Journal of Clinical Epidemiology 103 (2018) 1−9

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Design analysis indicates Potential overestimation of treatment effects in randomized controlled trials supporting Food and Drug Administration cancer drug approvals

Emily M. Lord, Isabelle R. Weir, Ludovic Trinquart*

Boston University School of Public Health, Department of Biostatistics, 801 Massachusetts Avenue, Boston, MA 02118, USA

Accepted 26 June 2018; Published online 2 July 2018

Abstract

Objective: Statistical significance drives interpretation of randomized controlled trials (RCTs). We examined the type S error risk—claiming a new drug is falsely beneficial—and exaggeration ratio—how estimated effects differ from true effects—to re-emphasize direction and magnitude of treatment effects.

Study Design and Setting: We systematically reviewed RCTs supporting Food and Drug Administration (FDA) approval of cancer drugs between 2007 and 2016. We extracted data for overall survival (OS), progression-free survival (PFS), and response outcomes from FDA reviews. We estimated type S error risks and exaggeration ratios by considering replicated RCTs of equal size and a range of true effects.

Results: We analyzed 42 RCTs for 39 approved drugs. Across 38 RCTs reporting OS, the median type S error risk was 0.00% (Q1−Q3, 0.00−0.01%) and 3.56% (0.40−6.74%), for true hazard ratios of 0.7 and 0.9, respectively, indicating confidence in effect direction. The corresponding exaggeration ratios were 1.09 (1.01−1.11) and 1.30 (1.13−1.42), indicating median overestimations of 9% and 30%. Similar results held for PFS and response outcomes.

Conclusions: The type S error risk and exaggeration ratio provide additional insights into the replicability of RCTs. Our analyses also quantify the winner's curse, in which pivotal RCTs tend toward overoptimism. © 2018 Elsevier Inc. All rights reserved.

Keywords: Randomized controlled trials; Drug approval; Reproducibility of results; Statistical data interpretation; Disease-free survival; Bias

1. Introduction

Approval of new cancer drugs is typically based on the results of pivotal trials [1]. Recent literature has shown that the quality of evidence provided by pivotal trials varies substantially [2−5]. Overall, the design and interpretation of pivotal trials are subject to conventional methods that place emphasis on statistical significance. Pivotal trials are deemed positive when they show a statistically significant difference ($P < 0.05$) in the primary endpoint between treatment groups. However, a small *P*-value does not necessarily suggests strong evidence against the null, and similarly, a large *P*-value does not necessarily favors the null [6,7]. The maintained focus on statistical significance is linked with a number of problematic consequences [8−10]. In particular, concerns have been raised about the reproducibility of trials [6,11,12].

Novel metrics can provide more focus on the direction and magnitude of treatment effect estimates [13−15]. The type S (''sign'') error risk is the probability that a statistically significant treatment effect estimate is in the wrong direction as compared to the true effect. The exaggeration ratio is the factor by which the magnitude of the estimated effect differs from the true effect, given that the estimated treatment effect is statistically significant. These new design analysis concepts have been illustrated in low-powered areas of research, such as social and behavioral science, ecological studies, and developmental economics [14,16,17]. Our objective was to explore the utility of these error quantifications in pivotal randomized controlled trials (RCTs) supporting the approval of cancer drugs by the United States Food and Drug Administration (FDA).

Conflict of interest: None.

Funding/Support: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. I.W. was supported by the National Institute of General Medical (NIGMS) Interdisciplinary Training Grant for Biostatisticians (T32 GM74905). This funding source had no involvement in the conduct of the research or the preparation of the article.

* Corresponding author. Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118. Tel.: 617-638-5878; fax: 617-638-6484.

E-mail address: ludovic@bu.edu (L. Trinquart).

https://doi.org/10.1016/j.jclinepi.2018.06.012

0895-4356/© 2018 Elsevier Inc. All rights reserved.

What is new?

Key findings

- We examined the type S error risk—claiming that a new drug is falsely beneficial—and exaggeration ratio—how estimated effects differ from true effects—to re-emphasize the direction and magnitude of treatment effects.
- Among pivotal randomized controlled trials (RCTs) supporting Food and Drug Administration approval for cancer drugs, RCTs with larger standard errors on the treatment effect estimates were likely to overestimate the magnitude of the true treatment effect.
- We observed this finding consistently across a range of plausible cancer treatment effects for overall survival, progression-free survival, and response outcomes.

What this adds to what was known?

- Design analyses using the type S error risk and exaggeration ratio provide additional insights into the replicability of RCTs, beyond statistical significance.
- Our findings offer a quantification of the ''winner's curse'' in RCTs.

What is the implication and what should change now?

- Investigators should report design analyses by using the type S error risk and exaggeration ratio over a range of plausible treatment effects to better inform the interpretation of RCTs.

2. Methods

We performed a systematic review of pivotal RCTs supporting cancer drugs approved by the FDA between 2007 and 2016. Across included RCTs, we estimated the power, type S error risk, and exaggeration ratio by considering replicated RCTs of equal size across a range of underlying true treatment effects for overall survival (OS), progression-free survival (PFS), and response outcome data.

2.1. Relationship between type S error risk and exaggeration ratio with power and P-value

We described the methods to calculate the type S error risk and exaggeration ratio in Appendix A. We further illustrated the relationships between both type S error risk and the exaggeration ratio with the power and *P*-value in a modest simulation study (Fig. 1). We show that the type S error risk decreases as power and true treatment effect increase. Across all scenarios, type S error risk is small, except for low-powered trials. Similarly, the exaggeration ratio decreases when power and true treatment effect increase. But, even RCTs with higher power can still show substantial overestimation for true treatment effects. Finally, we show that the type S error risk and exaggeration ratio give additional insight as compared to the *P*-value. In particular, both RCTs with significant and nonsignificant *P*-values can yield large exaggeration ratios. In addition, we show that increasing the significance level alpha results in larger type S error risks and smaller exaggeration ratios; however, the changes are minimal. On the other hand, both type S error risk and exaggeration ratio increase substantially as the standard error (SE) on the treatment effect estimate increases. (Appendix Fig. A.1).

2.2. Trial selection

We identified all FDA-approved cancer drugs between 2007 and 2016 [18−21]. For each drug, we searched for new drug application (NDA) reviews used for the basis of drug approvals that were publicly available via Drugs@FDA (https://www.accessdata.fda.gov/scripts/cder/daf/). We then searched for the statistical review used for each NDA. If not available, we used the medical or summary reviews. We examined the executive summary from each review to determine exclusion of noncancer indications and examined the full-text to identify eligible RCTs.

We included pivotal phase II and III RCTs reporting primary and/or secondary time-to-event or response outcome measures. We excluded supportive RCTs that were not used to determine approval, and any studies that were single-arm, nonrandomized, noncomparative, or RCTs with multiple treatment arms. Noncomparative studies were defined as those in which the control arm treatment was the same drug at a different dosage, or in which the analysis was done separately for each treatment arm.

2.3. Data extraction

The first author extracted data from NDA reviews, and the last author independently compared all extracted data with the reviews. For each RCT, we collected information about the drug indication, trial sample size, randomization ratio, drug name and class, whether the RCT was stopped early, and outcome data. We classified drug indications according to National Cancer Institute classification (https://www.cancer.gov/types/by-body-location).

Time-to-event outcome measures included OS, PFS, event-free survival (EFS), and time to progression (TTP). We extracted the hazard ratio (HR) and associated 95% confidence interval (CI), then converted them into the log HR and associated SE. In two cases, the associated 95% CIs were not reported, so we used the exact *P*-values to derive the SE. Response outcomes included change in tumor burden (overall/objective response, best overall

Download English Version:

https://daneshyari.com/en/article/7518316

Download Persian Version:

https://daneshyari.com/article/7518316

Daneshyari.com