OXFORD

Optimism Bias in the Design of Phase III Randomized Control Trials Evaluating PD-1/PD-L1 Targeting Monoclonal Antibodies

Laith Al-Showbaki[1,2], Fahad A. Almugbel[1,2,3], Husam A. Alqaisi[1,2], Eitan Amir[1,2], Eric X. Chen[1,2,*,]

[1]Department of Medical Oncology and Hematology, Princess Margaret Cancer Center, University Health Network, Toronto, Canada
[2]Division of Medical Oncology, Faculty of Medicine, University of Toronto, Toronto, Canada
[3]Medical Oncology Section, King Abdullah Center for Oncology and Liver Disease, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

*Corresponding author: Eric X. Chen, Department of Medical Oncology and Hematology, Princess Margaret Cancer Center, University Health Network, 7-824, 700 University Avenue, Toronto, ON M5G 2M9, Canada. Tel: 416-946-2263; Fax: 416-946-4467; Email: eric.chen@uhn.ca

Abstract

Background: Many randomized control trials (RCTs) evaluating programmed death receptor-1 (PD-1)/programmed death ligand-1 (PD-L1) targeting monoclonal antibodies (mAbs) have been completed or are in progress. We examined hypothesized hazard ratios (HHRs) and observed hazard ratios (OHRs) from published RCTs evaluating these mAbs.

Methods: Publications of RCTs evaluating at least one PD-1/PD-L1 targeting mAbs approved by the US Food and Drug Administration were identified through PubMed searches. The primary reports of RCTs were retrieved. Two investigators extracted HHR, OHR for the primary endpoint among other data elements independently. The differences (ΔHR) in HHR and OHR were analyzed statistically. A separate search was conducted for secondary reports after longer follow-ups, the updated OHR was extracted.

Results: Forty-nine RCTs enrolling 36 867 patients were included. The mean HHR and OHR were 0.672 and 0.738 respectively. The mean ΔHR was 0.067 (range: –0.300 to 0.895; 95% confidence interval (CI), 0.003-0.130). HHR was met or exceeded in 22 (45%) RCTs. OHR was ≥ 1.0 in 6 RCTs (12%). PD-L1 expression was not associated with the magnitude of effect. Of 18 RCTs with follow-up reports, the magnitude of benefit decreased in 8 RCTs with extended follow-ups.

Conclusion: The majority of published RCTs evaluating PD-1/PD-L1 targeting mAbs did not achieve their hypothesized magnitude of benefit. The optimism bias requires attention from the cancer clinical research community given the number of these agents in development and the intense interest in evaluating these agents in a variety of disease settings.

Key words: outcomes; randomized control trials; PD-1/PD-L1; hypothesized; observed; hazard ratio

---

Implications for Practice

There are many monoclonal antibodies targeting programmed death receptor-1/programmed death ligand-1 in development, and several hundred large phase III clinical trials have been conducted or are in progress. In phase III clinical trials evaluating these agents published so far, the majority of trials did not achieve their hypothesized magnitudes of benefit. Researchers seem to be overly optimistic in designing these phase III clinical trials. This optimism bias requires attention from the cancer clinical research community given the number of these agents in development and the intense interest in evaluating these agents in a variety of disease settings.

Background

The development of immune checkpoint inhibitors has revolutionized cancer treatment in the last decade. Since the initial approval of ipilimumab, a monoclonal antibody (mAb) targeting cytotoxic T-cell lymphocyte antigen-4 in 2011, several agents targeting either programmed death receptor-1(PD-1) or programmed death ligand-1 (PD-L1) have been approved and quickly became the standards of care in 16 different types of cancer, including melanoma, non–small cell lung cancer (NSCLC), and renal cell carcinoma.[1] Globally, 9 mAbs targeting PD-1/PD-L1 have been approved by various regulatory agencies. The development of these agents is supported by unprecedented numbers of clinical trials. It has been reported that 3362 trials have been initiated to evaluate PD-1/PD-L1 targeting mAbs alone or in combination since 2006.[2] Approximately 3000 trials were active as of September 2019, expecting to recruit over 500 000 patients. Among of these trials, as of January 2021, approximately 500 are phase III randomized control trials (RCTs) according to the commonly used trial registration site clinicaltrials.gov.

Results from properly designed and conducted phase III RCTs frequently provide the basis for regulatory approvals

and are considered the cornerstone of evidence-based treatment guidelines and clinical decisions. However, each RCT requires the enrollment of hundreds to thousands of patients necessitating the participation of many centers from different countries, and long durations of treatment and follow-up especially if overall survival is the primary endpoint. It is critical that RCTs are designed appropriately to maximize their chances of meeting their primary endpoints and reduce financial and opportunity costs.

A critical consideration in designing RCTs is the potential efficacy of interventions under investigation.[3] The potential efficacy not only can affect the desired number of patients to enroll in an RCT but also influence funding decisions. Treatments with a larger potential efficacy may be prioritized over those with smaller efficacies.[4] Recent research suggested that experts may be influenced by biases when designing and conducting RCTs, among which is the optimism bias. Optimism bias is defined as the "unwarranted belief in the efficacy of new therapies."[5] For example, a recent review showed that an effect size at least as large as the one projected in the protocol was observed in only 9.8% of trials sponsored by National Clinical Trials Network between 2007 and 2017.[6]

Given the intense interests and enthusiasms in developing PD-1/PD-L1 targeting mAbs in a wide variety of malignancies, we conducted a systemic review of the literature to investigate the optimism bias in RCTs evaluating these agents.

Methods

Literature Search and Data Collection

The National Library of Medicine online database (www.pubmed.ncbi.nlm.nih.gov) was searched in October 2020 for publications of phase III RCTs conducted in patients with cancer. A separate search for each Food and Drug Administration (FDA)-approved anti-PD-1/PD-L1 mAb (Drugs@FDA) was conducted. Drugs included in the search were pembrolizumab, nivolumab, atezolizumab, avelumab, durvalumab, and cemiplimab. Results from each search were directly imported to Covidence (www.covidence.org). Each abstract was screened. Publications were excluded for the following reasons: pediatric studies (participants were <18 years of age), no time-to-event primary endpoint, self-reported as phase II studies, meta-analysis or other forms of pooled analysis, and secondary reports of previously published studies, such as sub-group analysis or reports after longer durations of follow-ups. Full publications were then retrieved and data extraction was performed by 2 investigators independently for each publication. A third investigator reviewed all extracted data independently and discrepancy was resolved by consensus or arbitrated by the third investigator.

Characteristics of each RCT were extracted, including publication year, journal of publication, source of funding (industry or government), tumor type, treatment setting (adjuvant or metastatic disease), types of experimental treatments (monotherapy or combination with other agents), control arm, total number of arms, total of number of patients enrolled, PD-L1 positivity used in patient selection if any, hypothesized hazard ratio (HHR), and observed hazard ratio (OHR) of the primary outcome. The primary endpoint was taken as stated explicitly in the publication, or based on the sample size calculation if it was not stated explicitly or if there were more than one primary endpoint. For trials that

randomized patients to more than 2 treatment arms, only results based on the stated primary endpoint or based on the sample size calculation were included. The study protocol was checked if the statistical assumption was not detailed in the primary publication.

A separate literature search was conducted in February 2021 specifically for publications reporting updated outcomes after longer patient follow-ups for each RCT included in this analysis. The updated OHR (uOHR) was extracted from these publications.

Statistical Analysis

The difference between HHR and OHR (ΔHR = OHR – HHR) was calculated for each RCT. A negative ΔHR value indicates that the magnitude of benefit was greater than what was hypothesized. In contrast, a positive ΔHR denotes that the observed magnitude of benefit is lower than the hypothesized. Mean and confidence intervals were calculated. Differences between/among subgroups were compared with t-tests or analysis of variances. Distributions in HHR and OHR were compared with the Kolmogorov-Smirnov test. A P-value < .05 was considered to be of statistical significance. All statistical analyses were performed using Prism version 9.0 (GraphPad Software, San Diego, CA).

Results

A total of 49 RCTs with 36 837 patients enrolled were included in this analysis (Fig. 1, and Supplementary Table S1), all were industry sponsored. These included 17 trials for pembrolizumab, 14 trials for nivolumab, 12 trials for atezolizumab, 4 trials for avelumab, and 2 trials for durvalumab (Table 1). There were no eligible RCTs for cemiplimab. The majority of RCTs (91%) were in the palliative settings. The most common disease sites were NSCLC (16 RCTs), followed by melanoma (8 RCTs) and renal cell carcinoma (5 RCTs). Single agent PD-1/PD-L1 targeting mAb was evaluated in 27 RCTs, while combinations with ipilimumab were evaluated in 3 RCTs, with chemotherapy in 10 RCTs, and with other target agents (such as bevacizumab, tyrosine kinase inhibitors, and others) in 9 RCTs.
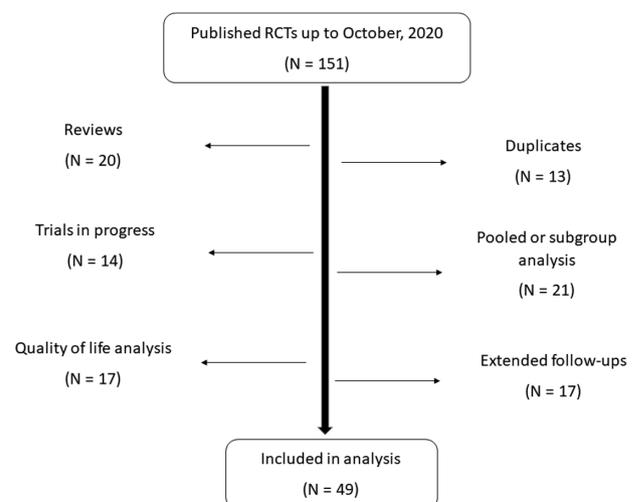

Figure 1. The PRISMA flowchart.

Table 1. Characteristics of included RCTs.

Characteristic	Number
Disease site	
Lung cancer, non–small cell	16
Melanoma	8
Renal cell carcinoma	5
Urothelial carcinoma	4
Gastroesophageal	4
Squamous cell carcinoma of head and neck	3
Others	9
Agents	
Pembrolizumab	17
Nivolumab	14
Atezolizumab	12
Avelumab	4
Durvalumab	2
Treatment setting	
Palliative: first line	26
Palliative: second line	15
Palliative: third line or beyond	4
Adjuvant	3
Neoadjuvant	1
Year of initiation	
2012	4
2013	4
2014	10
2015	14
2016	13
2017	2
2018	2
PD-L1 positive patients only	
Yes	10
No	39
Primary endpoints	
Overall survival	30
Progression-free survival	16
Recurrence free survival	3
Control arm	
Placebo/best supportive care	12
Active treatment	37

Abbreviations: PD-L1, programmed death ligand-1; RCTs, randomized control trials.

The mean HHR was 0.672 (range: 0.50 to 0.78; 95% CI, 0.655-0.689), while the mean OHR was 0.738 (range: 0.420 to 1.530; 95% CI, 0.679-0.798). The distribution for HHR was significantly wider than that for OHR (P = .01, Fig. 2). For 16 RCTs in NSCLC, HHR ranged from 0.50 to 0.74 and OHR ranged from 0.52 to 1.15.

The mean ΔHR was 0.067 (range: –0.300 to 0.895; 95% CI, 0.003-0.130) (Supplementary Fig. S1. Among 49 RCTs, OHRs were equal to (defined as having the same numbers as reported in each publication) or lower than corresponding HHRs in 22 (45%) RCTs (Fig. 3). Seventeen of these 22 RCTs were concentrated in 3 most common disease types,

9/16 (56%) for NSCLC, 5/8 (63%) for melanoma, and 3/5 (60%) for renal cell carcinoma. For 20 RCTs conducted in other cancer types, OHRs were equal or lower than HHRs in only 5 (25%), 2 in advanced urothelial carcinoma, one each in hepatocellular carcinoma, gastroesophageal cancer, and neoadjuvant breast cancer. The mean ΔHR was 0.004 (range: –0.300 to 0.440; 95% CI, –0.062-0.070) in 29 RCTs in NSCLC, melanoma, and renal cell carcinoma, and 0.157 in other cancer types (range: –0.140 to 0.895; 95% CI: 0.038 to 0.277) (P < .02).

OHRs were equal or lower than HHRs in 7/17 (41%) RCTs evaluating pembrolizumab, 7/14 (50%) for nivolumab, 5/12 (42%) for atezolizumab, and 3/6 (50%) for durvalumab and avelumab.

The OHR was ≥ 1 in 6 RCTs (12%) with 3326 patients enrolled. Two RCTs in multiple myeloma had the largest ΔHRs of 0.895 and 0.57, respectively. These 2 RCTs were the only RCTs in hematological malignancies in this analysis. Excluding these two RCTs, the mean ΔHR was 0.038 (range: –0.300 to 0.440; 95% CI, –0.012-0.089).

Only 4 RCTs were conducted in the adjuvant/neoadjuvant setting. Of 45 RCTs in the metastatic setting, 26 were for previously untreated patients, 15 for one line of prior therapy, and 4 for 2 or more lines of prior therapies. The mean ΔHR increased with more prior therapies, being 0.017 (range: –0.300 to 0.57; 95% CI: –0.064-0.098) in the first-line setting, 0.106 (range: –0.05 to 0.300; 95% CI, 0.043-0.168) in the second-line setting, and 0.426 (range: –0.020 to 0.895; 95% CI: –0.169-1.022) in more advanced settings (Fig. 4).

HHR, OHR, and ΔHR by the year of each RCT initiation were shown in Fig. 5. HHR remained relatively constant through time. Excluding 4 RCTs initiated in 2017 and 2018, OHR increased over time (P = .047) (Fig. 5).

In 10 RCTs (20%), eligible patients needed to have PD-L1 expression. There were no differences in HHR, OHR, and ΔHR between PD-L1 selected and unselected RCTs.

The majority of 49 RCTs (78%) were published in two journals, New England Journal of Medicine (NEJM) (26) and Lancet (12). OHRs were equal or lower than corresponding HHR in 73% of RCTs published in NEJM, 8% in Lancet, and 9% other journals.

There were 18 secondary publications with updated results after longer patient follow-ups (Supplementary Table S2). The mean OHR and uOHR were 0.631 (range: 0.420 to 0.730; 95% CI, 0.581-0.682) and 0.642 (range: 0.420 to 0.780; 95% CI, 0.587-0.698) respectively. The uOHR was higher than OHR in 8 RCTs (44%), the difference between uOHR and OHR in all 18 RCTs ranged from –0.04 to 0.07.

Discussion

Targeting PD-1/PD-L1 has resulted in significant improvements in patient outcomes over the last decade. It is surprising that OHR exceeded HHR in only 45% published of RCTs evaluating PD-1/PD-L1 targeting mAbs. The distribution of OHR was significantly wider than that of HHR, with OHRs being ≥ 1 in 6 RCTs. Furthermore, most RCTs with a higher observed magnitude of benefit than hypothesized were concentrated in 3 of the so-called immune "hot" tumor types, NSCLC, melanoma, and renal cell carcinoma, while the hypothesized benefit was achieved or exceeded only in 25% of RCTs in other cancer types. The presence of optimism bias in RCTs evaluating PD-1/PD-L1 targeting mAbs is consistent
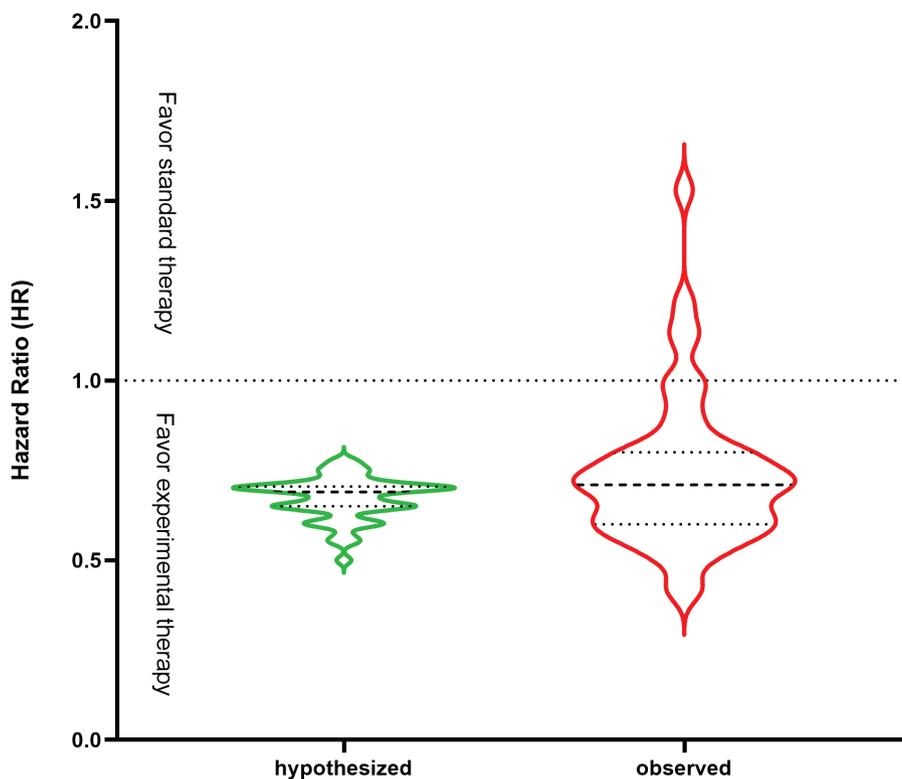
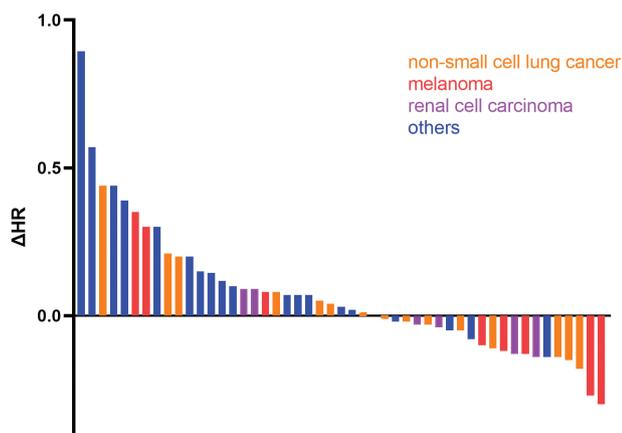**Figure 2.** Violin plots comparing distributions of hypothesized and observed hazard ratios among 49 RCTs.

Figure 3. Waterfall plot showing difference in hypothesized and observed hazard ratios (ΔHR) in 49 RCTs.
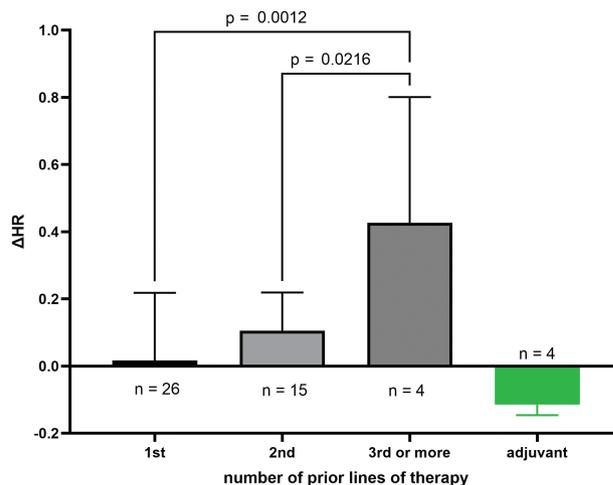

Figure 4. Differences in hypothesized and observed hazard ratios (ΔHR) by prior therapies.

with earlier reports in cancer clinical trials and other disease settings.[6-9]

Despite the importance of HHR in determining the required number of patients to be enrolled in an RCT, there is minimal research and little consensus on the best approach to determine HHR. HHRs are most often determined by expert opinions, reflecting their subjective clinical judgments. However, individual experts are known to be inaccurate in predicting the efficacy of experimental cancer therapies, even those evaluated in RCTs.[4,10] In this group of contemporary RCTs in the most exciting area of cancer research, the OHR was ≥ 1 in 12% RCTs. Among these trials, 3326 patients derived little to no benefit from treatment. Others have reported similar findings, indicating that this issue

has not been adequately addressed by the cancer research community.[6,7,11]

Given the intense interest in establishing the role of PD-1/PD-L1 targeting mAbs in multiple cancer sites, the fierce competition for potential patients to be enrolled in RCTs, and the financial costs associated with RCTs, it is critical that RCTs be designed with appropriate but perhaps more importantly realistic expectations of magnitudes of benefits. We would argue that the largest magnitudes of benefits for PD-1/PD-L1 targeting mAbs are for the immune "hot" cancers.[12] As investigators and pharmaceutical companies begin to explore the role of PD-1/PD-L1 targeting mAbs in other cancers, magnitudes of benefits from these agents will inevitably decrease without
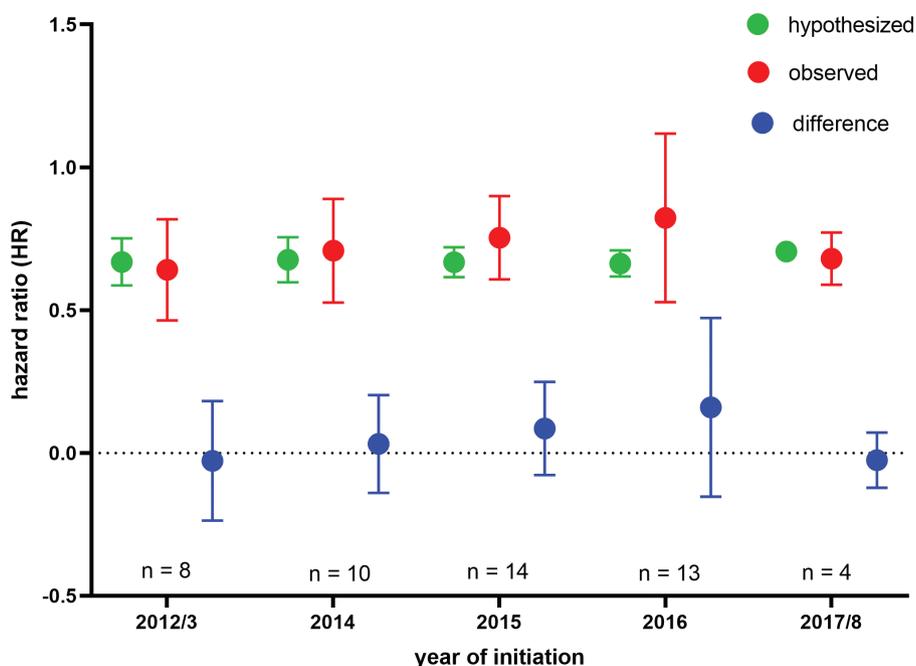
Figure 5. Hypothesized, observed hazard ratios and difference (ΔHR) by year of initiations.

additional breakthroughs in our understanding of immune regulation, tumor microenvironment, and biomarker validations. The mean HHR in this study, 0.672, is likely not achievable for the majority of future RCTs. This conclusion is further supported by the fact that OHR seemed to increase with time.

American Society of Clinical Oncology (ASCO) working groups have recommended meaningful goals in clinical trials for 4 common cancer types.[13] For NSCLC, the target HR was 0.76 (0.77 for squamous cell lung cancer) to 0.8. For 16 RCTs in NSCLC included in this analysis, HHR ranged from 0.50 to 0.74, all better than the recommended target. OHR was < 0.8 in 14 of these 16 RCTs, 2 RCTs had OHR of 0.90 and 1.15, respectively.

It is interesting that enrolling only PD-L1-positive patients did not result in a better approximation of HHR. This is likely related to the small number of published RCTs so far, varying cutoff values in defining positive PD-L1 expression and different antibodies/techniques deployed in RCTs.[14] RCTs published in NEJM significantly outperformed RCTs published in other journals in terms of OHR. This observation is likely explained by positive cancer RCTs were more likely to be published in journals with high impact factors.[15]

Like others, we also observed that updated results with more events often led to lower magnitudes of effects.[16] In 8/18 RCTs with more mature results, uOHR was higher than OHR, emphasizing the importance of adequate follow-up times.

This study has limitations. First, only fully published RCTs were included. Results presented at meetings frequently include nonfinal analysis and they are often discordant with subsequent publications.[17] This inclusion criterion ensured that only mature and peer-reviewed results were included but limited the number of RCTs in this analysis. Second, it has been shown that negative RCTs take longer time to be published and tended to be published in journals perceived to have less influence in clinical practices.[18,19] Therefore, it is likely that OHRs are higher than HHRs in unpublished RCTs. Including unpublished RCTs in our analysis will likely show an even larger difference between HHRs and OHRs. Third,

the current analysis included RCTs in a variety of disease settings, including both adjuvant, neoadjuvant and metastatic settings. However, 45/49 RCTs were conducted in patients with metastatic disease. Lastly, investigators may have to compromise in HHR when designing an RCT to reduce the required number of patients to fit within the resources available. It is not possible for delineate these issues with information provided in published manuscripts.

Despite the fact that the observed magnitudes of benefits were lower than the hypothesized effect sizes in > 50% published RCTs, the majority of 49 RCTs included in this analysis achieved statistical significance. We did not evaluate whether these OHRs were of meaningful clinical significance.

In conclusion, the majority of published RCTs evaluating PD-1/PD-L1 targeting mAbs did not achieve their hypothesized benefits. Investigators' optimism regarding these agents should be combined with more realistic expectations. The optimism bias requires attention from the cancer clinical research community given the number of these agents in development and the intense interest in evaluating these agents in disease settings with a lower expected benefit.

Conflict of Interest

Eitan Amir: Genentech/Roche (ET); Novartis, Sandoz, Exact Sciences (C/A). **Eric X. Chen**: Merck, Roche, BMS, Novartis, AZ, Zymeworks, Mirati (Clinical Trials), Eisai, Bayer, Ipsen (C/A). The other authors indicated no financial relationships.

(C/A) Consulting/advisory relationship; (RF) Research funding; (E) Employment; (ET) Expert testimony; (H) Honoraria received; (OI) Ownership interests; (IP) Intellectual property rights/inventor/patent holder; (SAB) Scientific advisory board.

Author Contributions

Conception/design: L.A.-S., E.C. Collection and/or assembly of data: L.A.-S., F.A.A., H.A.A. Data analysis and interpretation: All authors. Manuscript writing: All authors. Final approval: All authors.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary Material

Supplementary material is available at *The Oncologist* online.

References

1. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nat Commun*. 2020;11(1):3801.
2. Yu J X, Hodge JP, Oliva C, et al. Trends in clinical development for PD-1/PD-L1 inhibitors. *Nat Rev Drug Discov*. 2020;19(3): 163-164.
3. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet*. 2005;365(9467):1348-1353. https://doi.org/10.1016/s0140-6736(05)61034-3
4. Benjamin DM, Mandel DR, Barnes T, et al. Can oncologists predict the efficacy of treatments in randomized trials? *Oncologist*. 2021;26(1):56-62. https://doi.org/10.1634/theoncologist.2020-0054
5. Chalmers I, Matthews R. What are the implications of optimism bias in clinical research? *Lancet*. 2006;367(9509):449-450. https://doi.org/10.1016/S0140-6736(06)68153-1
6. Zakeri K, Noticewala S, Vitzthum L, et al. 'Optimism bias' in contemporary national clinical trial network phase III trials: are we improving? *Ann Oncol*. 2018;29(10):2135-2139. https://doi.org/10.1093/annonc/mdy340
7. Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *J Natl Cancer Inst*. 2012;104(8):590-598.
8. Noticewala SS, Patel RR, Ludmir EB. Absence of optimism bias in industry-sponsored cancer trials. *J Clin Oncol*. 2020;38(suppl 15):2020-2020. https://doi.org/10.1200/jco.2020.38.15_suppl.2020
9. Djulbegovic B, Kumar A, Magazin A, et al. Optimism bias leads to inconclusive results-an empirical study. *J Clin Epidemiol*. 2011;64(6):583-593. https://doi.org/10.1016/j.jclinepi.2010.09.007
10. Atanasov P, Diamantaras A, MacPherson A, et al. Wisdom of the expert crowd prediction of response for 3 neurology randomized trials. *Neurology*. 2020;95(5):e488-e498. https://doi.org/10.1212/WNL.0000000000009819
11. Lawrence NJ, Roncolato F, Martin A, Simes RJ, Stockler MR. Effect sizes hypothesized and observed in contemporary phase III trials of targeted and immunological therapies for advanced cancer. *JNCI Cancer Spectr*. 2018;2(4):pky037.
12. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. *Immunity*. 2013;39(1):1-10. https://doi.org/10.1016/j.immuni.2013.07.012
13. Ellis LM, Bernstein DS, Voest EE, et al. American Society of Clinical Oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. *J Clin Oncol*. 2014;32(12):1277-1280. https://doi.org/10.1200/JCO.2013.53.8009
14. Ancevski Hunter K, Socinski MA, Villaruz LC. PD-L1 testing in guiding patient selection for PD-1/PD-L1 inhibitor therapy in lung cancer. *Mol Diagn Ther*. 2018;22(1):1-10. https://doi.org/10.1007/s40291-017-0308-6
15. Unger JM, Barlow WE, Ramsey SD, et al. The scientific impact of positive and negative phase 3 cancer clinical trials. *JAMA Oncol*. 2016;2(7):875-881. https://doi.org/10.1001/jamaoncol.2015.6487
16. Elimova E, Moignard S, Li X, et al. Updating reports of phase 3 clinical trials for cancer. *JAMA Oncol*. 2021;7(4):593-596.
17. Booth CM, Maître AL, Ding K, et al. Presentation of nonfinal results of randomized controlled trials at major oncology meetings. *J Clin Oncol*. 2009;27(24):3938-3944.
18. Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. *JAMA* 2003;290(4):495-501. https://doi.org/10.1001/jama.290.4.495
19. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3(8):e3081. https://doi.org/10.1371/journal.pone.0003081