



Original Investigation | Oncology

# Correlation Between Surrogate End Points and Overall Survival in a Multi-institutional Clinicogenomic Cohort of Patients With Non-Small Cell Lung or Colorectal Cancer

Kenneth L. Kehl, MD, MPH; Gregory J. Riely, MD, PhD; Eva M. Lepisto, MA, MSc; Jessica A. Lavery, MS; Jeremy L. Warner, MD, MS; Michele L. LeNoue-Newton, PhD; Shawn M. Sweeney, PhD; Julia E. Rudolph, MPA; Samantha Brown, MS; Celeste Yu, MS; Philippe L. Bedard, MD; Deborah Schrag, MD, MPH; Katherine S. Panageas, DrPH; for the American Association of Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange (GENIE) Consortium

## Abstract

**IMPORTANCE** Contemporary observational cancer research requires associating genomic biomarkers with reproducible end points; overall survival (OS) is a key end point, but interpretation can be challenging when multiple lines of therapy and prolonged survival are common. Progression-free survival (PFS), time to treatment discontinuation (TTD), and time to next treatment (TTNT) are alternative end points, but their utility as surrogates for OS in real-world clinicogenomic data sets has not been well characterized.

**OBJECTIVE** To measure correlations between candidate surrogate end points and OS in a multi-institutional clinicogenomic data set.

**DESIGN, SETTING, AND PARTICIPANTS** A retrospective cohort study was conducted of patients with non-small cell lung cancer (NSCLC) or colorectal cancer (CRC) whose tumors were genotyped at 4 academic centers from January 1, 2014, to December 31, 2017, and who initiated systemic therapy for advanced disease. Patients were followed up through August 31, 2020 (NSCLC), and October 31, 2020 (CRC). Statistical analyses were conducted on January 5, 2021.

**EXPOSURES** Candidate surrogate end points included TTD; TTNT; PFS based on imaging reports only; PFS based on medical oncologist ascertainment only; PFS based on either imaging or medical oncologist ascertainment, whichever came first; and PFS defined by a requirement that both imaging and medical oncologist ascertainment have indicated progression.

**MAIN OUTCOMES AND MEASURES** The primary outcome was the correlation between candidate surrogate end points and OS.

**RESULTS** There were 1161 patients with NSCLC (648 women [55.8%]; mean [SD] age, 63 [11] years) and 1150 with CRC (647 men [56.3%]; mean [SD] age, 54 [12] years) identified for analysis. Progression-free survival based on both imaging and medical oncologist documentation was most correlated with OS (NSCLC:  $\rho = 0.76$ ; 95% CI, 0.73-0.79; CRC:  $\rho = 0.73$ ; 95% CI, 0.69-0.75). Time to treatment discontinuation was least associated with OS (NSCLC:  $\rho = 0.45$ ; 95% CI, 0.40-0.50; CRC:  $\rho = 0.13$ ; 95% CI, 0.06-0.19). Time to next treatment was modestly associated with OS (NSCLC:  $\rho = 0.60$ ; 0.55-0.64; CRC:  $\rho = 0.39$ ; 95% CI, 0.32-0.46).

**CONCLUSIONS AND RELEVANCE** This cohort study suggests that PFS based on both a radiologist and a treating oncologist determining that a progression event has occurred was the surrogate end point most highly correlated with OS for analysis of observational clinicogenomic data.

JAMA Network Open. 2021;4(7):e2117547. doi:10.1001/jamanetworkopen.2021.17547

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

## Key Points

**Question** What surrogate end point for capturing worsening disease is most correlated with overall survival (OS) in large linked clinicogenomic data sets?

**Findings** In this cohort study of patients with non-small cell lung cancer or colorectal cancer who initiated systemic therapy for advanced disease, progression-free survival based on both radiologist and medical oncologist assessment was more consistently correlated with OS than other candidate end points, including time to treatment discontinuation and time to next treatment.

**Meaning** This study suggests that, based on its correlation with OS, progression-free survival based on both radiologist and medical oncologist assessment may be an optimal surrogate end point for analysis of observational clinicogenomic data in cancer research.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

## Introduction

For patients with cancer enrolled in clinical trials, the Response Evaluation Criteria in Solid Tumors, version 1.1,<sup>1</sup> are applied to ascertain treatment response and disease progression. However, outside of clinical trials, clinical outcomes are typically recorded only in the unstructured text of radiology reports and clinician progress notes.<sup>2</sup> This can present challenges to the reproducibility of cancer research that incorporates large quantities of molecular and genomic data now routinely generated across institutions.<sup>3</sup>

Although overall survival (OS) constitutes a key outcome in cancer research, end points such as progression-free survival (PFS; time to progression or death), time to treatment discontinuation (TTD), and time to next treatment (TTNT) may also be relevant in some contexts. Progression can be assessed earlier than mortality, and PFS is less associated with subsequent lines of therapy, which can confound ascertainment of the association of any individual treatment with OS. However, the process of extracting PFS outcomes from electronic health record (EHR) data at scale is resource intensive, requiring review of thousands of clinical documents. Measurement of PFS in observational contexts is also not standardized regarding the definition of cancer progression. In contrast, TTD, defined as the time from initiation of a systemic therapy regimen to the date of treatment discontinuation or death, can be rapidly extracted from structured pharmacy data and has therefore been proposed as an alternative end point in the observational setting.<sup>4,5</sup> Time to next treatment, defined as time from initiation of a systemic therapy to the date of initiation of the first subsequent systemic therapy regimen<sup>6-8</sup> or death, can similarly be extracted from pharmacy data.

The utility of end points such as PFS, TTD, or TTNT, particularly when they are applied in lieu of, or as a surrogate for, OS, may be evaluated by measuring the correlation between the end point and OS.<sup>9</sup> For example, disease-free survival has become an accepted surrogate end point in adjuvant therapy clinical trials for colorectal cancer (CRC) owing to its high correlation with OS.<sup>10,11</sup>

Understanding the correlation between PFS measures and OS, and other pragmatic end points (TTD or TTNT) and OS, specifically in the observational context may have increasing implications given the growing role of real-world evidence for regulatory purposes.<sup>12</sup> A structured framework is necessary to define progression outcomes using such real-world data. The PRISSMM (pathology, radiology and imaging, documentation of signs and symptoms, medical oncologist notes, and tumor markers) framework consists of directives for abstracting clinical outcomes from individual components of the medical record.<sup>13</sup> This system fosters transparency and reproducibility by defining outcome measures with specific attention to data provenance. To specifically inform research derived from linked clinical and genomic data, this study focused on a multi-institutional cohort of patients with advanced non-small cell lung cancer (NSCLC) or CRC whose tumor sequencing results were submitted to the American Association for Cancer Research's Project GENIE (Genomics Evidence Neoplasia Information Exchange).<sup>3</sup> The specific objective of this analysis was to report correlations between OS and TTD, TTNT, and PFS, as systematically ascertained from the EHR using the PRISSMM framework for medical record curation.

## Methods

### Cohort

The cohort for this analysis included patients with stage I to stage IV NSCLC or CRC whose tumors underwent genomic sequencing at Dana-Farber Cancer Institute, Memorial Sloan Kettering Cancer Center, Princess Margaret Cancer Centre, or the Vanderbilt-Ingram Cancer Center between January 1, 2014, and December 31, 2017. Patients consented to medical record review and genomic profiling of their tumor tissue at each institution; this supplemental retrospective analysis was approved by the Dana-Farber/Harvard Cancer Center institutional review board under a waiver of informed consent because this study presented no more than minimal risk to the participants. Outcomes were analyzed for patients who were either diagnosed initially with stage IV NSCLC or CRC and received at

least 1 systemic therapy regimen, or who initially received a diagnosis of stage I to stage III NSCLC or CRC and received at least 1 systemic therapy regimen that began at least 6 months after initial diagnosis (assumed to represent treatment for recurrent disease). Patients were followed up through August 31, 2020 (NSCLC), and October 31, 2020 (CRC). This study is reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.<sup>14</sup>

### Medical Record Curation

Curation of imaging reports and medical oncologist notes was performed according to the PRISMM framework.<sup>13</sup> For each imaging report and medical oncologist assessment, curators were asked to record whether the radiologist or clinician described the presence of cancer, and if so, whether the cancer was improving and responding, stable, mixed, or worsening or progressing. Curators reviewed the text of radiologists' reports for computed tomography, magnetic resonance imaging, positron emission tomography, and nuclear medicine evaluations. Curators reviewed the first note per month from a medical oncologist; if one was not available, a note from an advanced practice clinician (nurse practitioner or physician assistant) was reviewed.

### Outcomes

#### Time to Treatment Discontinuation

The index date for TTD was defined as the start date of the first systemic therapy regimen, consisting of a drug or group of drugs for which there was documentation from the treating oncologist of a plan for simultaneous administration for recurrent or metastatic disease. Dates of initiation and final administration of each drug in infusional regimens were curated. For NSCLC, the most common first-line infusional regimens were given every 3 weeks, so to more closely capture the point when a decision was reached to discontinue the regimen, the end date of infusional regimens was defined for this analysis as 3 weeks after the last drug in the regimen was administered.<sup>15</sup> For CRC, the most common first-line infusional regimens were given every 2 weeks, so the end date for infusional regimens was defined as 2 weeks after the last drug in the regimen was administered. For oral therapy, the end date was defined as the date on which the prescription expired or the medical oncologist documented discontinuation of the regimen, whichever came first. In primary analyses, death also constituted a treatment discontinuation event. Censoring was performed on the date patients were last known alive and receiving treatment.

#### Time to Next Treatment

In primary analyses, TTNT was defined as time from first-line treatment start to initiation of subsequent systemic therapy or death. Censoring was performed at the date patients were last known alive and free of subsequent therapy.

#### Progression-Free Survival

Four definitions of PRISMM-derived PFS outcomes were evaluated: PFS-imaging (PFS-I; time to first worsening or progression documented in imaging report, or death), PFS-medical oncologist (PFS-M; time to first worsening or progression documented in medical oncologist assessment, or death), PFS-I-or-M (time to first indication of worsening or progression in imaging report or medical oncologist assessment, or death, whichever was earliest), and PFS-I-and-M (time from treatment start to worsening or progression having been documented in both an imaging report and a medical oncologist assessment, or death). The index date for PFS was defined as the start date of first-line therapy for recurrent or metastatic disease. Patients were censored on the date last known alive and free of disease progression.

## Statistical Analysis

### End Point Correlations

Analyses were conducted on January 5, 2021. Overall survival, TTD, TTNT, and PFS measures were estimated using the Kaplan-Meier method. Correlations and 95% CIs between OS and (1) TTD, (2) TTNT, or (3) each PRISMM-derived PFS outcome were measured using normal scores rank correlation, calculated using the iterative multiple imputation approach for analysis of correlations between 2 partially censored failure times.<sup>16</sup> Among patients with NSCLC, correlations between OS and alternative end points were further explored by systemic therapy regimen category, including (1) all regimens, (2) cytotoxic chemotherapy only (with or without an anti-vascular endothelial growth factor agent), (3) checkpoint inhibitor immunotherapy only, or (4) oral targeted therapy. Among patients with CRC, only the analysis of all regimens was performed, because most regimens in that cohort were chemotherapy based.

In sensitivity analyses, correlations were recalculated after restricting to patients who underwent genomic testing prior to starting first-line therapy. Because all patients in this cohort had genomic testing as an inclusion criterion, this procedure restricted calculations to the time during which patients were at risk for death, and it was performed in lieu of left truncation, which the statistical package used for correlation calculations could not incorporate. In a second set of sensitivity analyses, patients who died without experiencing disease progression or a treatment discontinuation or change event were excluded from the denominator to assess the extent to which correlations between candidate outcomes and OS were owing to mortality rather than progression or treatment discontinuation events. Calculations were performed using R, version 3.6.1 (R Group for Statistical Computing) and the SurvCorr R package, version 1.0.<sup>17</sup> All *P* values were from 2-sided tests and results were deemed statistically significant at *P* < .05.

## Results

### Cohorts

There were 1161 patients with NSCLC (648 women [55.8%]; mean [SD] age, 63 [11] years) and 1150 patients with CRC (647 men [56.3%]; mean [SD] age, 54 [12] years) (eFigures 1 and 2 in Supplement 1). Additional patient characteristics are provided in Table 1.

### Patient-Level End Point Correlations After First-Line Therapy

#### Non-Small Cell Lung Cancer

The median OS after initiation of any first-line therapy was 28.9 months (interquartile range [IQR], 12.0-66.6 months) (Table 2); in a sensitivity analysis restricted to 375 patients starting therapy after their genomic testing report, median OS was 22.8 months (IQR, 9.5-55.4 months) (eTable 1 in Supplement 1). Among all 1161 patients, TTD yielded the shortest median time to event, at 3.6 months (IQR, 1.6-8.5 months), while PFS-I-and-M yielded the longest median time to event at 9.6 months (IQR, 4.5-20.8 months) (Table 2).

Time to treatment discontinuation had the weakest correlation with OS ( $\rho = 0.45$ ; 95% CI, 0.40-0.50) and PFS-I-and-M had the strongest correlation ( $\rho = 0.76$ ; 95% CI, 0.73-0.79) (Table 2). Time to next treatment was modestly associated with OS ( $\rho = 0.60$ ; 0.55-0.64). A similar pattern held when the analysis was restricted to the 375 patients initiating therapy after genomic testing; eTable 1 in Supplement 1. Kaplan-Meier curves for OS, TTD, TTNT, and each variant of PFS are provided in Figure 1. The correlation between the 2 individual PFS measures themselves (PFS-I and PFS-M) was 0.69 (95% CI, 0.65-0.72).

The most common treatment regimens administered in this cohort are listed in eTable 2 in Supplement 1. Among 683 patients whose first-line therapy was cytotoxic chemotherapy only, TTD was poorly correlated with OS ( $\rho = 0.34$ ; 95% CI, 0.26-0.41), as was TTNT ( $\rho = 0.53$ ; 95% CI, 0.46-0.60) (Table 2). However, these end points were more correlated with OS among 124 patients who began first-line immunotherapy only (TTD:  $\rho = 0.76$ ; 95% CI, 0.64-0.84; TTNT:  $\rho = 0.70$ ; 95%

CI, 0.57-0.80). Among patients receiving cytotoxic chemotherapy only and those receiving immunotherapy only, PFS-I-and-M was similarly correlated with OS (chemotherapy:  $\rho = 0.76$ ; 95% CI, 0.72-0.79; immunotherapy:  $\rho = 0.78$ ; 95% CI, 0.65-0.86).

In a second sensitivity analysis in which patients who died before experiencing another outcome event were excluded from correlation calculations, death in the absence of progression within this cohort was uncommon, even for PFS-I-and-M, which by definition has the longest time to event among the candidate PFS measures (ie, 23 of 1161 patients [2.0%]) and for TTNT, which could be impacted by high rates of death without receiving a subsequent line of therapy (25 of 1161 patients [2.2%]). Correlation coefficients were similar to the primary analysis (eTable 3 in Supplement 1).

**Colorectal Cancer**

The median OS after initiation of any first-line therapy for patients with CRC was 42.0 months (IQR, 22.8-83.8 months) (Table 2); in a sensitivity analysis restricted to 160 patients starting therapy after their genomic testing report to remove time when patients were not at risk for death, it was not

**Table 1. Cohort Characteristics**

Characteristic	Patients, No. (%)	
	NSCLC (n = 1161)	CRC (n = 1150)
Histologic characteristics		
NSCLC		
Nonsquamous	1009 (86.9)	NA
Squamous	98 (8.4)	NA
Not otherwise specified	54 (4.7)	NA
Molecular alterations		
Alteration		
<i>KRAS</i>	316 (27.2)	518 (45.0)
<i>NRAS</i>	NA	62 (5.4)
<i>EGFR</i>	306 (26.4)	NA
<i>BRAF</i>	57 (4.9)	116 (10.1)
<i>MET</i>	56 (4.8)	NA
Rearrangement		
<i>ALK</i>	43 (3.7)	NA
<i>ROS1</i>	21 (1.8)	NA
<i>RET</i>	13 (1.1)	NA
None of the above	482 (34.6)	482 (41.9)
Sex		
Male	513 (44.2)	647 (56.3)
Female	648 (55.8)	503 (43.7)
Age, y		
<40	32 (2.8)	143 (12.4)
40-49	104 (9.0)	274 (23.8)
50-59	293 (25.2)	340 (29.6)
60-69	401 (34.5)	281 (24.4)
70-79	282 (24.3)	88 (7.7)
≥80	49 (4.2)	24 (2.1)
Stage at diagnosis		
I-III	404 (34.8)	461 (40.1)
IV	757 (65.2)	689 (59.9)
Treating institution		
Dana-Farber Cancer Institute	409 (35.2)	390 (33.9)
Memorial Sloan Kettering Cancer Center	552 (47.5)	558 (48.5)
Princess Margaret Cancer Centre	41 (3.5)	NA
Vanderbilt University Medical Center	159 (13.7)	202 (17.6)

Abbreviations: CRC, colorectal cancer; NA, not applicable; NSCLC, non-small cell lung cancer.

reached (IQR, 18.2 months to not reached) (eTable 1 in Supplement 1). Among all 1150 patients, TTD yielded the shortest median time to event, at 4.3 months (IQR, 2.3-6.5 months), while PFS-I-and-M yielded the longest median time to event at 14.4 months (IQR, 8.6-28.9 months) (Table 2).

Patterns of correlation between end points for CRC were similar to those in the NSCLC cohort. For CRC, TTD had the weakest correlation with OS ( $\rho = 0.13$ ; 95% CI, 0.06-0.19) and PFS-I-and-M had the strongest correlation ( $\rho = 0.73$ ; 95% CI, 0.69-0.75) (Table 2). Time to next treatment was

**Table 2. Correlations Between Clinical End Points and OS by Therapy Regimen Category**

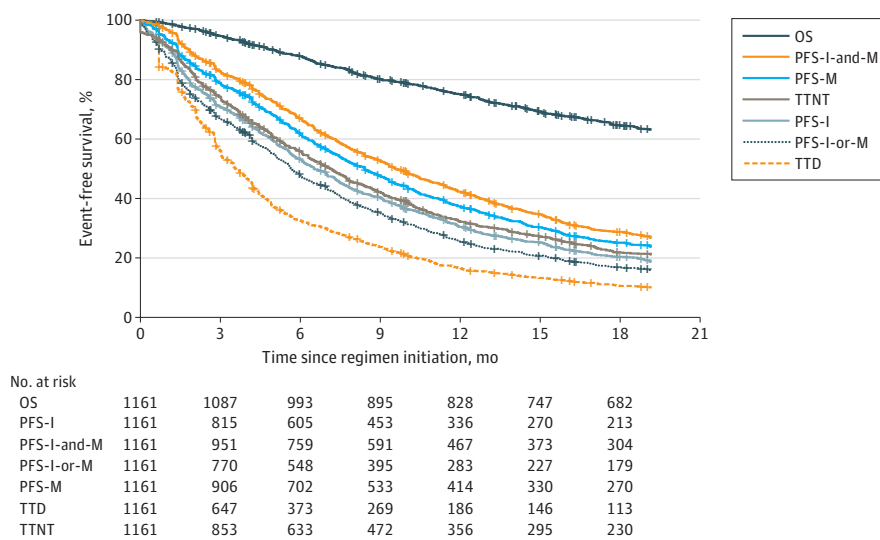
Characteristic	Time to event, median (IQR), mo	$\rho$ (95% CI)
<b>NSCLC: all regimens</b>		
No.	1161	NA
OS	28.9 (12.0-66.6)	1 [Reference]
TTD	3.6 (1.6-8.5)	0.45 (0.40-0.50)
TTNT	7.1 (2.8-16.2)	0.60 (0.55-0.64)
PFS-I	6.5 (2.5-15.1)	0.63 (0.59-0.67)
PFS-M	8.4 (3.8-18.2)	0.71 (0.67-0.74)
PFS-I-or-M	5.6 (1.9-12.3)	0.61 (0.56-0.65)
PFS-I-and-M	9.6 (4.5-20.8)	0.76 (0.73-0.79)
<b>CRC: all regimens</b>		
No.	1150	NA
OS	42.0 (22.8-83.8)	1 [Reference]
TTD	4.3 (2.3-6.5)	0.13 (0.06-0.19)
TTNT	7.2 (3.9-6.5)	0.39 (0.32-0.46)
PFS-I	10.5 (5.6-19.1)	0.66 (0.61-0.70)
PFS-M	13.5 (7.7-26.3)	0.71 (0.67-0.74)
PFS-I-or-M	10.0 (5.1-17.4)	0.65 (0.61-0.69)
PFS-I-and-M	14.4 (8.6-28.9)	0.73 (0.69-0.75)
<b>NSCLC: cytotoxic chemotherapy only</b>		
No.	683	NA
OS	27.3 (11.0-66.5)	1 [Reference]
TTD	3.0 (1.9-5.5)	0.34 (0.26-0.41)
TTNT	6.1 (2.4-13.4)	0.53 (0.46-0.60)
PFS-I	6.9 (2.8-15.4)	0.66 (0.61-0.70)
PFS-M	8.1 (4.0-17.1)	0.70 (0.66-0.74)
PFS-I-or-M	5.9 (2.3-12.3)	0.63 (0.57-0.67)
PFS-I-and-M	9.3 (4.6-21.0)	0.76 (0.72-0.79)
<b>NSCLC: immunotherapy only</b>		
No.	124	NA
OS	18.6 (7.1-43.9)	1 [Reference]
TTD	4.1 (1.6-8.9)	0.76 (0.64-0.84)
TTNT	5.2 (2.1-12.8)	0.70 (0.57-0.80)
PFS-I	2.6 (1.4-10.2)	0.63 (0.46-0.75)
PFS-M	2.0 (1.4-13.7)	0.69 (0.52-0.80)
PFS-I-or-M	1.9 (0.9-6.8)	0.59 (0.43-0.72)
PFS-I-and-M	4.5 (1.7-14.5)	0.78 (0.65-0.86)
<b>NSCLC: oral targeted therapy</b>		
No.	207	NA
OS	47.0 (22.8-91.4)	1 [Reference]
TTD	12.0 (7.4-26.0)	0.63 (0.50-0.72)
TTNT	12.6 (7.6-26.9)	0.62 (0.51-0.71)
PFS-I	9.7 (5.1-19.6)	0.52 (0.38-0.64)
PFS-M	12.7 (7.1-27.4)	0.63 (0.51-0.73)
PFS-I-or-M	8.5 (4.7-19.1)	0.50 (0.36-0.62)
PFS-I-and-M	14.4 (7.7-30.8)	0.71 (0.59-0.80)

Abbreviations: CRC, colorectal cancer; IQR, interquartile range; NA, not applicable; NSCLC, non-small cell lung cancer; OS, overall survival; PFS-I, PFS based on imaging reports only; PFS-I-and-M, PFS defined by a requirement that both imaging and medical oncologist ascertainment have indicated progression; PFS-I-or-M, PFS based on either imaging or medical oncologist ascertainment, whichever came first; PFS-M, PFS based on medical oncologist ascertainment only; TTD, time to treatment discontinuation (or death); TTNT, time to next treatment (or death).

modestly associated with OS ( $\rho = 0.39$ ; 95% CI, 0.32-0.46). A similar pattern held when the analysis was restricted to the 160 patients initiating therapy after genomic testing (eTable 1 in Supplement 1). Kaplan-Meier curves for OS, TTD, TTNT, and each variant of PFS are provided in Figure 2. The correlation between the 2 individual PFS measures (PFS-I and PFS-M) was 0.73 (95% CI, 0.70-0.75).

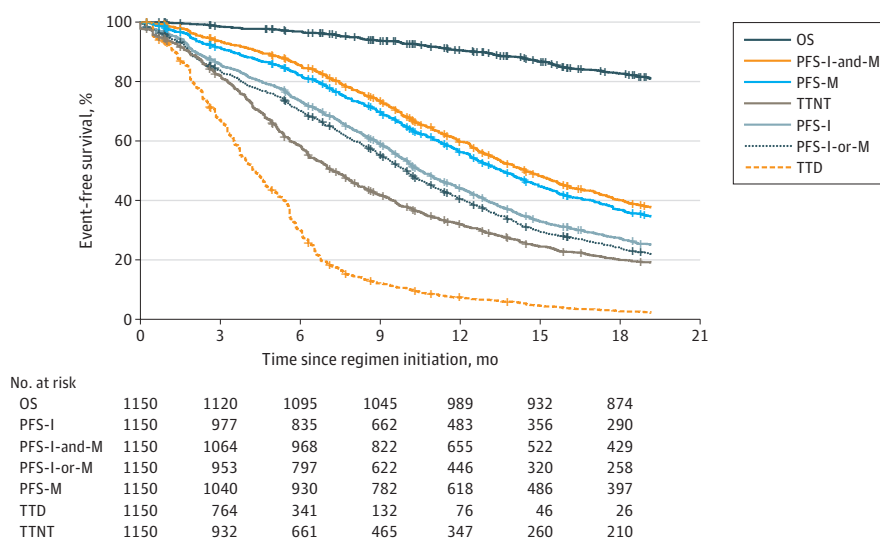
In a second sensitivity analysis in which patients who died before experiencing another outcome event were excluded from correlation calculations, death in the absence of progression remained uncommon, even for PFS-I-and-M (ie, 8 of 1150 patients [0.7%]) and for TTNT (7 of 1150 patients [0.6%]). Correlation coefficients were similar to the primary analysis (eTable 3 in Supplement 1).

Figure 1. Kaplan-Meier Curves for Candidate Outcome Measures After First-Line Systemic Therapy for Recurrent or Metastatic Non-Small Cell Lung Cancer



OS indicates overall survival; PFS, progression-free survival; PFS-I, PFS based on imaging reports only; PFS-I-and-M, PFS defined by a requirement that both imaging and medical oncologist ascertainment have indicated progression; PFS-I-or-M, PFS based on either imaging or medical oncologist ascertainment, whichever came first; PFS-M, PFS based on medical oncologist ascertainment only; TTD, time to treatment discontinuation (or death); and TTNT, time to next treatment (or death).

Figure 2. Kaplan-Meier Curves for Candidate Outcome Measures After First-Line Systemic Therapy for Recurrent or Metastatic Colorectal Cancer



OS indicates overall survival; PFS, progression-free survival; PFS-I, PFS based on imaging reports only; PFS-I-and-M, PFS defined by a requirement that both imaging and medical oncologist ascertainment have indicated progression; PFS-I-or-M, PFS based on either imaging or medical oncologist ascertainment, whichever came first; PFS-M, PFS based on medical oncologist ascertainment only; TTD, time to treatment discontinuation (or death); and TTNT, time to next treatment (or death).

## Discussion

In this multi-institutional analysis of candidate clinical end points among patients undergoing systemic therapy for advanced, genomically characterized NSCLC or CRC, TTD was poorly correlated with OS and TTNT was modestly correlated with OS. Progression-free survival estimated based on abstraction of both imaging and medical oncologist notes (PFS-I-and-M) was most consistently correlated with OS. The magnitude of these patient-level correlations between alternative end points and OS was similar to that observed in analyses of clinical trial data<sup>4</sup> and an observational cohort<sup>8,18</sup> of patients with NSCLC, although the examination of outcomes derived from specific components of the health record across cancer types was a novel feature of the present study. Surrogate outcomes may be useful in observational data sets even when OS data are available, particularly when researchers study a specific line of therapy in contexts in which survival through multiple lines of therapy is common. The consistency of the correlation between PFS-I-and-M and OS in this analysis implies that when a surrogate outcome in clinicogenomic analyses for patients with NSCLC or CRC is required, PFS-I-and-M may be an optimal way to define real-world PFS.

Although this analysis focused on a clinicogenomic data set, it has implications for analysis of real-world evidence in observational cancer research in general. Time to treatment discontinuation and TTNT are attractive end points because they can often be measured with minimal manual review of medical records using structured pharmacy records. In contrast, PFS-I and PFS-M require review of radiology reports and oncologist notes to identify evidence of progression. These results suggest that, in some contexts, a consistent framework for abstraction of such clinical end points, with attention to data provenance<sup>13</sup>—whether progression is documented on imaging results, by a clinician, or both—may be required for observational cancer research, rather than relying on TTD or TTNT. This is a substantial challenge because the careful manual abstraction of EHR data is very resource intensive. One solution is to develop validated natural language processing methods for extracting these end points from the EHR<sup>19</sup>; the feasibility of such approaches for both imaging reports<sup>20,21</sup> and clinician notes<sup>22,23</sup> has been previously demonstrated.

Correlations between surrogate end points and OS likely varied by treatment modality owing to several clinical factors. For example, planned discontinuation of therapy after a given number of cycles, prior to clinical progression, could partially account for a poor correlation between TTD and OS among patients receiving chemotherapy. An extreme example of this dynamic would include therapy delivered once, without immediate plans for regular administration of the drug, as might occur when floxuridine is used as local therapy for liver metastases in patients with CRC.<sup>24</sup> In addition, discontinuation of treatment for toxic effects may further diminish the correlation between TTD and OS if toxic effects are less associated with mortality risk than is progressive disease. This dynamic may explain the modest correlation between TTNT and OS as well because treatment discontinued for toxic effects may be followed by initiation of other treatment prior to progression. On the other hand, for patients with NSCLC, TTD and OS may have been correlated for patients receiving immunotherapy in this cohort if single-agent checkpoint inhibitor treatment had a low rate of severe adverse events<sup>18,25</sup> and therefore less early discontinuation. In this analysis, the patient-level correlation between TTD and OS among patients receiving oral targeted therapy for NSCLC was higher than that among patients receiving chemotherapy, again potentially owing to lower severe adverse event rates with targeted therapy,<sup>26,27</sup> but PFS-I-and-M remained numerically most correlated with OS after oral targeted therapy as well.

## Strengths and Limitations

This study has some strengths, including its multi-institutional cohort of patients with 2 types of common solid tumors who had linked clinical and genomic data. In such data sets, thorough evaluation of clinical end points is particularly critical, because the data are rich enough to enable researchers to ask a wide variety of questions about the association between genomic markers, treatment exposures, and outcomes.<sup>13</sup>



This study also has some limitations, including forms of selection bias. Contributing institutions were 4 North American academic centers, which may not be representative of institutions where patients receive care in the community or in other parts of the world. These patients were also selected for tumor genomic profiling, such that, particularly for NSCLC, those whose tumors may be less likely to harbor targetable mutations may have been underrepresented.

Patients do not always receive all their care within 1 institution or health system. Data on outcome events ascertained outside the primary academic centers for this analysis could potentially have been less complete than data for internally ascertained outcomes. In addition, patients who are included in a clinicogenomic cohort, by definition, very rarely have tumor genomic profiling performed after death. This can introduce lead time between diagnosis and genomic sequencing during which patients included in the data set could not have died, inflating survival estimates. Applying methods for handling left truncation—entering patients into the risk set of a time-to-event analysis only after they have undergone genomic profiling<sup>28</sup>—is a solution to this problem, but measuring correlations between left-truncated time-to-event outcomes requires the development of novel statistical methods. This is especially a challenge if tumor genotyping is performed selectively at the point of clinical progression, which can result in informative left truncation, or temporal selection bias, in which patients selectively enter the cohort specifically because they are at increasing risk of a poor outcome.<sup>29</sup> Nevertheless, these results were similar in sensitivity analyses that were restricted to patients receiving treatments initiated after genomic testing, such that all patients were in the risk set at the index time point. In addition, in observational data sets, the frequency of clinical end point ascertainment is based on the judgment of a patient's clinician and may itself be associated with clinical risk over time; the association of this phenomenon with outcome estimates requires further study.

Despite the relatively large size of the cohorts analyzed in this study, there were overlapping 95% CIs for some comparisons, particularly between PFS-I-and-M and PFS-M for both NSCLC and CRC, and by category of systemic therapy, in which each category had a more limited sample size. An even larger study would be needed to formally compare the correlations between these 2 particular end points and OS and evaluate the generalizability of all observed correlations to additional cancer types and categories of treatment, because the validity of surrogate end points may be specific to disease contexts<sup>30</sup> and clinical questions.<sup>31</sup> More broadly, substantial literature exists on approaches to evaluating the validity of surrogate end points in the context of clinical trials,<sup>31,32</sup> focused largely on measuring correlations between treatment effects. Many observational research questions, particularly in the clinicogenomic context, are focused less on treatment effect comparisons than on biomarker associations. More robust frameworks for evaluating surrogate end points in the observational context<sup>33</sup> are needed.

---

## Conclusions

In this observational cohort of patients with genomically profiled advanced NSCLC or CRC, TTD was inconsistently correlated, and TTNT moderately correlated, with OS on a per-patient basis. Progression-free survival based on review of both imaging reports and medical oncologist notes was most correlated with OS among patients with either type of cancer. Although TTD and TTNT are straightforward to calculate owing to the availability of structured pharmacy data in the EHR, these results indicate that researchers should be cautious about applying such end points if a surrogate outcome consistently correlated with OS is required. In such contexts, PFS-I-and-M may be an optimal choice.

## ARTICLE INFORMATION

**Accepted for Publication:** May 5, 2021.

**Published:** July 26, 2021. doi:10.1001/jamanetworkopen.2021.17547

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2021 Kehl KL et al. *JAMA Network Open*.

**Corresponding Author:** Kenneth L. Kehl, MD, MPH, Department of Medical Oncology, Division of Population Sciences, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Ave, Boston, MA 02215 ([kenneth.kehl@dfci.harvard.edu](mailto:kenneth.kehl@dfci.harvard.edu)).

**Author Affiliations:** Department of Medical Oncology, Division of Population Sciences, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts (Kehl, Lepisto, Schrag); Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York (Riely, Lavery, Rudolph, Brown, Panageas); Department of Medicine, Division of Hematology/Oncology, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee (Warner); Department of Biomedical Informatics, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee (Warner); Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee (LeNoue-Newton); American Association for Cancer Research, Philadelphia, Pennsylvania (Sweeney); Division of Medical Oncology & Hematology, Princess Margaret Cancer Centre/University Health Network, Toronto, Ontario, Canada (Yu, Bedard); Department of Medicine, University of Toronto, Toronto, Ontario, Canada (Bedard); Associate Editor, *JAMA* (Schrag).

**Author Contributions:** Dr Kehl had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Kehl, Riely, Schrag, Panageas.

*Acquisition, analysis, or interpretation of data:* All authors.

*Drafting of the manuscript:* Kehl, Lepisto, Warner, Brown, Schrag.

*Critical revision of the manuscript for important intellectual content:* Kehl, Riely, Lavery, Warner, LeNoue-Newton, Sweeney, Rudolph, Brown, Yu, Bedard, Schrag, Panageas.

*Statistical analysis:* Kehl, Lavery, Brown, Schrag, Panageas.

*Obtained funding:* Kehl, Riely, Sweeney.

*Administrative, technical, or material support:* Riely, Lepisto, Rudolph, Bedard, Schrag.

*Supervision:* Panageas.

**Conflict of Interest Disclosures:** Dr Kehl reported receiving grants from Wong Family Foundation, American Association of Cancer Research (AACR), Simeon J. Fortin Foundation, Doris Duke Foundation, and the National Cancer Institute during the conduct of the study; grants from IBM; and personal fees from Aetion and Roche outside the submitted work. Dr Riely reported receiving grants from AACR during the conduct of the study; grants from Novartis, Pfizer, Merck, Mirati, Takeda, and Roche; and nonfinancial support from Pfizer and Merck outside the submitted work. Ms Lepisto reported receiving grants from AACR during the conduct of the study. Ms Lavery reported receiving grants from AACR during the conduct of the study. Dr Warner reported receiving grants from AACR during the conduct of the study; grants from the National Institutes of Health; consulting fees from Westat and IBM; and having equity in HemOnc.org LLC. Dr LeNoue-Newton reported receiving research funding from AACR during the conduct of the study; and research funding from GE Healthcare outside the submitted work. Dr Sweeney reported receiving grants from Amgen Inc, AstraZeneca UK Ltd, Bristol Myers Squibb, Bayer Healthcare Pharmaceuticals Inc, Boehringer Ingelheim, Genentech, Janssen Pharmaceuticals Inc, Merck Sharp & Dohme Corp, Novartis, and Pfizer Inc during the conduct of the study. Ms Brown reported receiving grants from AACR during the conduct of the study. Dr Bedard reported receiving grants from AACR during the conduct of the study; consulting fees from Seattle Genetics, Eli Lilly and Co, Amgen Inc, Merck Sharp & Dohme Corp, Bristol Myers Squibb, Sanofi, and Pizer; serving on the advisory boards for Bristol Myers Squibb, Eli Lilly and Co, Amgen, Seattle Genetics, Merck, Pfizer, and Gilead; and receiving research funding directed to his institution from Bristol Myers Squibb, Sanofi, AstraZeneca, Genentech, Servier, GlaxoSmithKline, Novartis, SignalChem, PTC Therapeutics, Nektar, Merck Sharp & Dohme Corp, Seattle Genetics, Mersana, Immunomedics, and Eli Lilly and Co. Dr Schrag reported receiving fees from *JAMA* for editorial services; speaking fees from Pfizer; research funding given directly to the institution from AACR, Patient-Centered Outcomes Research Institute, GRAIL, and the National Cancer Institute (NCI). Dr Panageas reported receiving grants from NCI Cancer Center Support Grant and research funding from AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE) Consortium during the conduct of the study. AACR Project GENIE receives funding from Amgen Inc, Bristol Myers Squibb Co, Merck Sharp & Dohme Corp, AstraZeneca UK Ltd, Genentech, Novartis, Bayer Healthcare Pharmaceuticals Inc, H3 Biomedicine, Puma Biotechnology, Boehringer Ingelheim, and Janssen Pharmaceuticals Inc. No other disclosures were reported.

**Funding/Support:** This study was supported by the AACR; Wong Family Foundation, Simeon J. Fortin Foundation, Doris Duke Foundation, and NCI grant K99CA245899-01 (Dr Kehl); and Cancer Center support grant P30 CA08748 and the NCI (Dr Riely, Mss Lavery, Rudolph, Brown, and Dr Panageas).

**Role of the Funder/Sponsor:** The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Group Members:** The AACR Project GENIE Consortium members are listed in [Supplement 2](#).

**Disclaimer:** Interpretations are the responsibility of study authors.

**Additional Contributions:** The authors would like to acknowledge the AACR and its financial and material support in the development of the AACR Project GENIE registry as well as members of the AACR Project GENIE consortium for their commitment to data sharing.

## REFERENCES

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026
2. Bertagnoli MM, Anderson B, Norsworthy K, et al. Status update on data required to build a learning health system. *J Clin Oncol*. 2020;38(14):1602-1607. doi:10.1200/JCO.19.03094
3. AACR Project GENIE Consortium. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discov*. 2017;7(8):818-831. doi:10.1158/2159-8290.CD-17-0151
4. Blumenthal GM, Gong Y, Kehl K, et al. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non-small-cell lung cancer. *Ann Oncol*. 2019;30(5):830-838. doi:10.1093/annonc/mdz060
5. Chang E, Gong Y, Weinstock C, et al. FDA pooled analysis of time to treatment discontinuation (TTD) in frontline advanced renal cell carcinoma trials. *J Clin Oncol*. 2020;38(15\_suppl):5081. doi:10.1200/JCO.2020.38.15\_suppl.5081
6. Marshall J, Schwartzberg LS, Bepler G, et al. Novel panomic validation of time to next treatment (TNT) as an effective surrogate outcome measure in 4,729 patients. *J Clin Oncol*. 2016;34(15\_suppl):11521. doi:10.1200/JCO.2016.34.15\_suppl.11521
7. Feinberg B, Kish J, Dokubo I, Wojtynek J, Gajra A, Lord K. Comparative effectiveness of palliative chemotherapy in metastatic breast cancer: a real-world evidence analysis. *Oncologist*. 2020;25(4):319-326. doi:10.1634/theoncologist.2019-0699
8. Griffith SD, Miksad RA, Calkins G, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small-cell lung cancer data set. *JCO Clin Cancer Inform*. 2019;3:1-13. doi:10.1200/CCI.19.00013
9. Belin L, Tan A, De Rycke Y, Dechartres A. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. *Br J Cancer*. 2020;122(11):1707-1714. doi:10.1038/s41416-020-0805-y
10. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol*. 2005;23(34):8664-8670. doi:10.1200/JCO.2005.01.6071
11. Sargent D, Shi Q, Yothers G, et al; Adjuvant Colon Cancer End-points (ACCENT) Group. Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C-07 and C89803. *Eur J Cancer*. 2011;47(7):990-996. doi:10.1016/j.ejca.2010.12.015
12. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin Pharmacol Ther*. 2018;103(2):202-205. doi:10.1002/cpt.946
13. Schrag D. GENIE: real-world application. Presented at: ASCO Annual Meeting; June 7, 2018; Chicago, Illinois.
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X
15. Ettinger DS, Wood DE, Aggarwal C, et al; OCN. NCCN guidelines insights: non-small cell lung cancer, version 1.2020. *J Natl Compr Canc Netw*. 2019;17(12):1464-1472. doi:10.6004/jnccn.2019.0059
16. Schemper M, Kaider A, Wakounig S, Heinze G. Estimating the correlation of bivariate failure times under censoring. *Stat Med*. 2013;32(27):4781-4790. doi:10.1002/sim.5874
17. Ploner M, Kaider A, Heinze G. SurvCorr: correlation of bivariate survival times. R package, version 1.0. R Group for Statistical Computing; 2015.

18. Khozin S, Miksad RA, Adami J, et al. Real-world progression, treatment, and survival outcomes during rapid adoption of immunotherapy for advanced non-small cell lung cancer. *Cancer*. 2019;125(22):4019-4032. doi:10.1002/ncr.32383
19. Savova GK, Danciu I, Alamudun F, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. *Cancer Res*. 2019;79(21):5463-5470. doi:10.1158/0008-5472.CAN-19-0579
20. Kehl KL, Elmarakeby H, Nishino M, et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol*. 2019;5(10):1421-1429. doi:10.1001/jamaoncol.2019.1800
21. Arbour KC, Luu AT, Luo J, et al. Deep learning to estimate RECIST in patients with NSCLC treated with PD-1 blockade. *Cancer Discov*. 2021;11(1):59-67. doi:10.1158/2159-8290.CD-20-0419
22. Kehl KL, Xu W, Lepisto E, et al. Natural language processing to ascertain cancer outcomes from medical oncologist notes. *JCO Clin Cancer Inform*. 2020;4:680-690. doi:10.1200/CCI.20.00020
23. Li Y, Luo Y-H, Wampfler JA, et al. Efficient and accurate extracting of unstructured EHRs on cancer therapy responses for the development of RECIST natural language processing tools, part I: the corpus. *JCO Clin Cancer Inform*. 2020;4(4):383-391. doi:10.1200/CCI.19.00147
24. Power DG, Kemeny NE. The role of floxuridine in metastatic liver disease. *Mol Cancer Ther*. 2009;8(5):1015-1025. doi:10.1158/1535-7163.MCT-08-0709
25. Mok TSK, Wu Y-L, Kudaba I, et al; KEYNOTE-042 Investigators. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7
26. Rosell R, Carcereny E, Gervais R, et al; Spanish Lung Cancer Group in collaboration with Groupe Français de Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica. 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(3):239-246. doi:10.1016/S1470-2045(11)70393-X
27. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*. 2009;361(10):947-957. doi:10.1056/NEJMoa0810699
28. Cain KC, Harlow SD, Little RJ, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am J Epidemiol*. 2011;173(9):1078-1084. doi:10.1093/aje/kwq481
29. Kehl KL, Schrag D, Hassett MJ, Uno H. Assessment of temporal selection bias in genomic testing in a cohort of patients with cancer. *JAMA Netw Open*. 2020;3(6):e206976. doi:10.1001/jamanetworkopen.2020.6976
30. Validity of surrogate endpoints in oncology: executive summary of rapid report A10-05, version 1.1. 2005. Accessed January 5, 2021. <https://www.ncbi.nlm.nih.gov/pubmed/24783303>
31. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989;8(4):431-440. doi:10.1002/sim.4780080407
32. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613. doi:10.7326/0003-4819-125-7-199610010-00011
33. Duke Margolis Center for Health Policy. A roadmap for developing study endpoints in real-world settings. 2020. Accessed January 5, 2021. <https://healthpolicy.duke.edu/publications/roadmap-developing-study-endpoints-real-world-settings>

#### SUPPLEMENT 1.

**eTable 1.** Correlations Between Surrogate Outcomes and Overall Survival by Therapy Regimen Category: Sensitivity Analysis Restricted to Patients Who Started First-Line Therapy After Genomic Testing

**eTable 2.** Most Common Systemic Therapy Regimens Administered as First-Line Therapy

**eTable 3.** Correlations Between Surrogate Outcomes and Overall Survival by Therapy Regimen Category: Sensitivity Analysis Excluding Patients Who Died Without Experiencing a Surrogate Outcome Event

**eFigure 1.** Cohort Derivation: Non-Small Cell Lung Cancer (NSCLC)

**eFigure 2.** Cohort Derivation: Colorectal Cancer (CRC)

#### SUPPLEMENT 2.

**Nonauthor Collaborators.** AACR Project GENIE Consortium Members