Comparison of Solid Tumor Treatment Response Observed in Clinical Practice With Response Reported in Clinical Trials

Bruce A. Feinberg, DO; Marjorie E. Zettler, PhD, MPH; Andrew J. Klink, PhD, MPH; Choo H. Lee, MS; Ajeet Gajra, MD; Jonathan K. Kish, PhD, MPH

Abstract

**IMPORTANCE** In clinical trials supporting the regulatory approval of oncology drugs, solid tumor response is assessed using Response Evaluation Criteria in Solid Tumors (RECIST). Calculation of RECIST-based responses requires sequential, timed imaging data, which presents challenges to the method's application in real-world evidence research.

**OBJECTIVE** To evaluate the feasibility and validity of a novel real-world RECIST method in assessing tumor burden associated with therapy for a large heterogeneous patient population undergoing treatment in routine clinical practice.

**DESIGN, SETTING, AND PARTICIPANTS** This cohort study used physician-abstracted data pooled from retrospective, multisite electronic health record (EHR) review studies of patients treated with anticancer drugs at US oncology practices from 2014 through 2017. Included patients were receiving first-line treatment for thyroid cancer, breast cancer, or metastatic melanoma. Data were analyzed from March through August 2020.

**EXPOSURES** Undergoing treatment with immunotherapy or targeted therapy.

**MAIN OUTCOMES AND MEASURES** Tumor response was classified according to RECIST guidelines (ie, change in sum diameter of target lesions) post hoc with measurements derived from imaging scans and reports.

**RESULTS** Among 1308 completed electronic case report forms, 956 forms (73.1%) had adequate data to classify tumor response using a novel method, real-world RECIST response. The greatest difference between physician-recorded responses and real-world RECIST-based responses was found in the proportion of complete responses: 118 responses (12.3%) vs 46 responses (4.8%) ($P < .001$). Among 609 patients in the metastatic melanoma population, complete responses were reported in 112 physician-recorded responses (18.4%) vs 44 real-world RECIST-based responses (7.2%) ($P < .001$), compared with 11 of 247 responses (4.5%) to 31 of 192 responses (16.1%) across pivotal trials of the same melanoma therapies.

**CONCLUSIONS AND RELEVANCE** These findings suggest that comparing tumor lesion sizes and categorizing treatment response according to RECIST guidelines may be feasible using real-world data. This study found that physician-recorded assessments were associated with overestimation of treatment response, with the largest overestimation among complete responses. Real-world RECIST-based assessments were associated with better approximations of tumor response reported in clinical trials compared with those reported in EHRs.

Introduction

Real-world data provides an opportunity to gain valuable insight into the clinical effectiveness and safety associated with oncology drugs in a broader patient population than that enrolled in clinical trials, under the less structured and stringent real-world circumstances of clinical practice. Interest in real-world data to generate efficacy data associated with oncology drugs has increased significantly in recent years as real-world evidence is increasingly accepted as a complement to randomized clinical trials in supporting regulatory approval for new drug indications. As a result, there is increasing interest in determining the validity and practicality of estimating traditional efficacy end points for oncology drugs in routine clinical settings. This is a critical need, as less than 5% of adult patients with cancer in the United States participate in clinical trials, and the patients that do are younger, healthier, and less diverse.

In clinical trials supporting the Food and Drug Administration (FDA) approval of oncology drugs, the most common primary end point over the past few years has been response rate, followed by progression-free survival. To estimate solid tumor response, or progression of disease, the predominant standard used in oncology clinical trials is Response Evaluation Criteria in Solid Tumors (RECIST). These guidelines require serial imaging: a baseline or pretreatment assessment and interval posttreatment assessments for response or progression, with protocol-specified frequency and radiologic modality. Such a rigid structure can be challenging to replicate with real-world data, for which imaging is performed at variable frequencies, using a variety of different diagnostic modalities, at differing sites of care, and interpreted by different radiologists.

Owing to these complexities, some studies using real-world data rely on the treating physician’s assessment of tumor response, as recorded in the narrative of the patients’ electronic health records (EHRs), using manual or technology-enabled (eg, natural language processing) EHR abstraction, as the measure of clinical outcomes. Alternatively, some real-world data researchers elect to evaluate more easily obtained surrogate end points, such as time to treatment failure or time to treatment discontinuation (ie, the length of time from treatment initiation to treatment discontinuation for any reason). These options have significant limitations, as physicians’ estimates of tumor response may be subject to bias and surrogate end points may not be directly comparable to clinical trial end points.

The validity of real-world end points depends on the accuracy, measurability, and reproducibility of the underlying real-world data and the methods used to derive the real-world evidence. In this study, we present a novel real-world method for calculating tumor response using lesion measurement data abstracted into an electronic case report form from imaging reports (or made directly from the images themselves) post hoc by a physician who treated the patient. Using these measurements of target lesions, a real-world RECIST response can be calculated using the RECIST version 1.1 guidelines on the extent of lesion size changes and the development of new lesions. The objective of our study was to compare real-world RECIST response with physician-recorded response in a large, heterogenous population of patients undergoing treatment with anticancer drugs for several different solid tumor indications. To internally validate our findings, a separate analysis was performed in a subset of this population limited to patients treated for a single indication (ie, metastatic melanoma), comparing results with responses in health records and with the results of clinical trials for melanoma agents.

Methods

This cohort study pooled data from 4 retrospective, multisite patient EHR review studies for 3 different indications (ie, metastatic melanoma, metastatic breast cancer, and metastatic differentiated thyroid cancer) to describe outcomes, including disease response, for patients undergoing systemic treatment at oncology clinics in the United States. An independent institutional review board reviewed and approved each study protocol and electronic case report.
form and provided waivers for informed consent under 45 CFR 46.116(f) (2018 requirements) and 45 CFR 46.116(d) (pre-2018 requirements). The Cardinal Health Specialty Solutions Ethics Committee determined that no formal review or approval were required, as the study used only deidentified, aggregated data, and that no informed consent was required, per the requirements of 45 CFR 46.116(f). The study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Physicians in the Cardinal Health Oncology Provider Extended Network were asked to identify patients undergoing treatment for specific indications between 2014 and 2017 at their practices. (This network is a community of more than 7000 oncologists geographically distributed across the United States, of which approximately 800 comprise the real-world research community, with 300 unique investigators having contributed patient-level data to retrospective EHR abstraction research studies since 2016.) Deidentified patient-level data were abstracted by the physicians from the EHRs into electronic case report forms. Physicians were asked to identify the earliest patient meeting the selection criteria and select patients chronologically forward in time until submitting all eligible patients or the maximum number of patients allowed per physician for that study (typically 10). Data collection for all studies occurred in 2018.

Physicians abstracting the data were asked to indicate each patients’ best response to therapy based on the disease response in the EHR narrative: complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). Physicians were also asked to abstract lesion measurements from available imaging reports or from the images themselves (accessed via picture archiving and communication systems, a medical imaging technology that provides storage, retrieval, and distribution of medical images of multiple modalities that is in near universal use in US hospitals) at initiation of treatment and at time of best response to therapy. Physicians were instructed to abstract measurements (ie, the longest diameters of the lesions) and locations for 5 target lesions (up to 2 per organ). The definition of a measurable lesion per RECIST version 1.1 (ie, ≥10 mm or ≥15 mm short axis for lymph nodes) was provided to the physicians. The time span in which we collected the data precedes the development of immune RECIST, a modified version of the RECIST criteria that was developed to measure tumor response specifically in patients receiving immunotherapy, so it was not feasible to analyze the data using immune RECIST.

Source documents were not evaluated; however, quality control audits of submitted electronic case report forms were conducted to evaluate accuracy. Submitted data were reviewed by clinical research staff and the study statistician (C.H.L.) for missing data or outliers, such as implausible dates or radiology results inconsistent with known clinical parameters. The physician who abstracted the data was contacted and asked to verify flagged data entries. Additionally, a random sample of all submitted electronic case report forms was validated through physician follow-up. Patient records of physicians who did not respond to requests for validation or who did not accurately verify data from the audits were removed from the data set.

Statistical Analysis
The study statistician (C.H.L.) performed real-world RECIST classification using the tumor measurements at baseline and best response reported in the electronic case report form based on RECIST version 1.1 guidelines, assigning response as CR, PR, SD, or PD. Descriptive measures, including counts and frequencies for categorical variables and measures of centrality (ie, median) and spread (ie, minimum and maximum) for continuous variables, were used to summarize treatment responses and changes in response. \( \kappa \) coefficients were calculated to measure the magnitude of agreement between the best response to therapy reported in the narrative of the patient’s EHRs and that retrospectively classified by the research team using RECIST. Weighted \( \kappa \) coefficients, which consider categories as ordered and account for how far apart classifications are, were also calculated. The primary outcome, comparisons of the proportions of physician-recorded responses to real-world RECIST responses for each of the classifications (ie, CR, PR, SD, and PD), was evaluated using the \( \chi^2 \) test. Statistical significance was determined at 2-sided \( \alpha = .05 \), and statistical analysis was
performed using SAS statistical software version 9.4 (SAS Institute) from March through August 2020.

As an internal validation, a second analysis was conducted for a subset of the overall pooled data, limited to those patients treated for a single indication (ie, metastatic melanoma). Physician-recorded responses and real-world RECIST-based responses were then indirectly compared with those from the pivotal trials for the same treatments approved for metastatic melanoma. Among 6 pivotal trials, responses were assessed by the investigator in 4 trials, by central review in 1 trial, and by the investigator and central review in 1 trial.

Results

Among 1308 patients with electronic case report forms submitted by 175 physicians, forms for 80 patients had baseline image measurements unavailable, 11 had best response image measurements unavailable, and 261 had baseline and best response image measurements unavailable. Reasons for missing scans included extensive or nonvisceral (eg, bone or brain) metastases and individual study design, including the breast cancer study analysis, which had 135 patients categorized as too early to determine best response. This left a sample size of forms from 956 patients (73.1%) with complete image measurements available. The median (interquartile range) time to best response was 15.1 (11.0-24.6) weeks. Of the patients represented, 609 patients (63.7%) underwent treatment for BRAF V600E metastatic melanoma, 239 patients (25.0%) underwent treatment for metastatic breast cancer, and 108 patients (11.3%) underwent treatment for metastatic differentiated thyroid cancer. Approximately half of the patients with metastatic melanoma were receiving first-line treatment with immunotherapy (the remainder received first-line treatment with BRAF/MEK combination therapy). Real-world RECIST calculations and classifications were performed for all 956 patients.

The tumor responses as reported in the patient EHR by the physician and as calculated according to RECIST are presented in Table 1; details of the CR and PR responses are described in the table. More physician-recorded responses than real world RECIST-based responses were categorized as CRs (118 responses [12.3%] vs 46 responses [4.8%]; \( P < .001 \)). Of the physician-recorded CRs, 43 responses (36.4%; 95% CI, 27.8%-45.8%) were also classified as CRs by real-world RECIST. Of the remaining 75 physician-recorded CRs, real-world RECIST classified 65 responses (55.1%) as PRs, 6 responses (5.1%) as SDs, and 4 responses (3.4%) as PDs. The proportion of responses categorized as PRs was similar for physician and real-world RECIST responses (571 responses [59.7%] vs 562 responses [58.8%]; \( P < .001 \)). However, of the PRs reported by physicians, real-world RECIST classified 470 responses (82.3%; 95% CI, 78.9%-85.4%) as PRs, 2 responses (0.4%) as CRs, 67 responses (11.7%) as SDs, and 32 responses (5.6%) as PDs. The \( \kappa \) coefficient was 0.58 (95% CI, 0.53-0.62), and the weighted \( \kappa \) was 0.64 (95% CI, 0.59-0.68).

Percentage change in lesion measurements between baseline and best response is plotted against physician-recorded responses in Figure 1. The greatest median (range) percent decrease in

| Table 1. Treatment Response by Assessment Method, Overall Patient Population |
|-----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Physician-recorded responses | Real-world RECIST responses, No. (%) | Total physician-recorded responses |
| CRA | PR | SD | PD | Total real-world RECIST-based responses |
|---|---|---|---|---|---|
| CR | 43 (36.4) | 65 (55.1) | 6 (5.1) | 4 (3.4) | 118 (12.3) |
| PR | 2 (0.4) | 470 (82.3) | 67 (11.7) | 32 (5.6) | 571 (59.7) |
| SD | 1 (0.7) | 20 (14.1) | 109 (76.8) | 12 (8.5) | 142 (14.9) |
| PD | 0 (0) | 7 (5.6) | 23 (18.4) | 95 (76.0) | 125 (13.1) |
| Total real-world RECIST-based responses | 46 (4.8) | 562 (58.8) | 205 (21.4) | 143 (15.0) | 956 (100) |

Abbreviations: CR, complete response; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease.
tumor lesion measurements was −87.2% (−100.0% to 328.6%), in patients for whom the physician-recorded responses were CRs. This was followed by −52.9% (−100.0% to 484.9%) for physician-recorded PRs and −6.6% for physician-recorded SDs (−100.0% to 233.3%). Physician-recorded PD median (range) percent increase in lesion size was 29.9% (−82.0% to 1293.4).

The same analyses were performed for 609 patients receiving first-line treatment for the BRAF V600E metastatic melanoma indication. Tumor responses as reported in the patient EHR by the physician and as calculated according to RECIST for the metastatic melanoma subset are presented in Table 2. Similar to results for the overall patient population, the greatest difference between physician-recorded responses and real-world RECIST-based responses was found in those categorized as CR (112 responses [18.4%] vs 44 responses [7.2%]; P < .001). Of physician-recorded CRs, 41 responses (36.6%; 95% CI, 27.7%–46.2%) were also classified as CRs by real-world RECIST. For the remaining 71 physician-recorded CRs, real-world RECIST classified 61 responses (54.5%) as PRs, 6 responses (5.4%) as SDs, and 4 responses (3.6%) as PDs. The proportion of responses categorized as PRs was approximately equivalent between physician-recorded responses and real-world RECIST responses (358 responses [58.8%] vs 383 responses [62.9%]; P < .001). Among physician-recorded PRs, real-world RECIST classified 305 responses (85.2%; 95% CI, 81.1%–88.7%) as PRs, 2 responses (0.6%) as CRs, 32 responses (8.9%) as SDs, and 19 responses (5.3%) as PDs. The κ coefficient for agreement of responses within the metastatic melanoma subset was 0.55 (95% CI, 0.49–0.61), and the weighted κ was 0.62 (95% CI, 0.57–0.68).

Figure 2 presents the percentage change in lesion measurements between baseline and best response plotted against physician-recorded responses for the metastatic melanoma subset. Similar to results for the overall patient population, the greatest median (range) percent decrease in tumor lesion measurements was −87.2% (−100.0% to 328.6%) for CRs, followed by −57.0% (−100.0% to 484.9%) for PRs and 27.2% (−82.0% to 1293.4%) for PDs, with the lowest value at −9.1% (−100.0% to 233.3%) for SDs.

Table 2. Treatment Response by Assessment Method, Metastatic Melanoma Subset

<table>
<thead>
<tr>
<th>Physician-recorded responses</th>
<th>Real-world RECIST responses , No. (%)</th>
<th>Total physician recorded responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>41 (36.6)</td>
<td>112 (18.4)</td>
</tr>
<tr>
<td>PR</td>
<td>2 (0.6)</td>
<td>358 (58.8)</td>
</tr>
<tr>
<td>SD</td>
<td>1 (1.5)</td>
<td>65 (10.7)</td>
</tr>
<tr>
<td>PD</td>
<td>0 (0)</td>
<td>74 (12.2)</td>
</tr>
<tr>
<td>Total real-world RECIST-based responses</td>
<td>44 (7.2)</td>
<td>609 (100)</td>
</tr>
</tbody>
</table>

Abbreviations: CR, complete response; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease.

CR indicates complete response; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease; blue dashed line, cutoff for classification as PD (ie, at least a 20% increase in the sum of diameters of target lesions); gray dashed line, cutoff for classification as PR (ie, at least a 30% decrease in the sum of diameters of target lesions); circles, outliers; diamonds, means; midlines of boxes, medians; tops of boxes, lower quartiles (Q1s); and whiskers, ranges for top and bottom 25% of data values, excluding outliers. Seven patients with a greater than 200% increase are not displayed.
The responses obtained in this analysis and RECIST-based responses from pivotal trials\textsuperscript{15-20} of agents approved for the treatment of metastatic melanoma are presented in Table 3. Variability was observed across different classes of agents and between investigator-assessed and central review-assessed responses, with CRs ranging from 11 of 247 responses (4.5\%) for investigator-assessed cobimetinib plus vemurafenib to 31 of 192 responses (16.1\%) for investigator-assessed encorafenib plus binimetinib.

### Discussion

This cohort study found that the application of RECIST to solid tumor measurements derived through retrospective EHR review may be a feasible approach to determine tumor response in a real-world setting across a range of cancer diagnoses. Although the frequency and timing of response assessment imaging was variable, participating physician abstractors were able to provide sufficient data to derive meaningful conclusions.

#### Figure 2. Change in Lesion Measurements for Metastatic Melanoma Subset

- CR indicates complete response; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease; blue dashed line, cutoff for classification as PD (ie, at least a 20\% increase in the sum of diameters of target lesions); gray dashed line, cutoff for classification as PR (ie, at least a 30\% decrease in the sum of diameters of target lesions); circles, outliers; diamonds, means; midlines of boxes, medians; tops of boxes, lower quartiles (Q1s); and whiskers, ranges for top and bottom 25\% of data values, excluding outliers. Five patients with a greater than 200\% increase are not displayed.

#### Table 3. Best Responses for First-Line Therapy of Metastatic Melanoma

<table>
<thead>
<tr>
<th>Drug</th>
<th>No. (%)</th>
<th>CR</th>
<th>PR</th>
<th>SD</th>
<th>PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled (n = 609)\textsuperscript{*}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician recorded</td>
<td>112 (18.4)</td>
<td>358 (58.8)</td>
<td>65 (10.7)</td>
<td>74 (12.2)</td>
<td></td>
</tr>
<tr>
<td>Real-world RECIST</td>
<td>44 (7.2)</td>
<td>383 (62.9)</td>
<td>91 (14.9)</td>
<td>91 (14.9)</td>
<td></td>
</tr>
<tr>
<td>Immunotherapy (monotherapy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nivolumab; Robert et al,\textsuperscript{16} 2015 (n = 210)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator assessed</td>
<td>16 (7.6)</td>
<td>68 (32.4)</td>
<td>35 (16.7)</td>
<td>69 (32.9)</td>
<td></td>
</tr>
<tr>
<td>Pembrolizumab; Robert et al,\textsuperscript{17} 2015 (n = 277)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central review</td>
<td>17 (6.1)</td>
<td>74 (26.7)</td>
<td>39 (14.1)</td>
<td>114 (41.2)</td>
<td></td>
</tr>
<tr>
<td>Immunotherapy (combination)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nivolumab plus ipilimumab; Larkin et al,\textsuperscript{18} 2015 (n = 314)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator assessed</td>
<td>36 (11.5)</td>
<td>145 (46.2)</td>
<td>41 (13.1)</td>
<td>71 (22.6)</td>
<td></td>
</tr>
<tr>
<td>Targeted therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dabrafenib plus trametinib; Long et al,\textsuperscript{19} 2014 (n = 211)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator assessed</td>
<td>22 (10.4)</td>
<td>118 (55.9)</td>
<td>54 (25.6)</td>
<td>13 (6.2)</td>
<td></td>
</tr>
<tr>
<td>Cobimetinib plus vemurafenib; Larkin et al,\textsuperscript{20} 2014 (n = 247)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator assessed</td>
<td>11 (4.5)</td>
<td>100 (40.5)</td>
<td>105 (42.5)</td>
<td>25 (10.1)</td>
<td></td>
</tr>
<tr>
<td>Encorafenib plus binimetinib; Dummer et al,\textsuperscript{15} 2018 (n = 192)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigator assessed</td>
<td>31 (16.1)</td>
<td>113 (58.9)</td>
<td>35 (18.2)</td>
<td>13 (6.8)</td>
<td></td>
</tr>
<tr>
<td>Central review</td>
<td>15 (7.8)</td>
<td>106 (55.2)</td>
<td>56 (29.2)</td>
<td>15 (7.8)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CR, complete response; PD, progressive disease; PR, partial response; RECIST, Response Evaluation Criteria in Solid Tumors; SD, stable disease.\textsuperscript{*} From this study. Includes nivolumab monotherapy, pembrolizumab monotherapy, nivolumab plus ipilimumab, dabrafenib plus trametinib, cobimetinib plus vemurafenib, and encorafenib plus binimetinib.
data for real-world RECIST calculation in nearly three-quarters of the pooled study cases. The inability
to provide sufficient data was due to multiple factors: absent measurements in imaging report,
inability to access digital image, extensive or non viscera (eg, bone or brain) metastases, and
individual study design (eg, the breast cancer study analysis included 135 patients who were
categorized as too early to determine best response).

Our study identified significant differences between the tumor responses noted in the patients’
EHRs by the physician and the real-world RECIST-based tumor responses classified by the research
team based on tumor measurements; this difference was greatest in the response categorizations of
CR. While this study did not assess the reasons for this variability, a 2018 study\(^4\) found that
differences in pretreatment and posttreatment imaging technology and inconsistency in the target
lesions imaged, measured, or reported were among the reasons associated with such variability. In
clinical trials, it is not uncommon for investigator-assessed responses to overestimate treatment
effect. An analysis\(^2\) of 28 phase 3 clinical trials for anticancer drugs in patients with solid tumors
found that central assessment consistently reported lower objective response rates and disease
control rates than local assessment, in blinded and unblinded trials and in control and experimental
arms. This analysis supported the conclusions of 2 meta-analyses\(^22,23\) with respect to the
overestimation of objective response rate by local investigators. This is likely owing to the subjective
nature of the local investigator’s assessment, which may draw from more than imaging data and take
into account other factors, such as patient-reported symptoms or clinical laboratory test
results.\(^14,22,23\) This known discordance between local and central review has important ramifications
for interpreting real-world studies that rely on tumor responses as recorded by physicians in
patients’ EHRs.

Other researchers have attempted to circumvent the perceived difficulty in using RECIST to
characterize tumor response in a real-world setting by using surrogate end points, such as time to
treatment failure or time to treatment discontinuation. Although these end points are easier to
derive from real-world data sources like EHRs and claims databases, they are not often evaluated in
pivotal clinical trials. The FDA does not recommend end points like time to treatment failure in clinical
trials for the approval of cancer drugs because of the inability to distinguish between patients who
discontinued treatment due to disease progression and those who discontinued for other reasons,
such as toxic effects.\(^24\) Analyses from 2019\(^25\) and 2018\(^26\) of clinical trials for metastatic non-small cell
lung cancer and melanoma found that time to discontinuation was associated with progression-free
survival.\(^25,26\) However, in a clinical trial, the treatment duration is often mandated by protocol and
investigators may not have the option of treating beyond disease progression as they do in clinical
practice. Evaluation of time to treatment discontinuation within these confines may be associated
with value of limited clinical validity that differs significantly from that which would be obtained in a
real-world-evidence study. Thus, these alternative end points may be problematic both from the
standpoint of their imprecision with respect to clinical efficacy and the inability to compare the
results in the real-world study with those in the pivotal clinical trials in a meaningful way.

For a subset of cases within a specific indication (ie, BRAF V600\(^+\) metastatic melanoma), real-
world RECIST-based responses were similar to those in the overall population. However, because the
patients with metastatic melanoma comprised most of the overall patient population, this was not a
surprising finding. To put the metastatic melanoma results in context, they were reviewed relative to
responses assessed in pivotal clinical trials for several agents approved by the FDA for the first-line
metastatic melanoma indication. In the pivotal trials, responses were primarily assessed by the
investigator; that is, the local investigator assigned 1 of the 4 RECIST responses to each patient, with
oversight from the trial sponsor. These responses may be expected to align more closely with
RECIST-based responses than the physician-recorded responses in our study but still allow for the
investigator to apply clinical judgement based on factors other than the lesion measurements (as
opposed to blinded independent central review, which relies solely on the lesion measurements).
This is apparent in the results of the COLUMBUS trial,\(^15\) in which blinded independent central review-
assessed CRs were reported in 8.0% of patients treated with encorafenib and binimetinib, while
investigator-assessed CRs were reported in 16.0% of patients. For comparison, 18.4% of physician-recorded responses in our metastatic melanoma analysis were CRs (2.5-fold the proportion of CRs identified by real-world RECIST [7.2%]). This finding supports those from a 2017 pooled analysis,\textsuperscript{21} a 2012 review,\textsuperscript{22} and a 2010 study\textsuperscript{23} of the tendency for local investigators to overestimate treatment outcomes and underscores the unreliability of determining real-world effectiveness solely by subjective physician assessments.

**Limitations**

This study has several limitations. First, the images used for real-world RECIST measurements did not undergo blinded independent central review, so the potential exists for imaging reader bias. Second, although half of the patients with metastatic melanoma evaluated in this study were treated with immunotherapy, most patients received therapy prior to the introduction of immune RECIST, and therefore the modified criteria were not used.\textsuperscript{27} However, the novel method described herein, in which raw data elements were abstracted rather than an EHR estimate recapitulated, can be expanded to iRECIST as well as other complex clinical status measurements, such as Lugano classification or Cheson criteria, or physiologic measures like the Child-Pugh score. Third, over a quarter of the patients in our study were missing 1 or both scans. As our intent was to realistically assess method feasibility, we did not require scan availability as an eligibility criterion. Fourth, the timing of the imaging was variable, which precludes precision in time to response outcomes. This variability may also reflect possible misclassification errors by physician abstractors.

Replicating clinical trial end points in the real world is a complicated endeavor. A 2019 analysis\textsuperscript{28} of 220 clinical trials found that 15% could be replicated using real-world data; the inability to reliably ascertain a primary end point from EHR or claims data was one of the barriers. Our real-world RECIST methods may add expense and time to collect and analyze data compared with abstracting the physician-recorded response from the EHR narrative or evaluating surrogate end points like time to treatment discontinuation; however, this method was associated with greater accuracy and an apples to apples comparison with clinical trial data for anticancer drugs in solid tumors. Use of this method may also help to surmount a major impediment to the use of oncology drug real-world evidence by regulatory agencies and could represent a new benchmark for assessment of tumor response outside of clinical trials. This study represents a first step, comparing real-world RECIST with the current standard used in real-world evidence studies (ie, physician-recorded response). Future studies will incorporate blinded independent central review as further validation of this methodology.

**Conclusions**

The findings of this cohort study suggest that a real-world RECIST approach may be feasible. The differences between local physician-recorded tumor responses and central review RECIST-based responses were similar to that previously reported in the clinical trial setting. Additionally, when real-world RECIST outcomes in a specific indication (ie, metastatic melanoma) were compared with RECIST outcomes from pivotal clinical trials of agents approved for that indication, results were similar, suggesting the validity of our method, despite variability in the timing of imaging. In sum, a real-world RECIST method may provide a clinically meaningful measure of tumor response in the real-world setting that may approximate the measure used in clinical trials.
Observed Solid Tumor Treatment Response vs Response Reported in Clinical Trials

Open Access: This is an open access article distributed under the terms of the CC-BY-NC-ND License. © 2021 Feinberg BA et al. JAMA Network Open.

Corresponding Author: Marjorie Zettler, PhD, MPH, Director of Research Strategy, Cardinal Health Specialty Solutions, 7000 Cardinal Pl, Dublin, OH 43017 (marjorie.zettler@cardinalhealth.com).

Author Affiliations: Cardinal Health Specialty Solutions, Dublin, Ohio.

Author Contributions: Drs Feinberg and Kish had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Feinberg, Zettler, Klink, Gajra, Kish.

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

Drafting of the manuscript: Feinberg, Zettler, Lee, Gajra.

Critical revision of the manuscript for important intellectual content: Feinberg, Zettler, Klink, Gajra, Kish.

Statistical analysis: Klink, Lee, Kish.

Administrative, technical, or material support: Klink, Gajra, Kish.

Supervision: Feinberg, Klink, Gajra, Kish.

Conflict of Interest Disclosures: All authors reported serving as employees of Cardinal Health, which receives funding to conduct research outside of this study from biopharmaceutical manufacturers. Dr Gajra reported serving as an employee of Icon outside the submitted work.

REFERENCES


