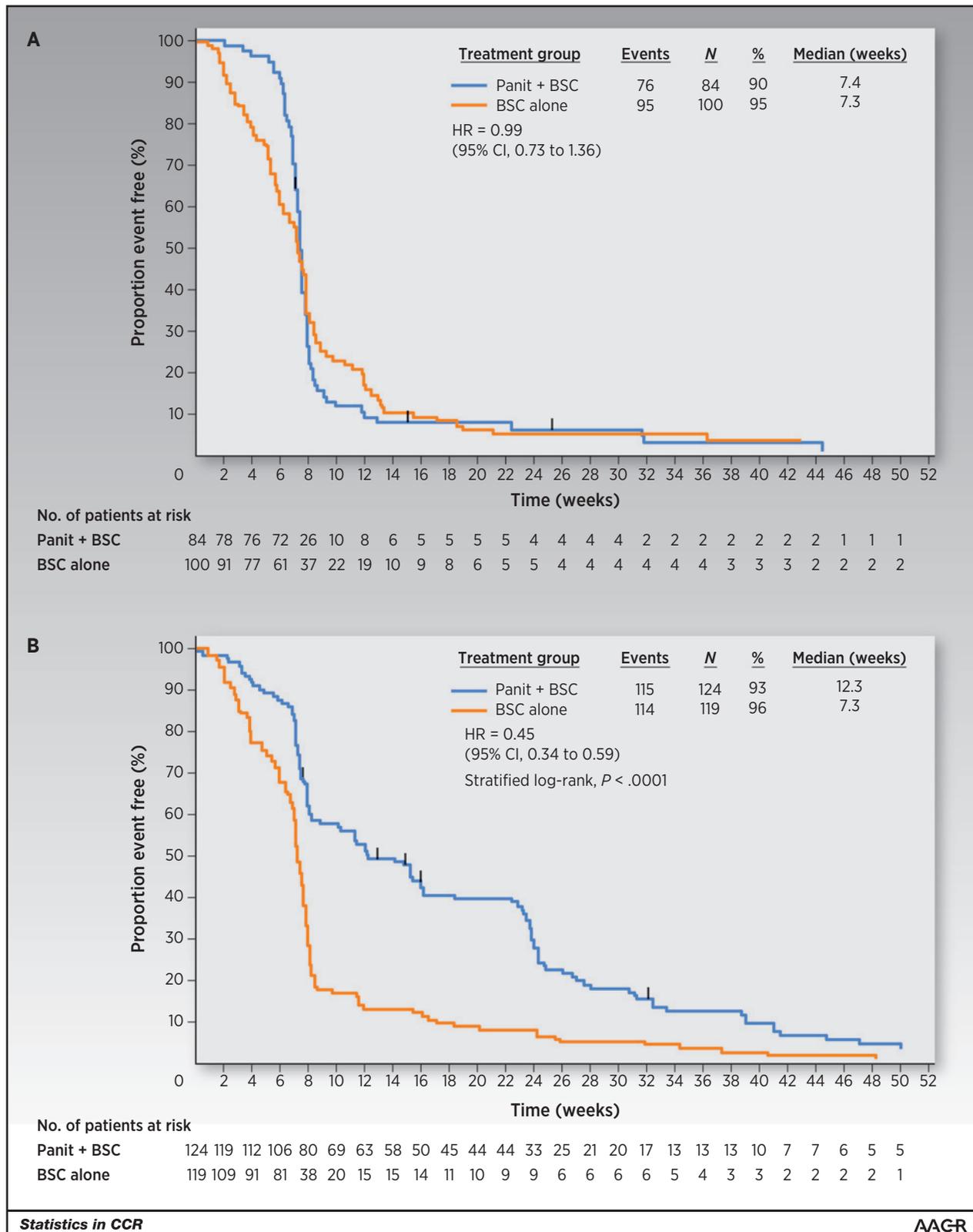
## Results

### Illustrative example, Trial 1

Using the conventional approach, Amado and colleagues (1) showed that *KRAS* wild-type status predicts the benefit of panitumumab treatment (as measured by PFS) in metastatic colorectal cancer patients. In their study, the TB as measured by HR was 0.45 and 0.99 in *KRAS* wild-type and mutant patients, respectively. Hence, the interaction effect was 2.2 (0.99/0.45) in a proportional hazards model, which was significantly different from 1 ( $P < 0.0001$ ; Fig. 1, Table 2). Using similar analyses, Karapetis and colleagues (2) showed a significant interaction between cetuximab treatment and *KRAS* mutation status (Supplementary Fig. S1), and concluded that *KRAS* wild-type status predicts the benefit of cetuximab treatment in patients with *EGFR*-positive metastatic colorectal cancer.

**Relative treatment benefit.** Using the data from Amado and colleagues (1), the 6-week PFS probabilities for treated and untreated patients were 93% and 65%, respectively, for carriers and 88% and 69%, respectively, for noncarriers. The 6-week TB is 1.43 (93/65) for carriers and 1.28 (88/69) for noncarriers. Hence, the RTB at 6 weeks is 1.12 (=1.43/1.28). Similar calculations show the RTB to be equal to 0.17 (=0.57/3.33) at 12 weeks. This means that the 6-week panitumumab benefit for patients with *KRAS* mutations is 12% better compared with the benefit in patients with wild-type *KRAS*, whereas at 12 weeks, panitumumab benefit for patients with *KRAS* mutations is 83% worse relative to panitumumab benefit in patients with wild-type *KRAS* (Table 2). Figure 1 also illustrates that TB changes over time, but the median alone as a summary measure does not capture this. The description of Trial 2 below shows an example in which the RTB is significant and in addition it depends on age.

**Absolute treatment benefit.** In the example from Amado and colleagues, the ATBs are 9% (= 93–65–88+69) and –41% (= 8–14–50+15) at 6 and 12 weeks, respectively (Table 2).



**Figure 1.** PFS by randomized treatment in (A) mutant and (B) wild-type KRAS groups. HRs are shown for panitumumab (panit) vs. best supportive care (BSC) adjusted for randomization factors (Eastern Cooperative Oncology Group score, geographic region). Reprinted with permission from ref. 1. © 2008 American Society of Clinical Oncology. All rights reserved. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 2008;26:1626-34.

**Table 2.** Quantifying treatment benefit (TB) in marker-specific subgroups: examples and clinical interpretation

<b>Trial 1: KRAS is predictive of panitumumab benefit in colon patients (data from ref. 1)</b>				
Measure	Marker negative (KRAS mutant)	Marker positive (KRAS wild-type)	Differential treatment (panitumumab) benefit	Interpretation
HR (log scale); PFS	0.99	0.45	Log (0.99/0.45; $P < 0.0001$ )	There is significant interaction of panitumumab benefit over supportive care and KRAS status over time. In the mutant group, the hazard of progressing is similar (0.99 is close to 1) in patients treated with panitumumab vs. supportive care. In the wild-type group, there is a 55% (1-0.45) decrease in the hazard of progressing in patients treated with panitumumab compared with the hazard of progressing in patients treated with supportive care.
At 6 weeks				
Ratio of PFS estimates at 6 weeks	1.43 (93/65)	1.28 (88/69)	1.12 ( <sup>a</sup> )	At 6 weeks, PFS was improved in both mutant and wild-type patients; panitumumab benefit for patients with KRAS mutations is 12% better compared with the benefit in wild-type KRAS patients.
Difference in proportion of patients progression-free at 6 weeks	28% (93%-65%)	19% (88%-69%)	9% ( <sup>a</sup> )	At 6 weeks, PFS was improved in both mutant and wild-type patients; panitumumab benefit is 9% higher in mutant patients than in wild-type KRAS patients.
At 12 wks				
Ratio of PFS estimates at 12 weeks	0.57 (8/14)	3.33 (50/15)	0.17 ( <sup>a</sup> )	At 12 weeks, panitumumab benefit for patients with KRAS mutations is 83% worse relative to panitumumab benefit in patients with wild-type KRAS
Difference in proportion of patients progression free at 12 weeks	-6% (8%-14%)	35% (50%-15%)	-41% ( <sup>a</sup> )	At 12 weeks, panitumumab benefit is 41% smaller in mutant patients than in wild-type KRAS
<b>Trial 2: PR and PFT treatment in breast patients</b>				
<b>PR, progesterone receptor; PFT, L-phenylalanine mustard, 5-fluorouracil, and tamoxifen (data<sup>b</sup> from ref. 8; subset of women ages &lt;50)</b>				
Measure	Marker negative (PR < 10)	Marker positive (PR ≥ 10)	Differential treatment (PFT) benefit	Interpretation
HR (log scale); disease-free-survival (DFS)	NR	NR	-0.52 ( $P = 0.014$ )	The difference between treatment effects in the PR < 10 and ≥ 10 groups on the HR scale is -0.523 units, regardless of the time point. TB varies according to PR status significantly.
DFS ratio (3 y)	0.73	1.07	0.68 ( $P = 0.035$ )	At 3 years, the benefit of PFT in women with PR < 10 is 32% worse relative to the benefit of PFT in women with PR > 10, and it is statistically significant.
Difference in proportion of patients disease-free at 3 years	-16%	5%	-21% ( $P = 0.035$ )	At 3 years, among women with PR < 10, disease free survival associated with PFT (i.e., PFT benefit) is worse than PF benefit by 16%. PFT benefit among women with PR > 10 is 5% greater than PF benefit. The ATB is 21% and there is a statistically significant difference in TB according to PR status.

Abbreviation: NR, not reported.

<sup>a</sup> $P$  values are not provided in examples where raw data were not available to estimate the SE.

<sup>b</sup>Details for the source of estimates of Trial 2 are shown in Supplementary Table S1.

These differences are interpreted as follows: Panitumumab benefit is 9% higher in mutant patients than in wild-type KRAS at 6 weeks; whereas at 12 weeks, panitumumab benefit is 41% smaller in patients with mutant KRAS than in patients with wild-type KRAS.

#### Illustrative example, Trial 2

This example uses published data from the National Surgical Adjuvant Breast and Bowel Project trial (NSABP; ref. 8). This trial evaluated L-phenylalanine mustard, 5-fluorouracil, and tamoxifen (PFT) versus L-phenylalanine mustard and 5-fluoro-

uracil (PF) and showed a statistically significant interaction between progesterone receptor (PR) status (PR < 10 vs. ≥ 10), treatment (PFT vs. PF), and age (<50 vs. ≥50) in relation to 3-year disease-free survival (8, 9). Detailed data are shown in Supplementary Table S1 and comparisons with proposed measures are shown in Table 2. Using the conventional approach and a proportional hazards model, the NSABP trial demonstrated significant evidence for an interaction between treatment and PR status (i.e., differential TB) among premenopausal women (interaction HR on the log scale = -0.523, SE = 0.212,  $P = 0.014$ ; reported in ref. 8).

Among premenopausal women (age < 50), the relative 3-year disease-free survival associated with PFT treatment (relative to PF treatment) was 0.73 (= 0.436/0.599) for the PR < 10 group and 1.07 (= 0.698/0.651) for the PR ≥ 10 group. The estimated RTB at 3 years is 0.68 (= 0.73/1.07, test statistic = -2.106; *P* = 0.035), in the age < 50 group, which is significantly different from 1 supporting that TB varies according to PR status in this age group. In contrast, the RTB at 3 years in the age ≥ 50 group is 0.98 (= 1.21/1.24), which was not significantly different from 1 (test statistic = -0.122; *P* = 0.90). The RTBs at 3 years among postmenopausal women (age ≥ 50) are 1.21 (= 0.639/0.526) and 1.24 (= 0.790/0.639) for the PR < 10 and ≥ 10 groups, respectively.

The differences in disease-free-survival estimates at 3 years between the two treatments were as follows: -0.163 (= 0.436-0.599) and 0.047 (= 0.698-0.651) in the PR < 10 and ≥ 10 groups, respectively, for premenopausal women; and 0.113 (= 0.639-0.526) and 0.151 (= 0.79-0.639) in the PR < 10 and ≥ 10 groups, respectively, for postmenopausal women. The ATB at 3 years was -21% (= -16% - 5%), which is significantly different from 0 (test statistic = -2.102; *P* = 0.035) in the age < 50 group. For women of age ≥ 50, the estimated ATB was -3.8%, which is not significantly different from 0 (test statistic = -0.432; *P* = 0.67). Thus, PR < 10 is predictive of poor 3-year disease-free survival with PFT treatment in women of age < 50. This holds under all our definitions of TB.

For additional examples, including an example of the benefit of combined nivolumab and ipilimumab versus nivolumab alone according to PD-L1 status, refer to Supplementary Data (Supplementary Fig. S2; Supplementary Table S2).

### Choice of time point

Both RTB and ATB quantify differential TB at a time point of interest. The choice of the landmark time point should not be based on the data under investigation. The time point will be different for different cancer types and should be chosen according to clinical considerations in a similar fashion to choosing an endpoint for a given definition of benefit (10). Another approach for choosing the time can be based on a certain proportion of patients surviving or being disease free: for example, the time at which 75% of the patients in the full cohort survive (or the time at which 50% of the patients survive, which is equivalent to the median survival time). The choice of time point also has important statistical implications. There is adequate sample size at earlier time points, while the sample size decreases due to censoring (and earlier failures) at later time points. Hence, we expect the estimated survival probabilities to have better precision and, thus, greater power to detect differential TB at earlier as opposed to later time points. We conducted simulation studies to confirm this (details not shown), and we have summarized the various statistical properties of the proposed measures in the Supplementary Data. In practice, after identifying a clinically meaningful time point, it is useful to understand whether this is an early or late time point in the context of the data at hand, keeping in mind that the time point must not be selected after looking into the data under investigation.

## Discussion

Recent studies have demonstrated that different individuals may respond differently to the same treatment, and biomarkers such as cancer genes play a role in understanding this interindi-

vidual variation in TBs. For example, *RAF* inhibitors are beneficial in patients with *BRAF*-mutated melanoma, but harmful in patients with *BRAF* wild-type tumors (11, 12); anti-PD1 antibodies improve PFS, but the magnitude of improvement differs according to PD-L1 status and possibly disease type (4); *EGFR* inhibitors improve PFS in *KRAS* wild-type, but not mutant, colon cancer patients (1, 2). Such empirical findings have initiated active statistical research to derive a quantitative definition of TB that is clinically meaningful and to develop formal statistical approaches for evaluating the predictive values of biomarkers. In this article, we have examined ways to quantify differential TB across biomarker subgroups for time-to-event endpoints such as PFS/OS. Our proposed measures are based on survival probabilities calculated at a certain time of clinical interest and, hence, are easy to interpret. In oncology, we may observe that the magnitude of treatment effect varies over time. The proposed measures have the advantage of quantifying benefit in molecular subgroups at time horizons that are relevant to a specific disease or patient. The approaches presented here may be used to compare various biomarkers within a specific disease type and could help prioritize research efforts among several biomarker evaluations. They may also be applied to compare the strength of a predictive biomarker across several disease types. There are settings where we know *a priori* that the drug targets specific molecular subgroups (biologic interaction) and settings where we are uncertain whether TB varies according to biomarker status. The proposed measures may be used to evaluate differential TB in each one of these situations.

The interpretation of differential TB depends upon the scale on which TB is measured (13, 14). Different scales can lead to different interpretations, as demonstrated in the PD-L1 example from a melanoma trial (Supplementary Data). While each of our proposed measures has a distinct interpretation, the measure based on survival differences (ATB) has the unique advantage of being interpretable on the probability scale. Furthermore, survival difference provides insights into the actual magnitude of risk reduction, which is more informative for understanding treatment effect with regards to PFS (or OS) endpoint. Although survival ratios are useful, ratios are sensitive to the magnitude of the denominator. Furthermore, a ratio can show considerable proportional increase or decrease in benefit even when the actual difference is small. We do not advocate providing various measures of TB during a dialog with patients; we recommend describing TB to patients in terms of survival probabilities and their differences as opposed to HR. This aligns with epidemiologic studies that have advocated the use of absolute risk scale for conveying benefits of screening and interventions (15-17). Hernan (18) and Spruance and colleagues (19) also recommend using survival estimates and not HR for the reasons we outlined above, mainly the difficulty in interpretation. In addition, Hernan (18) describes how to use adjusted or stratified survival estimates for observational or randomized studies when various stratification factors are available for analysis. The interpretation would be similar to the setting of our work; the measures would indicate which group has higher or lower benefit relative to another after adjusting for covariates. The work by Fagerlin and colleagues (20) discusses the challenges of communicating risk to patients and they recommend, although not strongly, that communication about cancer risk and TB be presented using frequency format.

Once estimates of TB and differential TB according to biomarker status are obtained, each patient may have her or his own conception of how large the benefit of treatment must be in order

to make it worthwhile to participate in a clinical trial or start a new treatment. Further work is needed before we can provide thresholds for what magnitude of TB and differential ought to be considered clinically meaningful and the role of the sample size required. We anticipate this threshold would depend upon disease type and prevalence of the biomarker of interest.

## Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Authors' Contributions

**Conception and design:** A. Iasonos, J.M. Satagopan

**Development of methodology:** A. Iasonos, J.M. Satagopan

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** A. Iasonos, J.M. Satagopan

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** A. Iasonos, P.B. Chapman, J.M. Satagopan

**Writing, review, and/or revision of the manuscript:** A. Iasonos, P.B. Chapman, J.M. Satagopan

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** A. Iasonos, J.M. Satagopan

**Study supervision:** J.M. Satagopan

## Grant Support

A. Iasonos was supported in part by the NIH under award number P30CA008748. P.B. Chapman was supported in part by the John K. Figge Research Fund. J.M. Satagopan was supported in part by the NIH under award numbers P30CA008748, UL1TR000457, and R01CA137420.

Received October 19, 2015; revised February 10, 2016; accepted February 19, 2016; published online May 2, 2016.

## References

- Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol* 2008;26:1626–34.
- Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 2008;359:1757–65.
- Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 2012;366:2443–54.
- Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015;373:23–34.
- Motzer RJ, Rini BI, McDermott DF, Redman BG, Kuzel TM, Harrison MR, et al. Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. *J Clin Oncol* 2015;33:1430–7.
- Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;32:2380–5.
- O'Quigley J. Proportional hazards regression. New York: Springer Science + Business Media; 2008.
- Fisher B, Redmond C, Brown A, Wickerham DL, Wolmark N, Allegra J, et al. Influence of tumor estrogen and progesterone receptor levels on the response to tamoxifen and chemotherapy in primary breast cancer. *J Clin Oncol* 1983;1:227–41.
- Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985;41:361–72.
- Wu W, Sargent D. Choice of endpoints in cancer clinical trials. In: Kelly WK, Halabi S, editors. *Oncology clinical trials: successful design, conduct, and analysis*. New York, NY: Demos Medical Publishing; 2010. p.35–42.
- Callahan MK, Rampal R, Harding JJ, Klimek VM, Chung YR, Merghoub T, et al. Progression of RAS-mutant leukemia during RAF inhibitor treatment. *N Engl J Med* 2012;367:2316–21.
- Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* 2010;468:973–7.
- Wang X, Elston RC, Zhu X. The meaning of interaction. *Hum Hered* 2010;70:269–77.
- Wang X, Elston RC, Zhu X. Statistical interaction in human genetics: how should we model it if we are looking for biological interaction? *Nat Rev Genet* 2011;12:74.
- Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. *Stat Med* 2011;30:1090–104.
- Helzlsouer KJ. The numbers game: the risky business of projecting risk. *J Natl Cancer Inst* 2011;103:992–3.
- Petracci E, Decarli A, Schairer C, Pfeiffer RM, Pee D, Masala G, et al. Risk factor modification and projections of absolute breast cancer risk. *J Natl Cancer Inst* 2011;103:1037–48.
- Hernan MA. The hazards of hazard ratios. *Epidemiology* 2010;21:13–5.
- Spruance S, Reid J, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004;48:2787–92.
- Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst* 2011;103:1436–43.

# Clinical Cancer Research

## Quantifying Treatment Benefit in Molecular Subgroups to Assess a Predictive Biomarker

Alexia Iasonos, Paul B. Chapman and Jaya M. Satagopan

*Clin Cancer Res* 2016;22:2114-2120.

<b>Updated version</b>	Access the most recent version of this article at: <a href="http://clincancerres.aacrjournals.org/content/22/9/2114">http://clincancerres.aacrjournals.org/content/22/9/2114</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://clincancerres.aacrjournals.org/content/suppl/2016/05/24/22.9.2114.DC2">http://clincancerres.aacrjournals.org/content/suppl/2016/05/24/22.9.2114.DC2</a>

<b>Cited articles</b>	This article cites 18 articles, 5 of which you can access for free at: <a href="http://clincancerres.aacrjournals.org/content/22/9/2114.full#ref-list-1">http://clincancerres.aacrjournals.org/content/22/9/2114.full#ref-list-1</a>
-----------------------	---

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, use this link <a href="http://clincancerres.aacrjournals.org/content/22/9/2114">http://clincancerres.aacrjournals.org/content/22/9/2114</a> . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.