# Response Rates and Durations of Response for Biomarker-Based Cancer Drugs in Nonrandomized Versus Randomized Trials

Bishal Gyawali, MD, PhD[a,b,c]; Elvira D'Andrea, MD, MPH[a]; Jessica M. Franklin, PhD[a]; and Aaron S. Kesselheim, MD, JD, MPH[a]

## ABSTRACT

**Background:** Many new targeted cancer drugs have received FDA approval based on durable responses in nonrandomized controlled trials (non-RCTs). The goal of this study was to evaluate whether the response rates (RRs) and durations of response (DoRs) of targeted cancer drugs observed in non-RCTs are consistent when these drugs are tested in RCTs. **Methods:** We used the FDA's Table of Pharmacogenomic Biomarkers in Drug Labeling to identify cancer drugs that were approved based on changes in biomarker endpoints through December 2017. We then identified the non-RCTs and RCTs for these drugs for the given indications and extracted the RRs and DoRs. We compared the RRs and median DoR in non-RCTs versus RCTs using the ratio of RRs and the ratio of DoRs, defined as the RRs (or DoRs) in non-RCTs divided by the RRs (or DoRs) in RCTs. The ratio of RRs or DoRs was pooled across the trial pairs using random-effects meta-analysis. **Results:** Of the 21 drug–indication pairs selected, both non-RCTs and RCTs were available for 19. The RRs and DoRs in non-RCTs were greater than those in RCTs in 63% and 87% of cases, respectively. The pooled ratio of RRs was 1.06 (95% CI, 0.95–1.20), and the pooled ratio of DoRs was 1.17 (95% CI, 1.03–1.33). RRs and DoRs derived from non-RCTs were also poor surrogates for overall survival derived from RCTs. **Conclusions:** The RRs were not different between non-RCTs and RCTs of cancer drugs approved based on changes to a biomarker, but the DoRs in non-RCTs were significantly higher than in RCTs. Caution must be exercised when approving or prescribing targeted drugs based on data on durable responses derived from non-RCTs, because the responses could be overestimates and poor predictors of survival benefit.

*J Natl Compr Canc Netw 2020;18(1):36–43*
*doi: 10.6004/jnccn.2019.7345*

[a]Program on Regulation, Therapeutics, and Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; and [b]Department of Oncology, Division of Cancer Care and Epidemiology, and [c]Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada.

## Background

In recent years, the FDA has approved numerous precision oncology drugs—anticancer drugs with an identified genomic target—based on early-stage (phase I or II) nonrandomized controlled trials (non-RCTs). The most commonly used endpoints in these pivotal trials are the surrogate measures of response rate (RR) and duration of response (DoR), sometimes collectively referred to as durable response rates (DRRs). In some cases, the RRs or DoRs in these early trials are impressive; for example, crizotinib had an RR of 72% with a median DoR of 17.6 months among patients with *ROS1*-positive lung cancer in a single-arm trial.[1]

However, RRs and DoRs derived from early-phase single-arm trials may overestimate the RRs and DoRs observed in later-phase RCTs because these surrogate measures are subject to measurement bias. A 2005 study showed that single-arm phase II trials exaggerated the RR for cytotoxic chemotherapies when compared with the RR in phase III RCTs of the same drugs for similar indications.[2] DRRs also may not be correlated with true clinical responses to therapy, such as overall survival (OS) and quality of life. Caution in relying on surrogate measures such as RR and DoR to guide treatment choices is particularly relevant when the trials are early-stage trials and lack other features commonly associated with high quality trials, such as randomization and blinding.

Investigators have argued that such concerns do not hold for precision oncology drugs and that patients, physicians, and regulators should be more confident in relying on RRs and DoRs derived from single-arm trials in this context.[3] Some have even argued that for targeted drugs, the need for randomized trials has been eliminated because the durable responses in single-arm trials are enough for regulatory approval.[4,5] For example, a previous meta-analysis concluded that precision oncology drugs are beneficial based on pooling data from phase I trials alone.[6] A recent paper also concluded that

See JNCCN.org for supplemental online content.

See page 113 for related commentary.

for precision oncology drugs, the intent of phase I trials has changed from testing a drug's safety to offering therapeutic potential because of improved responses.[7]

To review the reliability of RR and DoR as trial endpoints in early-phase trials of precision oncology products, we conducted a meta-epidemiologic study. We sought to assess the difference in RRs and DoRs between non-RCTs and RCTs of precision oncology drugs.

## Methods

This article describes a meta-epidemiologic study conducted according to the modified PRISMA guidelines for meta-epidemiologic studies.[8] The primary goal of this study was to compare the RRs and DoRs between non-RCTs and RCTs of anticancer drugs that have been approved with a biomarker-based indication (ie, precision oncology drugs). Secondarily, we also sought to assess the correlation of RRs and DoRs derived from non-RCTs with OS derived from RCTs.

### Selection of Drugs

We accessed the Table of Pharmacogenomic Biomarkers in Drug Labeling,[9] and downloaded a PDF of the table on March 23, 2018. The website reported that at the time of our download, the table had last been updated in December 2017. We selected the anticancer drugs from this table. To match the current understanding of precision oncology drugs, we excluded agents such as hormone therapies in breast cancer (eg, anastrozole) that could literally be considered biomarker-based therapies but are not considered precision oncology drugs by the oncology community. We also excluded drugs with biomarker-based labeling for safety (eg, dihydropyrimidine dehydrogenase gene testing and capecitabine).

We included precision oncology drugs with biomarker-based labeling for efficacy in advanced solid tumors, and excluded 4 drug–indication pairs because they were approved based on trials in unselected populations and an association with a biomarker was discovered later. The excluded drugs include erlotinib and gefitinib for non–small cell lung cancer (NSCLC), in which the initial trials were conducted broadly, but it was later found that these drugs benefited only patients with an *EGFR* mutation. The other 2 excluded drugs were cetuximab and panitumumab for colorectal cancer, which were first approved based on *EGFR* status but were revised to *RAS* status as a predictive marker for efficacy.

### Selection of Studies

For the drugs in the cohort, data were extracted based on biomarkers and disease setting (indication). We then retrieved the approval information to determine whether the drugs for the given indications were approved based on RCTs or single-arm trials. If the approval was received based on a single-arm trial, we searched the literature (PubMed, Google Scholar, and conference abstracts) for any subsequent RCTs in the same setting. If approval was received based on an RCT, we searched the literature for any single-arm trials of the drug in the same setting, which would likely have been conducted before the RCT. If more than one single-arm trial was available for the RCT, we chose the trial with patient and disease characteristics closest to those of the RCT. If a non-RCT contained multiple arms for different doses or different histologies, we chose the dose and tumor cohort with the nearest match to the RCT. If the trial characteristics were similar, we pooled the RRs and DoRs among the single-arm trials using fixed-effects meta-analysis and used the pooled value for our analysis. We did not anticipate finding more than one RCT with similar characteristics for a given single-arm trial, but planned to do the same in case we encountered such an occurrence.

With regard to immunotherapies, we included microsatellite instability (MSI)– or mismatch repair (MMR)–based approvals as biomarker-based approvals, but excluded PD-L1–based approvals. MSI- or MMR-based approval for immunotherapies is considered a classic example of precision oncology heralding the first FDA tumor-agnostic approval based on biomarkers. However, we did not consider PD-L1–based approval to be biomarker-based approval because some approvals are based on PD-L1 levels and some are not, the PD-L1 level cutoffs are not always uniform, and the consensus on PD-L1 as a biomarker and the correct method of quantifying PD-L1 levels are still being debated.

### Data Extraction and Statistical Analyses

We extracted data on RRs and DoRs for each single-arm trial and RCT from the original publication. We also extracted data on 95% CIs for the median RRs and DoRs. Whenever these data were unavailable, we retrieved them from results posted on ClinicalTrials.gov. Median values and 95% CIs were used to calculate the standard errors.

We visually compared the RRs and DoRs in non-RCTs versus RCTs using a scatterplot (Figure 1).

For statistical comparisons, we used the ratio of RRs (rRRs), defined as the RRs in non-RCTs divided by the those in RCTs, and the ratio of DoRs (rDoRs), defined as the DoRs in non-RCTs divided by those in RCTs. Thus, an rRR or rDoR $>1$ would mean that the RRs or DoRs were greater in non-RCTs than in RCTs, whereas an rRR or rDoR $<1$ would signify that the RRs or DoRs were greater in RCTs. We pooled the rRRs and rDoRs across the pairs by conducting random-effects meta-analyses because of the heterogeneity across trials.

For both RRs and DoRs derived from non-RCTs, we also estimated the relationship with OS derived from the RCTs using a joint random-effects model, as described by Korn et al.[10] This model estimates a linear model

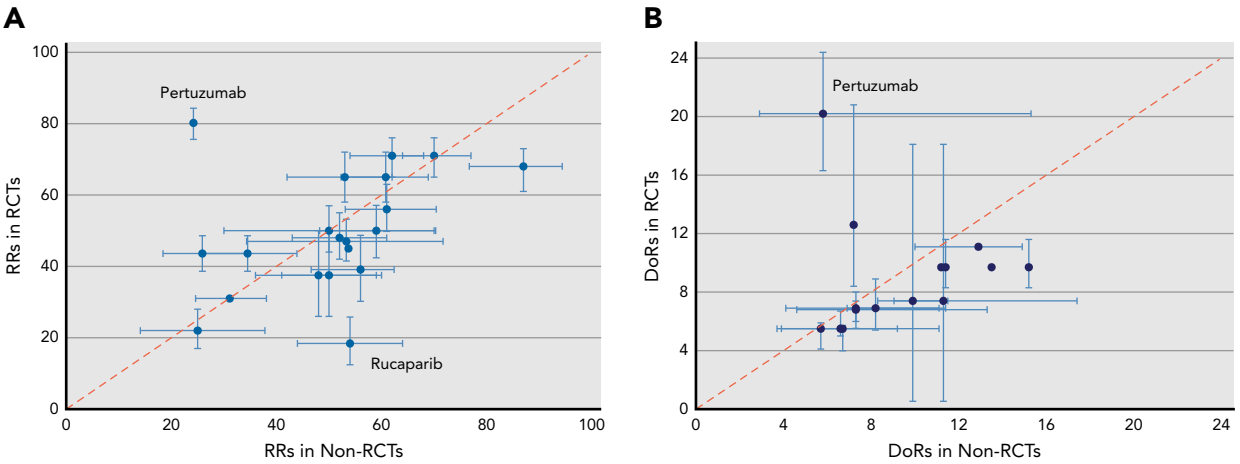



**Figure 1.** Scatterplots correlating (A) RRs and (B) DoRs in RCTs vs non-RCTs. The dashed line describes the function $y = x$.
Abbreviations: DoR, duration of response; RCT, randomized controlled trial; RR, response rate.

describing the relationship between the logarithmic OS hazard ratio (HR) and the logarithmic RR or DoR, accounting for the uncertainty in each of the estimates. Pooled analyses were conducted using STATA/SE 15.0 (StataCorp LLP), and other analyses were conducted using R version 3.2.3 (R Foundation for Statistical Computing).

## Results

The 21 drug–indication pairs in the cohort consisted of 19 drugs. Trastuzumab and crizotinib had 2 approvals each: trastuzumab for *HER2*-positive breast and gastric cancer, and crizotinib for *ALK*-positive and *ROS1*-positive NSCLC (Table 1). Fewer than half (n=9; 43%) were FDA-approved based on phase III RCTs, 10 approvals (48%) were based on phase II non-RCTs, and 2 (9%) were based on phase I single-arm trials—including accelerated approval of ceritinib for *ALK*-positive NSCLC and regular approval of crizotinib for *ROS1*-positive NSCLC. All of these drugs have approved companion diagnostics listed on the FDA website, except for nivolumab and pembrolizumab which do not have a companion diagnostic to establish MSI or MMR status.[11]

A total of 24 single-arm trials relating to these 21 drug–indication pairs were identified. For 4 drug–indication pairs, RCTs were still unavailable; for one indication (lapatinib), non-RCTs were unavailable (Table 2). Thus, 24 single-arm trials were paired with RCTs in 20 cases, involving 14 drugs for 15 indications. Imatinib was excluded from further analysis because the patients were randomized to 2 different doses of the same drug in both the phase II and III trials.

### RRs and DoRs in Non-RCTs Versus RCTs

RR values were identifiable for 19 RCT/non-RCT pairs. By contrast, median DoRs for both RCTs and non-RCTs allowing comparisons were available for 15 pairs.

The scatterplot for RRs (Figure 1A) shows no clear difference in clustering of pairs above or below the line of unity. RRs were greater in non-RCTs than in RCTs in most cases (63% of points below the line of unity), but the figure does not reveal any substantial differences.

For DoR, a clear signal emerged, with only 2 data points above the line of unity (Figure 1B). In 87% cases, DoRs were higher in non-RCTs than in RCTs.

The scatterplots for both RRs and DoRs reveal an outlier at the top left: pertuzumab in breast cancer. Pertuzumab was tested as a single agent in second-line treatment in the non-RCT and was tested in combination with trastuzumab and docetaxel in first-line treatment in the RCT. Similarly, the RR plot revealed rucaparib as another outlier. For rucaparib, although the non-RCT and initial approval were based on the presence of *BRCA* mutation, the RCT was conducted in an all-comer population as a maintenance therapy. Hence, for the pooled analysis, we decided to remove the pertuzumab and rucaparib pairs from the analysis.

### Pooled rRR and rDoR

Because ado-trastuzumab emtansine, alectinib, crizotinib, and osimertinib had 2 non-RCTs each that were matched to 1 RCT, first the 2 non-RCTs for each drug were pooled using fixed-effects meta-analysis to obtain a pooled RR and DoR, which were then compared with the RR and DoR from the randomized counterpart. Because 95% CIs or other measures of variation are required to pool across the trials, and because these data were not always available or computable, the sample sizes for evaluating rRRs and rDoRs were smaller (13 unique drug–indication pairs for rRRs and 11 unique drug–indication pairs for rDoRs).

The pooled rRR indicated that RRs in non-RCTs were 13% higher than in RCTs, but this difference was not

**Table 1. General Characteristics of Drug–Indication Pairs (N=21)**

| Drug | Biomarker | Indication | Phase of Trial on Which Approval Was Based |
|---|---|---|---|
| Ado-trastuzumab emtansine | *HER2* | Breast cancer | III |
| Afatinib | *EGFR* | NSCLC | III |
| Alectinib | *ALK* | NSCLC | II |
| Brigatinib | *ALK* | NSCLC | II |
| Ceritinib | *ALK* | NSCLC | I |
| Cobimetinib (+ vemurafenib) | *BRAF* | Melanoma | III |
| Crizotinib | *ALK* | NSCLC | II |
| Crizotinib | *ROS1* | NSCLC | I |
| Dabrafenib | *BRAF* | Melanoma | III |
| Lapatinib (+ capecitabine) | *HER2* | Breast cancer | III |
| Nivolumab | MSI-H or dMMR | Colorectal cancer | II |
| Olaparib | *BRCA* | Ovarian cancer | II |
| Osimertinib | *T790M* | NSCLC | II |
| Pembrolizumab | MSI-H or dMMR | Tissue agnostic | II |
| Pertuzumab | *HER2* | Breast cancer | III |
| Rucaparib | *BRCA* | Ovarian cancer | II |
| Trametinib | *BRAF* | Melanoma | III |
| Trastuzumab | *HER2* | Breast cancer | II |
| Trastuzumab | *HER2* | Gastric cancer | III |
| Vemurafenib | *BRAF* | Melanoma | III |
| Imatinib | *c-KIT* | GIST | II |

Abbreviations: dMMR, deficiency in mismatch repair; GIST, gastrointestinal stromal tumor; MSI-H, microsatellite instability-high; NSCLC, non–small cell lung cancer.

statistically significant (rRR, 1.06; 95% CI, 0.95–1.20; $I^2$=69%; $P$ for heterogeneity <.001) (Figure 2). The pooled rDoR showed that DoRs in non-RCTs were 17% higher than in RCTs, and this difference was statistically significant (rDoR, 1.17; 95% CI, 1.03–1.33; $I^2$=0%; $P$ for heterogeneity =.704) (Figure 3).

### Relationship Between RR or DoR From Non-RCTs and OS From RCTs

Because the ultimate marker of a drug's efficacy is improvement in OS, we also measured the magnitude of the linear association between logarithmic RR or DoR derived from non-RCTs and logarithmic OS HR derived from RCTs. The association was weak for RRs (n=12 pairs; β=0.165; 95% CI, −0.439 to 0.770) (supplemental eFigure 1, available with this article at JNCCN.org). The association was stronger for DoR, but was also not significant (n=9 pairs; β=0.746; 95% CI, −0.028 to 1.520) (supplemental eFigure 2).

## Discussion

In this systematic analysis of precision oncology drugs approved by the FDA, RRs in non-RCTs were not systematically higher than those seen in RCTs. However, DoRs for these drugs were higher in non-RCTs than in RCTs.

There are various reasons why the DoRs among the drugs in our study would be higher in non-RCTs than in RCTs. First, it could simply be the result of measurement error or bias, especially in non-RCTs. However, even RCTs were not blind in most of our sample. Second, many drugs with poor responses in phase I or II trials do not advance to testing in phase III trials, whereas those with impressive responses in single-arm trials are much more likely to be later tested in RCTs. The likelihood of regression to the mean is also supported by the clustering of points toward the right side of the scatterplots in Figure 1. Indeed, even hard endpoints such as OS have been known to be misleading in early phase trials, as in the recent case of olaratumab in sarcoma.[12]

The lack of significance in the analysis of rRRs might be reassuring, but it also might have been a result of our conservative study design that included only FDA-approved drugs. By definition, therefore, we excluded all the drugs that performed poorly in RCTs and did not gain FDA approval. Our finding that RRs and DoRs were higher in non-RCTs than in their randomized counterparts in most cases (63% and 87% higher, respectively) is consistent with a previous study on cytotoxic therapies that showed that RRs were higher in phase II versus phase III trials in most cases (81% higher). This previous study did not report on DoR or statistical significance.[2]

These findings have important implications for physicians and patients because of the large number of

**Table 2. RRs and DoRs for Included Drug–Indication Pairs (N=25)**

| Drug | Biomarker | Indication | RR in P III (RCT) | RR in P II or P I (Non-RCT) | DoR RCT (mo) | DoR Non-RCT (mo) | RCT Reference | Non-RCT Reference |
|---|---|---|---|---|---|---|---|---|
| Ado-trastuzumab emtansine[a] | *HER2* | Breast cancer | 43.6% | 34.5% | 12.6 | 7.2 | 22 | 23 |
| Ado-trastuzumab emtansine[a] | *HER2* | Breast cancer | 43.6% | 25.9% | 12.6 | NR | 22 | 24 |
| Afatinib | *EGFR* | NSCLC | 56% | 61% | 11.1 | 12.9 | 25 | 26 |
| Alectinib[a] | *ALK* | NSCLC | 37.5% | 48% | 9.7 | 13.5 | 27 | 28 |
| Alectinib[a] | *ALK* | NSCLC | 37.5% | 50% | 9.7 | 11.2 | 27 | 29 |
| Brigatinib | *ALK* | NSCLC | NA | 54% | NA | 11.1 | NA | 30 |
| Ceritinib | *ALK* | NSCLC | 39.1% | 56% | 6.9 | 8.2 | 31 | 32 |
| Cobimetinib (+ vemurafenib) | *BRAF* | Melanoma | 68% | 87% | NR | 12.5 | 33 | 34 |
| Crizotinib | *ALK* | NSCLC | 65% | 53% | 7.4 | 9.9 | 35 | PROFILE 1005[36] |
| Crizotinib[a] | *ALK* | NSCLC | 65% | 60.8% | 7.4 | 11.3 | 35 | 37 |
| Crizotinib[a] | *ROS1* | NSCLC | NA | 72% | NA | 17.6 | NA | 1 |
| Dabrafenib | *BRAF* | Melanoma | 50% | 59% | 5.5 | 6.6 | 38 | 39 |
| Imatinib | *c-KIT* | GIST | 45% | 53.7% | NA | NR | 40 | 41 |
| Lapatinib (+ capecitabine) | *HER2* | Breast cancer | 22% | NA | NA | NA | 42 | NA |
| Nivolumab | MSI-H or dMMR | Colorectal cancer | NA | 31.1% | NA | NR | NA | 43 |
| Olaparib | *BRCA* | Ovarian cancer | 31% | 31.1% | 6.8 | 7.3 | 44 | 45 |
| Osimertinib[a] | *T790M* | NSCLC | 71% | 70% | 9.7 | 11.4 | 46 | 47 |
| Osimertinib[a] | *T790M* | NSCLC | 71% | 62% | 9.7 | 15.2 | 46 | 48 |
| Pembrolizumab | MSI-H or dMMR | Tissue agnostic | NA | 39.6% | NA | NA | NA | 49 |
| Pertuzumab | *HER2* | Breast cancer | 80.2% | 24.2% | 20.2 | 5.8 | 50 | 51 |
| Rucaparib | *BRCA* | Ovarian cancer | 18.4% | 54% | NR | 9.2 | 52 | 53 |
| Trametinib | *BRAF* | Melanoma | 22% | 25% | 5.5 | 5.7 | 54 | 55 |
| Trastuzumab | *HER2* | Breast cancer | 50% | 50% | 9.1 | NA | 56 | 57 |
| Trastuzumab | *HER2* | Gastric cancer | 47% | 53.3% | 6.9 | 7.3 | 58 | 59 |
| Vemurafenib | *BRAF* | Melanoma | 48% | 52% | 5.5 | 6.7 | 60 | 61 |

Abbreviations: dMMR, deficiency in mismatch repair; DoR, duration-of-response rate; GIST, gastrointestinal stromal tumor; MSI-H, microsatellite instability-high; NA, not available; NR, not reached; non-RCT, single-arm nonrandomized controlled trial; NSCLC, non–small cell lung cancer; P, clinical trial phase; RCT, randomized controlled trial; RR, response rate.
[a]Drug–indication pairs with >1 non-RCT were matched with the same P III RCT.

precision oncology drugs approved in recent years based on durable responses in early-stage trials. Physicians and patients considering whether to use one of these drugs based on these metrics should recognize that the DoR predicted in the preapproval trial may decrease in subsequent studies.

These results also have important implications for clinical investigators in oncology. Responses are important markers of drug activity and serve as endpoints in early trials that are then used for making decisions about whether to proceed with phase III RCTs.[13] One investigation of cancer drugs that failed in phase III RCTs showed that 19% of such drugs had a negative phase II trial result.[14] In a recent editorial, the FDA questioned the rationale for conducting phase III RCTs of cancer drugs with poor responses in single-arm trials.[15] Resources could be better channeled by conducting RCTs of drugs that showed responses in non-RCTs impressive enough to remain useful even if they might be expected to decrease by a certain percentage in RCTs.

DRRs are not necessarily markers of clinical benefit themselves.[16] It is worth noting that responses observed even in RCTs are usually poor surrogates for the true clinical benefits—improved survival or quality of life resulting from use of cancer drugs.[17] For this reason, RCTs usually use RR or DoR only as a secondary endpoint, and it is unusual to see a phase III RCT in oncology with a primary endpoint of RR or DoR. Our analysis also showed that the RRs and DoRs derived from non-RCTs have weak and nonsignificant associations with OS derived from RCTs, the ultimate marker of clinical benefit.

A recent analysis by the FDA of 10 RCTs in patients with metastatic melanoma treated with targeted therapies, immunotherapies, and chemotherapies has suggested that depth of response may be a better marker for assessing response than RR or DoR.[18] In this analysis, patients with a deep response (>75% response) were associated with the best HR for OS. Further analyses with

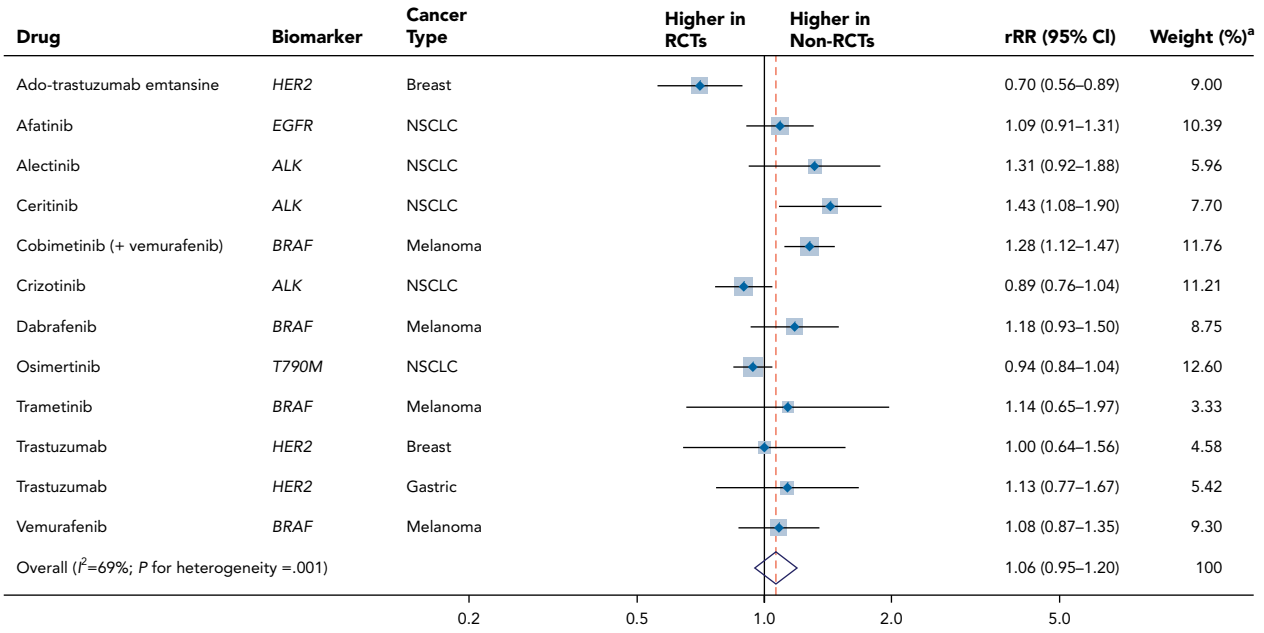



**Figure 2.** Forest plot of rRRs across drug–indication pairs.
Abbreviations: NSCLC, non–small cell lung cancer; RCTs, randomized controlled trials; rRRs, ratio of response rates.
[a]Weights are derived from random-effects analysis.

individual patient data are necessary to investigate the validity of this novel marker of response for correlation with OS across tumor types.

The FDA has an Accelerated Approval Program in which drugs can receive approval based on surrogate measures, but results must be confirmed with a clinical endpoint in a postapproval confirmatory trial.[19] The FDA then has the ability to withdraw the approval should the results not be replicated. The same postapproval authorities are not as strong for regular approvals. Although some precision oncology drugs have received accelerated approval based on durable responses in early non-RCTs, some of these drugs have also received full approval based on single-arm trials.[20] This reflects the increasing

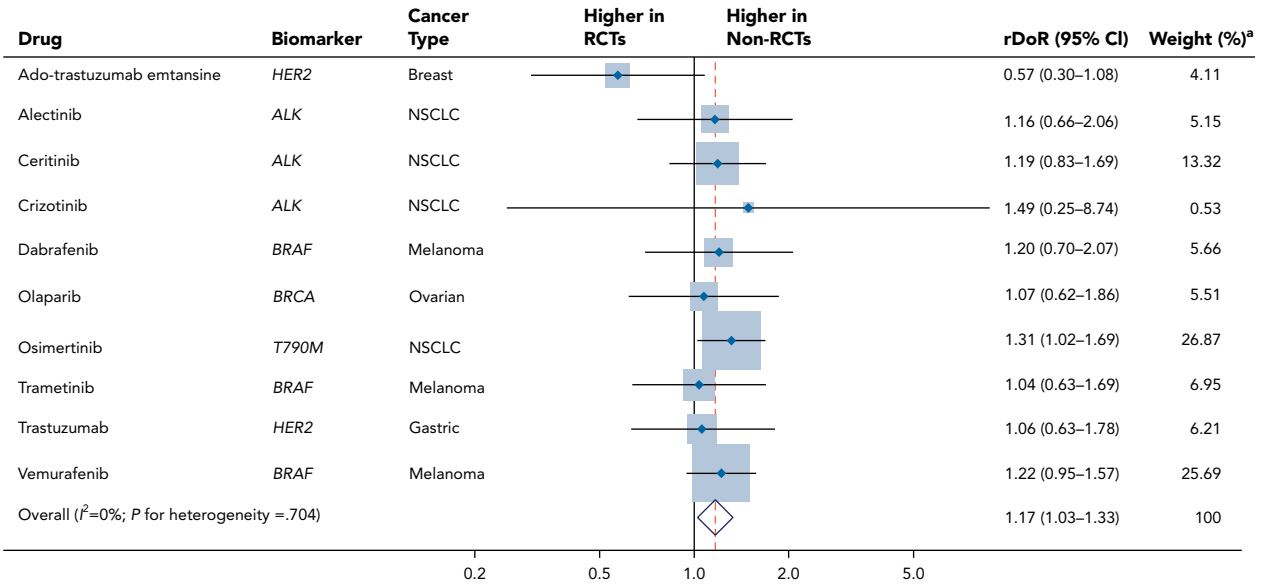



**Figure 3.** Forest plot of rDoRs across drug–indication pairs.
Abbreviations: NSCLC, non–small cell lung cancer; RCTs, randomized controlled trials; rDoRs, ratio of durations of response.
[a]Weights are derived from random-effects analysis.

confidence of the regulatory authorities in durable responses derived from single-arm trials as markers of clinical benefit. Given our study results showing the subsequent changes associated with observed responses in early single-arm trials, it seems prudent to grant accelerated approval—not regular approval—for biomarker-based drugs approved based on DRRs in single-arm trials to better ensure that the confirmatory RCTs are completed.

Our study has several limitations. First, we focused only on approved drugs, excluding all the failed drugs, because there is no database for all biomarker-based drugs, and many failed drugs also suffer from publication bias. As a result, our sample size was small. Furthermore, many published reports did not provide CIs for RRs and DoRs. Second, there is some heterogeneity in patient characteristics and eligibility criteria between non-RCTs and follow-up RCTs, although we tried to include the closest matching pair. The heterogeneous population between RCTs and non-RCTs for the same drug in the same disease is more prominent for drugs that have received accelerated approval based on non-RCTs. These drugs must be tested in confirmatory trials. However, when confirmatory RCTs are conducted, they are usually conducted in a slightly different patient population from that in the single-arm preapproval trial, because this could allow both full approval and an expanded indication. However, because meta-epidemiologic studies do not make treatment recommendations but rather analyze policy issues across the trials, such heterogeneity is inevitable. Third, we analyzed the correlation of RRs or DoRs in non-RCTs with OS in RCTs, although there has been some debate about the appropriate endpoint for targeted drugs.[21] Because ours is a meta-epidemiologic study across tumor types, conclusions for specific tumor types may be different.

## Conclusions

Many biomarker-based cancer drugs have improved outcomes for patients with cancer and represent important advances in the field of oncology. However, durable responses in non-RCTs may not be replicated in subsequent RCTs in the subset of drug–indication pairs that move on to phase III trials. Caution must be exercised when interpreting durable responses derived from single-arm trials of cancer drugs for regulatory and clinical purposes.

Submitted May 15, 2019; accepted for publication August 7, 2019.

**Author contributions:** *Study concept:* Gyawali, Kesselheim. *Statistical analysis:* D'Andrea, Franklin. *Writing, editing, and approval of the manuscript:* All authors.

**Disclosures:** The authors have disclosed that they have not received any financial consideration from any person or organization to support the preparation, analysis, results, or discussion of this article.

**Funding:** Work on this project was funded by Arnold Ventures.

**Disclaimer:** The funders 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; or the decision to submit the manuscript for publication.

**Correspondence:** Bishal Gyawali, MD, PhD, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute, 10 Stuart Street, Level 2, Queen's University, Kingston, ON, Canada K7L 3N6.
Email: bg.bishalgyawali@gmail.com

## References

1. Shaw AT, Ou SH, Bang YJ, et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med 2014;371:1963–1971.
2. Zia MI, Siu LL, Pond GR, et al. Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. J Clin Oncol 2005;23:6982–6991.
3. Chabner BA. Approval after phase I: ceritinib runs the three-minute mile. Oncologist 2014;19:577–578.
4. Selaru P, Tang Y, Huang B, et al. Sufficiency of single-arm studies to support registration of targeted agents in molecularly selected patients with cancer: lessons from the clinical development of crizotinib. Clin Transl Sci 2016;9:63–73.
5. Sharma MR, Schilsky RL. Role of randomized phase III trials in an era of effective targeted therapies. Nat Rev Clin Oncol 2011;9:208–214.
6. Schwaederle M, Zhao M, Lee JJ, et al. Association of biomarker-based treatment strategies with response rates and progression-free survival in refractory malignant neoplasms: a meta-analysis. JAMA Oncol 2016;2:1452–1459.
7. Chakiba C, Grellety T, Bellera C, et al. Encouraging trends in modern phase 1 oncology trials. N Engl J Med 2018;378:2242–2243.
8. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med 2017;22:139–142.
9. US Food and Drug Administration. Table of Pharmacogenomic Biomarkers in Drug Labeling. Available at: https://www.fda.gov/drugs/scienceresearch/ucm572698.htm. Accessed March 23, 2018.
10. Korn EL, Sachs MC, McShane LM. Statistical controversies in clinical research: assessing pathologic complete response as a trial-level surrogate end point for early-stage breast cancer. Ann Oncol 2016;27:10–15.
11. US Food and Drug Administration. List of cleared or approved companion diagnostic devices (in vitro and imaging tools). Available at: https://www.fda.gov/medical-devices/vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-vitro-and-imaging-tools. Accessed July 2, 2019.
12. Eli Lilly & Co. Lilly reports results of phase 3 soft tissue sarcoma study of LARTRUVO [news release]. Indianapolis, IN: Lilly Oncology; January 18, 2019. Available at: https://investor.lilly.com/news-releases/news-release-details/lilly-reports-results-phase-3-soft-tissue-sarcoma-study. Accessed January 22, 2019.
13. Vickers AJ, Ballen V, Scher HI. Setting the bar in phase II trials: the use of historical data for determining "go/no go" decision for definitive phase III testing. Clin Cancer Res 2007;13:972–976.
14. Gyawali B, Addeo A. Negative phase 3 randomized controlled trials: why cancer drugs fail the last barrier? Int J Cancer 2018;143:2079–2081.
15. Gormley NJ, Pazdur R. Immunotherapy combinations in multiple myeloma—known unknowns. N Engl J Med 2018;379:1791–1795.
16. George SL. Response rate as an endpoint in clinical trials. J Natl Cancer Inst 2007;99:98–99.
17. Haslam A, Hey SP, Gill J, et al. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate endpoints and overall survival in oncology. Eur J Cancer 2019;106:196–211.
18. Osgood C, Mulkey F, Mishra-Kalyani PS, et al. FDA analysis of depth of response (DpR) and survival across 10 randomized controlled trials in patients with previously untreated unresectable or metastatic melanoma

(UMM) by therapy type [abstract]. J Clin Oncol 2019;37(Suppl):Abstract 9508.

19. Gyawali B, Kesselheim AS. Reinforcing the social compromise of accelerated approval. Nat Rev Clin Oncol 2018;15:596–597.
20. DeLoughery EP, Prasad V. The US Food and Drug Administration's use of regular approval for cancer drugs based on single-arm studies: implications for subsequent evidence generation. Ann Oncol 2018;29: 527–529.
21. Robinson AG, Booth CM, Eisenhauer EA. Progression-free survival as an end-point in solid tumours – perspectives from clinical trials and clinical practice. Eur J Cancer 2014;50:2303–2308.
22. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012;367:1783–1791.
23. Krop IE, LoRusso P, Miller KD, et al. A phase II study of trastuzumab emtansine in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer who were previously treated with trastuzumab, lapatinib, an anthracycline, a taxane, and capecitabine. J Clin Oncol 2012;30:3234–3241.
24. Burris HA III, Rugo HS, Vukelja SJ, et al. Phase II study of the antibody drug conjugate trastuzumab-DM1 for the treatment of human epidermal growth factor receptor 2 (HER2)-positive breast cancer after prior HER2-directed therapy. J Clin Oncol 2011;29:398–405.
25. Sequist LV, Yang JC, Yamamoto N, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 2013;31:3327–3334.
26. Yang JC, Shih JY, Su WC, et al. Afatinib for patients with lung adenocarcinoma and epidermal growth factor receptor mutations (LUX-Lung 2): a phase 2 trial. Lancet Oncol 2012;13:539–548.
27. Novello S, Mazières J, Oh IJ, et al. Alectinib versus chemotherapy in crizotinib-pretreated anaplastic lymphoma kinase (ALK)-positive non-small-cell lung cancer: results from the phase III ALUR study. Ann Oncol 2018;29:1409–1416.
28. Shaw AT, Gandhi L, Gadgeel S, et al. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. Lancet Oncol 2016;17:234–242.
29. Ou SH, Ahn JS, De Petris L, et al. Alectinib in crizotinib-refractory ALK-rearranged non-small-cell lung cancer: a phase II global study. J Clin Oncol 2016;34:661–668.
30. Kim DW, Tiseo M, Ahn MJ, et al. Brigatinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small-cell lung cancer: a randomized, multicenter phase II trial. J Clin Oncol 2017;35:2490–2498.
31. Shaw AT, Kim TM, Crinò L, et al. Ceritinib versus chemotherapy in patients with ALK-rearranged non-small-cell lung cancer previously given chemotherapy and crizotinib (ASCEND-5): a randomised, controlled, open-label, phase 3 trial. Lancet Oncol 2017;18:874–886.
32. Shaw AT, Kim DW, Mehra R, et al. Ceritinib in ALK-rearranged non-small-cell lung cancer. N Engl J Med 2014;370:1189–1197.
33. Larkin J, Ascierto PA, Dréno B, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. N Engl J Med 2014;371:1867–1876.
34. Ribas A, Gonzalez R, Pavlick A, et al. Combination of vemurafenib and cobimetinib in patients with advanced BRAF(V600)-mutated melanoma: a phase 1b study. Lancet Oncol 2014;15:954–965.
35. Shaw AT, Kim DW, Nakagawa K, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med 2013;368:2385–2394.
36. Kim DW, Ahn MJ, Shi Y, et al. Results of a global phase II study with crizotinib in advanced ALK-positive non-small cell lung cancer (NSCLC) [abstract]. J Clin Oncol 2012;30(Suppl):Abstract 7533.
37. Camidge DR, Bang YJ, Kwak EL, et al. Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: updated results from a phase 1 study. Lancet Oncol 2012;13:1011–1019.
38. Hauschild A, Grob JJ, Demidov LV, et al. Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. Lancet 2012;380:358–365.
39. Ascierto PA, Minor D, Ribas A, et al. Phase II trial (BREAK-2) of the BRAF inhibitor dabrafenib (GSK2118436) in patients with metastatic melanoma. J Clin Oncol 2013;31:3205–3211.
40. Cohen MH, Farrell A, Justice R, Pazdur R. Approval summary: imatinib mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors. Oncologist 2009;14:174–180.
41. Demetri GD, von Mehren M, Blanke CD, et al. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 2002;347:472–480.
42. Geyer CE, Forster J, Lindquist D, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med 2006;355: 2733–2743.
43. Overman MJ, McDermott R, Leach JL, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. Lancet Oncol 2017;18:1182–1191.
44. Kaye SB, Lubinski J, Matulonis U, et al. Phase II, open-label, randomized, multicenter study comparing the efficacy and safety of olaparib, a poly (ADP-ribose) polymerase inhibitor, and pegylated liposomal doxorubicin in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer. J Clin Oncol 2012;30:372–379.
45. Kaufman B, Shapira-Frommer R, Schmutzler RK, et al. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. J Clin Oncol 2015;33:244–250.
46. Mok TS, Wu YL, Ahn MJ, et al. Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. N Engl J Med 2017;376:629–640.
47. Goss G, Tsai CM, Shepherd FA, et al. Osimertinib for pretreated EGFR Thr790Met-positive advanced non-small-cell lung cancer (AURA2): a multicentre, open-label, single-arm, phase 2 study. Lancet Oncol 2016; 17:1643–1652.
48. Yang JC, Ahn MJ, Kim DW, et al. Osimertinib in pretreated T790M-positive advanced non-small-cell lung cancer: AURA study phase II extension component. J Clin Oncol 2017;35:1288–1296.
49. Lemery S, Keegan P, Pazdur R. First FDA approval agnostic of cancer site – when a biomarker defines the indication. N Engl J Med 2017;377: 1409–1412.
50. Baselga J, Cortés J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 2012;366:109–119.
51. Baselga J, Gelmon KA, Verma S, et al. Phase II trial of pertuzumab and trastuzumab in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer that progressed during prior trastuzumab therapy. J Clin Oncol 2010;28:1138–1144.
52. Coleman RL, Oza AM, Lorusso D, et al. Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017;390:1949–1961.
53. Balasubramaniam S, Beaver JA, Horton S, et al. FDA approval summary: rucaparib for the treatment of patients with deleterious BRCA mutation-associated advanced ovarian cancer. Clin Cancer Res 2017;23:7165–7170.
54. Flaherty KT, Robert C, Hersey P, et al. Improved survival with MEK inhibition in BRAF-mutated melanoma. N Engl J Med 2012;367:107–114.
55. Kim KB, Kefford R, Pavlick AC, et al. Phase II study of the MEK1/MEK2 inhibitor trametinib in patients with metastatic BRAF-mutant cutaneous melanoma previously treated with or without a BRAF inhibitor. J Clin Oncol 2013;31:482–489.
56. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 2001;344:783–792.
57. Tedesco KL, Thor AD, Johnson DH, et al. Docetaxel combined with trastuzumab is an active regimen in HER-2 3+ overexpressing and fluorescent in situ hybridization-positive metastatic breast cancer: a multi-institutional phase II trial. J Clin Oncol 2004;22:1071–1077.
58. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet 2010;376:687–697.
59. Chua C, Tan IB, Yamada Y, et al. Phase II study of trastuzumab in combination with S-1 and cisplatin in the first-line treatment of human epidermal growth factor receptor HER2-positive advanced gastric cancer. Cancer Chemother Pharmacol 2015;76:397–408.
60. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;364:2507–2516.
61. Ribas A, Kim KB, Schuchter LM, et al. BRIM-2: an open-label, multicenter phase II study of vemurafenib in previously treated patients with BRAF V600E mutation-positive metastatic melanoma [abstract]. J Clin Oncol 2011;29(Suppl):Abstract 8509.

See JNCCN.org for supplemental online content.

JNCCN

National Comprehensive Cancer Network®

JOURNAL OF THE NATIONAL COMPREHENSIVE CANCER NETWORK

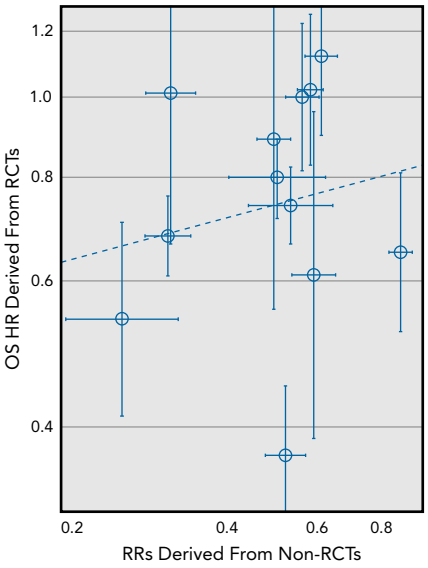
Supplemental online content for:

# Response Rates and Durations of Response for Biomarker-Based Cancer Drugs in Nonrandomized Versus Randomized Trials

Bishal Gyawali, MD, PhD; Elvira D'Andrea, MD, MPH; Jessica M. Franklin, PhD; and Aaron S. Kesselheim, MD, JD, MPH

*J Natl Compr Canc Netw 2020;18(1):36–43*

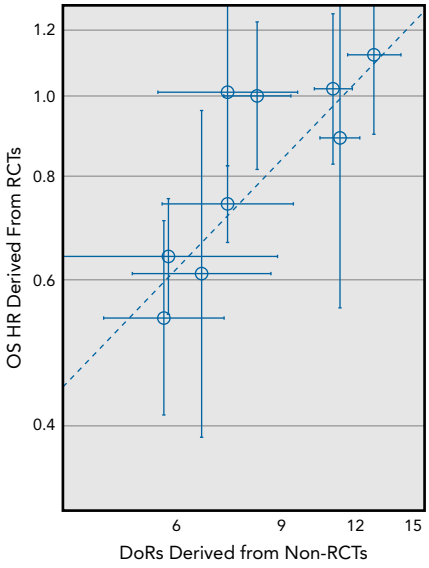
| | Estimate | SE |
|---|---|---|
| Mean of log RRs | −0.7002 | 0.0934 |
| Intercept for OS model | −0.1843 | 0.2327 |
| Slope of OS model | 0.1652 | 0.3084 |
| Variance of log RRs | 0.0974 | 0.0467 |
| Variance of log OS HRs | 0.0587 | 0.0517 |




**eFigure 1.** Correlation of RRs derived from non-RCTs with OS derived from RCTs. For the analysis of RR, there were 12 non-RCTs with RRs measured and corresponding RCTs with OS HRs measured. We fit the model to the data derived from these trials on the logarithmic scale. Model results are presented in the table above the figure. The association between logarithmic RR and logarithmic OS HR was weak: β=0.165 (95% CI, −0.439 to 0.770).

Abbreviations: HR, hazard ratio; OS, overall survival; RCT, randomized controlled trial; RRs, response rates.

| | Estimate | SE |
|---|---|---|
| Mean of log RRs | 2.1644 | 0.1051 |
| Intercept for OS model | −1.8163 | 0.8659 |
| Slope of OS model | 0.7460 | 0.3949 |
| Variance of log RRs | 0.0655 | 0.0470 |
| Variance of log OS HRs | 0.0000 | NA |




**eFigure 2.** Correlation of DoRs derived from non-RCTs with OS derived from RCTs. For the analysis of DoR, there were 9 non-RCTs with DoR measured and corresponding RCTs with OS HRs measured. We fit the model to the data derived from these trials on the logarithmic scale. Model results are presented in the table above the figure. The association between logarithmic DoR and logarithmic OS HR was much stronger but still not statistically significant: β=0.746 (95% CI, −0.028 to 1.520).

Abbreviations: DoR, duration of response; HR, hazard ratio; NA, not applicable; OS, overall survival; RCT, randomized controlled trial.