

US Food and Drug Administration Approvals of Drugs and Devices Based on Nonrandomized Clinical Trials: A Systematic Review and Meta-analysis

Razavi, Marianne; Glasziou, Paul; Klocksieben, Farina A; Ioannidis, John P A; Chalmers, Iain; Djulbegovic, Benjamin

Published in:
JAMA network open

DOI:
[10.1001/jamanetworkopen.2019.11111](https://doi.org/10.1001/jamanetworkopen.2019.11111)
[10.1001/jamanetworkopen.2019.11111](https://doi.org/10.1001/jamanetworkopen.2019.11111)

Licence:
CC BY-NC-ND

[Link to output in Bond University research repository.](#)

Recommended citation(APA):
Razavi, M., Glasziou, P., Klocksieben, F. A., Ioannidis, J. P. A., Chalmers, I., & Djulbegovic, B. (2019). US Food and Drug Administration Approvals of Drugs and Devices Based on Nonrandomized Clinical Trials: A Systematic Review and Meta-analysis. *JAMA network open*, 2(9), Article e1911111.
<https://doi.org/10.1001/jamanetworkopen.2019.11111>, <https://doi.org/10.1001/jamanetworkopen.2019.11111>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.



US Food and Drug Administration Approvals of Drugs and Devices Based on Nonrandomized Clinical Trials: A Systematic Review and Meta-analysis

Marianne Razavi, PhD; Paul Glasziou, MD, PhD; Farina A. Klocksieben, MS; John P. A. Ioannidis, MD, DSC; Iain Chalmers, MD; Benjamin Djulbegovic, MD, PhD

Abstract

IMPORTANCE The size of estimated treatment effects on the basis of which the US Food and Drug Administration (FDA) has approved drugs and devices with data from nonrandomized clinical trials (non-RCTs) remains unknown.

OBJECTIVES To determine how often the FDA has authorized novel interventions based on non-RCTs and to assess whether there is an association of the magnitude of treatment effects with FDA requirements for additional testing in randomized clinical trials (RCTs).

DATA SOURCES Overall, 606 drug applications for the Breakthrough Therapy designation from its inception in January 2012 were downloaded from the FDA website in January 2017 and August 2018, and 71 medical device applications for the Humanitarian Device Exemption from its inception in June 1996 were downloaded in August 2017.

STUDY SELECTION Approved applications based on non-RCTs were included; RCTs, studies with insufficient information, duplicates, and safety data were excluded.

DATA EXTRACTION AND SYNTHESIS Data were extracted by 2 independent investigators. A statistical association of the magnitude of estimated effect (expressed as an odds ratio) with FDA requests for RCTs was assessed. The data were also meta-analyzed to evaluate the differences in odds ratios between applications that required further testing and those that did not. The results are reported according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines.

MAIN OUTCOMES AND MEASURES Disease, laboratory, and patient-related outcomes, including disease response or patient survival, were considered.

RESULTS Among 677 drug and medical device applications, 68 (10.0%) were approved by the FDA based on non-RCTs. Estimates of effects were larger when no further RCTs were required (mean natural logarithm of the odds ratios, 2.18 vs 1.12; odds ratios, 8.85 vs 3.06; $P = .03$). The meta-analysis results confirmed these findings: estimated effects were approximately 2.5-fold higher for treatments or devices that were approved based on non-RCTs than for treatments or devices for which further testing in RCTs was required (6.30 [95% CI, 4.38-9.06] vs 2.46 [95% CI, 1.70-3.56]; $P < .001$). Overall, 9 of 677 total applications (1.3%) that were approved on the basis of non-RCTs had relative risks of 10 or greater and 12 (1.7%) had relative risks of 5 or greater. No clear threshold above which the FDA approved interventions based on the magnitude of estimated effect alone was detected.

(continued)

Key Points

Question Are US Food and Drug Administration (FDA) decisions to approve new therapies based on nonrandomized clinical trials (non-RCTs) associated with the sizes of treatment effects?

Findings In this systematic review and meta-analysis, the magnitude of treatment effects was approximately 2.5-fold higher among treatments that were approved based on non-RCTs than among treatments that required further testing in RCTs, but only approximately 1% to 2% of the drugs approved by the FDA based on non-RCTs displayed so-called dramatic effects. No clear threshold of magnitude of treatment effect above which the FDA would not request further RCTs was found.

Meaning Based on this study's findings, the size of therapeutic effects was important to FDA approval of treatments based on non-RCTs.

+ Supplemental content

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

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

Abstract (continued)

CONCLUSIONS AND RELEVANCE In this study, estimated magnitudes of effect were larger among studies for which the FDA did not require RCTs compared with studies for which it did. There was no clear threshold of treatment effect above which no RCTs were requested.

JAMA Network Open. 2019;2(9):e1911111. doi:10.1001/jamanetworkopen.2019.11111

Introduction

A well-designed randomized clinical trial (RCT) is widely considered the most reliable approach for evaluating the efficacy of clinical interventions or assessing the probable benefit of a medical device. However, RCTs often face practical limitations, such as difficulties in accruing patients, or other logistical or feasibility issues. As a result, alternative methods for the evaluation of treatments are increasingly considered, such as the assessment of a health interventions' effects in nonrandomized, observational studies.¹ Although such evaluations often provide results similar to those seen in RCTs,²⁻³ exceptions are not uncommon.⁴ The main concern about non-RCTs is that it is currently challenging to predict with confidence when the results of RCTs will be concordant with those of observational studies. The consequences of such a discordance can sometimes be disastrous.^{5,6}

It is essential to explore when observational studies are considered sufficiently credible. Regulatory agencies, such as the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), have accepted that large treatment effects can sometimes obviate the need for RCTs. They have accepted that non-RCTs can be considered reliable sources of evidence, ie, produce results with a low risk of bias, when they generate sufficiently large or dramatic effects that are thought to rule out the combined effects of bias and random errors.⁷ Therefore, an approach for evaluating when findings in observational studies can be considered credible because of sufficiently dramatic treatment effects is to analyze the approval decisions of the drug licensing authorities. Based on theoretical and simulation studies, several definitions have been proposed to categorize dramatic effects as those treatments displaying relative risks (RRs) of 5 or greater⁸ or 10 or greater⁹ or odds ratios (ORs) of 12 or greater.¹⁰

Using these definitions, we showed that the EMA approved 7% of treatments based on non-RCTs; between 2% and 4% of these approvals displayed dramatic effects.¹⁰ Like the EMA Priority Medicines and Adaptive Pathways programs, the FDA has a program called the Breakthrough Therapy designation (BTD), designed to support the approval of drugs demonstrating substantial (ie, dramatic) improvement over existing therapies and which may not require further testing in RCTs.⁷

In this study, we evaluated the association of the magnitude of treatment effects with FDA approval based on non-RCTs of drugs within the BTD program and medical devices within the Humanitarian Device Exemption (HDE) program. Because the FDA has rarely required devices be tested in RCTs, it appeared to have implicitly judged that further testing of devices in RCTs may be unnecessary. However, the FDA has also acknowledged the importance of magnitude of effect in evaluating devices in the context of the Breakthrough Devices Program.¹¹

We compared the magnitudes of effect of the interventions that were approved without a request for further randomized evidence vs interventions for which such a request was made. We hypothesized that the magnitude of treatment effect would be larger among studies for which the FDA had not requested further RCTs. We also investigated the existence of a possible threshold of treatment effect, beyond which further testing in RCTs would seem unnecessary.

Methods

Study Selection

We downloaded 606 drug applications in January 2017 (with updated downloads in August 2018) for treatments within the BTD program from its inception in January 2012 from the websites of the FDA Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. In August 2017, we downloaded similar data referring to 71 medical devices from the inception of the HDE program in June 1996. We included consecutive nonrandomized studies from FDA databases (BTD drugs and HDE devices). We excluded RCTs, studies with nonextractable data, and studies for which the applications were denied. **Figure 1** shows the flow diagrams depicting the inclusion and exclusion process. Because this analysis is based on publicly available data with no patient information disclosed, institutional review board approval was not needed. This study followed Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline.

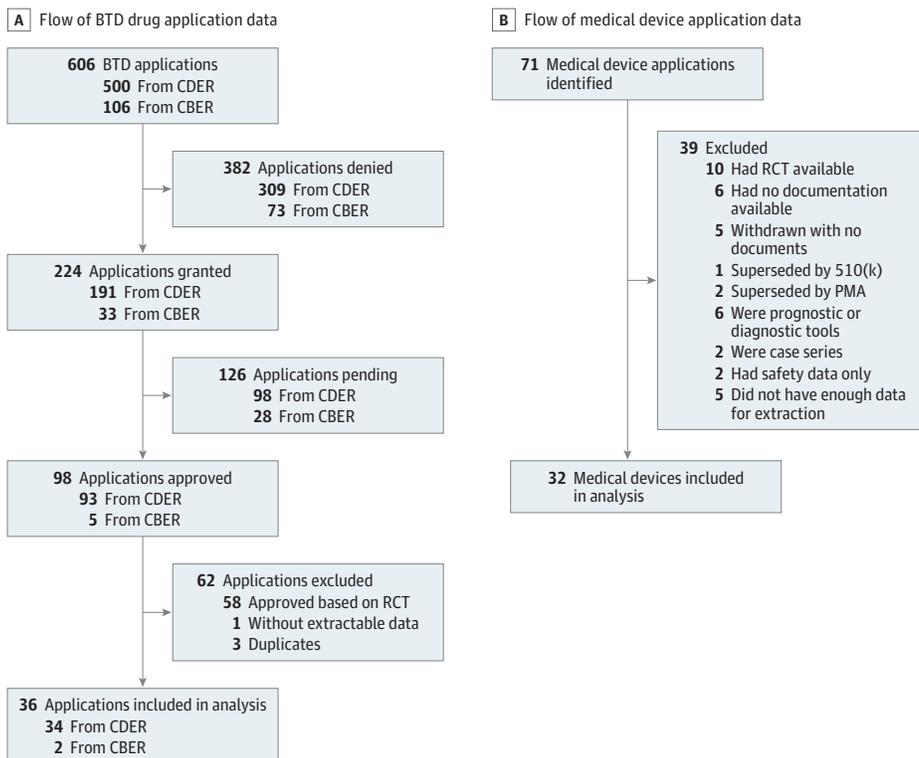
Data Extraction

Data were extracted independently by 2 of us (M.R. and F.A.K.), using a standardized data extraction form. Additionally, 25% of the data was randomly verified by 1 of us (M.R.). No major disagreement was detected. The following data were extracted: study design, primary outcome, disease characteristics, interventions, and comparators. All outcomes were dichotomous except for 4 interventions. We recorded any FDA mention of whether further RCTs should be desirable, required, or neither.

Selection of Comparators

Comparator data were extracted from FDA documents. They included control data reported in the literature associated with no active or standard treatment. When explicit details for the comparators

Figure 1. Flow Diagram of Breakthrough Therapy Designation (BTD) Drug and Humanitarian Device Exemption Medical Device Applications



A, Applications for 606 BTD drugs were downloaded from the US Food and Drug Administration websites. B, Applications for 71 devices were downloaded from the US Food and Drug Administration website. Consecutive nonrandomized studies from the US Food and Drug Administration databases were included for BTD drugs and devices. Exclusions referred to applications for which randomized clinical trials (RCTs) were used, case series, denied applications, or studies with nonextractable data. CBER indicates Center for Biologics Evaluation and Research; CDER, Center for Drug Evaluation and Research; and PMA, premarket approval.

(ie, events of interest over the number of participants in experimental and control groups) were not provided, we searched the literature and attempted to match experimental with control arms using a patients-intervention-comparator-outcome framework.¹²

Estimation of the Magnitude of Effect

The magnitude of effect was based on the following primary outcomes: overall response rate, overall survival, objective response rate, and remission rate. For devices, other primary outcomes, such as clinical improvement and reoperation rate, were also used. We used ORs and/or natural logarithm of OR (ln[ORs]) as the main metric of magnitude of effect; however, we also presented RRs because most definitions of dramatic effects were based on RRs. When a very high or very low number of successes are observed in the control group, RRs cannot have large values, while ORs can range from 0 to infinity. To enable pooling of data from the 4 studies with dichotomous outcomes with the other studies, we computed the standardized mean differences and converted them to dichotomous outcomes, as recommended in *Cochrane Handbook for Systematic Reviews of Interventions*.¹³

Statistical Analysis

We compared treatment effects of drugs and devices for which the FDA required further testing in RCTs vs those for which it did not. Our hypothesis was that the magnitude of estimated effect as expressed as an OR or ln(OR) would be statistically higher in the trials for which the FDA required no further randomized testing.

We checked the normality of the response variable and then used *t* tests or the Kruskal-Wallis test, respectively, to assess for ln(ORs) between the subgroups for which further RCTs would be required and not required. Related box plots were produced. Furthermore, to compare the average magnitude of effect between the 2 groups, we meta-analyzed the data and presented forest plots. We used random-effects modeling to summarize data. A test for subgroup differences was also used.

We applied the above analysis to the drug data, then to the drug and device data combined. Because the FDA did not express a desire for further RCTs within the device database, we could not compare the magnitude of effect between the subgroups separately for the device group. Instead, we computed ORs for the entire group as not requiring further RCTs.

To explore whether FDA decisions had reflected dramatic effects, we defined dramatic effect as a magnitude of effects with an RR of 5 or greater⁸ or of 10 or greater.⁹ We assessed the frequencies at which drugs or devices were authorized without requesting further RCTs under each definition. Insufficient data were available to test whether an OR was 12 or greater, which we had identified as an empirical value for dramatic effects in the EMA study.¹⁰

Statistical analyses were conducted in R version 3.4.1 (R Project for Statistical Computing) and SPSS version 25 (IBM). No prespecified level of statistical significance was set, and all tests were 2-tailed.

Results

Study Selection and Data Sources

We examined 606 applications for drugs with BTD as well as 71 applications for medical devices. Our final sample consisted of 68 interventions (36 BTD drugs, accounting for 6% of all drug applications, and 32 devices, accounting for 45% of all device applications) (Figure 1). Within the sample, 17 studies (25%) required further testing via RCTs and 51 (75%) did not. eTable 1 and eTable 2 in the [Supplement](#) show the characteristics of the included studies.

Breakthrough Drug Studies

Overall, 36 drugs authorized by FDA were based on non-RCTs. All studies used dichotomous data. Data from prospective studies were used in 34 cases (94%); in the remaining 2 cases, 1 analysis (3%) was based on retrospective studies and 1 (3%) combined evaluation from prospective and

retrospective data (Table 1). Most approvals (32 of 36 [89%]) were based on data from single arms. Comparators were extracted from the main FDA approval documents or were based on historical controls cited in FDA documents (Table 2). Explicit numbers for both numerator and denominator for control arms were only available for 10 studies (28%). Of the 36 comparators, 31 (86%) were active treatments; of those, 14 (45%) included reference to standard treatments. Five studies (14%) were based on no active treatment (ie, natural history of the disease) (Table 2). Overall response rate was

Table 1. Design and Outcomes of 68 Studies

Design and Outcome	No. (%)	
	Drugs (n = 36)	Devices (n = 32)
Study designs		
Prospective	34 (94)	25 (78)
Retrospective	1 (3)	6 (19)
Prospective and retrospective	1 (3)	1 (3)
Single-arm study	32 (89)	30 (94)
Comparator treatments		
Active	31 (86)	18 (56)
No active	5 (14)	14 (44)
Outcomes		
Overall response rate	24 (67)	...
Survival	2 (6)	12 (38)
Objective response rate	5 (14)	1 (3)
Remission	1 (3)	1 (3)
Reoperation	NA	1 (3)
Perforation closed	NA	1 (3)
Aneurysm	NA	3 (9)
Systolic blood pressure	NA	1 (3)
Fecal incontinence episodes per week	NA	1 (3)
Mean prosthetic use score	NA	1 (3)
Yale-Brown Obsessive Compulsive Scale score	NA	1 (3)
Other outcomes	4 (11)	9 (28)

Abbreviation: ellipses, not available; NA, not applicable.

Table 2. Control Arms in 68 Drug and Device Studies

Control Arm	No. (%)
Breakthrough Drug Studies (n = 36)	
Explicitly mentioned in FDA document	10 (28)
Data found in the published manuscript or current guidelines	16 (44)
Mentioned in FDA document but no explicit denominator, ie, referred to historical control or efficacy threshold	5 (14)
Found using literature search	5 (14)
Medical Device Studies (n = 32)	
Data explicitly mentioned in FDA document	14 (44)
Natural history data	2 (6)
Mentioned in FDA document but denominator was not explicitly given	4 (13)
Found in references cited by the FDA	3 (9)
Found using literature search	2 (6)
Found in the published manuscript for the device	7 (22)

Abbreviation: FDA, US Food and Drug Administration.

the most commonly used outcome (Table 1). The median number of patients in the experimental arms was 106 (range, 8-411).

Of 36 FDA drug approvals, the following 5 drugs were also in the EMA database: ofatumumab, blinatumomab, idarucizumab, asfotase alfa, and ceritinib. However, asfotase alfa is counted once in the EMA database and split into 2 indications in the FDA database. Therefore, 30 drugs were unique to the FDA data, and 46 were unique to EMA data.

Studies on Devices

A dichotomous outcome was used in 28 of 32 device studies (88%), and a continuous outcome in 4 (13%). Data from prospective studies were used in 25 approvals (78%), from retrospective studies in 6 (19%), and from retrospective and prospective data in 1 (3%) (Table 1). Most studies (30 of 32 [94%]) were single-arm trials. Comparators were often based on baseline values, historical data, or cited in approval documents. Explicit numbers for control arms were available for 14 devices (44%) (Table 2). Overall, 18 comparators (56%) were active treatments, and 14 (44%) were no active treatments (ie, natural history or comparison with baseline values). Twelve outcomes (38%) consisted of survival data at different periods (Table 1). The 4 continuous outcomes included systolic blood pressure, weekly fecal incontinence episodes, mean prosthetic use score, and Yale-Brown Obsessive Compulsive Scale score at 12 months. The median number of patients in experimental arm was 31 (range, 12-597), excluding the 4 studies using continuous variables.

Magnitude of Effects According to Requirement for Further RCTs

The eFigure in the Supplement displays box plots of the estimated magnitude of effects based on raw data. The distribution of magnitude of the estimated effects ranged from an OR of 0.66 to 2566.00 in the drug group and from an OR of 0.32 to 2807.00 for the combined group. Based on the raw data, the magnitude of estimated effects was higher in the group that did not require RCTs compared with the group that did require RCTs (drug-only group: mean $\ln[\text{OR}]$, 2.23 [OR, 9.34; SD, 1.83] vs mean $\ln[\text{OR}]$, 1.12 [OR, 3.06; SD, 1.32]; $P = .02$; combined group: mean $\ln[\text{OR}]$, 2.18 [OR, 8.82; SD, 1.98] vs mean $\ln[\text{OR}]$, 1.12 [OR, 3.06; SD, 1.32]; $P = .03$).

Because the FDA did not express a desire to confirm the observed effects in additional RCTs for the devices, a separate analysis according to the request for future RCTs of devices was not possible. However, the magnitude of effect observed for all devices had a mean (SD) $\ln(\text{OR})$ of 2.14 (2.09) with a mean (SD) OR of 8.49 (8.10), which was similar to the treatment effects noted for the drugs for which no further testing in RCTs was required.

The meta-analysis also yielded a statistically larger magnitude of estimated effects associated with the group that did not require RCTs than for the group that did require further RCTs (BTD drugs: OR, 6.79 [95% CI, 3.90-11.80] vs OR, 2.46 [95% CI, 1.70-3.56]; $P = .003$; combined group: OR, 6.30 [95% CI, 4.38-9.06] vs OR, 2.46 [95% CI, 1.70-3.56]; $P < .001$) (Figure 2). The meta-analytic estimate of ORs for the devices based on all data was an OR of 6.09 (95% CI, 3.70-10.03).

Assessment of Dramatic Effect Threshold

Regardless of which criterion for dramatic effect was used (Figure 3), we were unable to identify any clear threshold beyond which the FDA would not require additional randomized evidence. Overall, 12 studies (26%) for which the FDA did not request a subsequent RCT noted RRs of 5 or greater vs 2 studies (12%) for which it did ($P = .27$). We observed RRs of 10 or greater in 9 studies (19%) for which the FDA did not request a subsequent RCT vs 1 (6%) for which it did ($P = .38$). Overall, 9 of 677 total applications (1.3%) that were approved on the basis of non-RCTs had RRs of 10 or greater and 12 (1.7%) had RRs of 5 or greater.

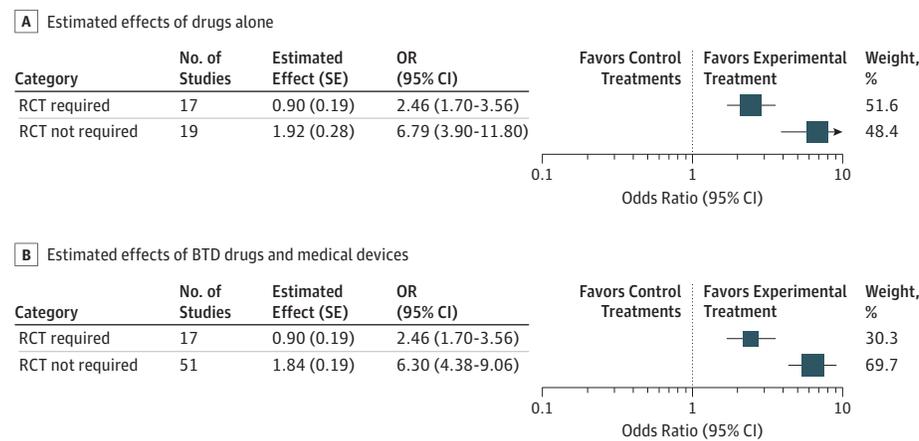
We did not perform detailed qualitative analyses of FDA reports. However, our reading of FDA documents indicated that 2 other factors may be associated with the decision to approve drugs or devices based on non-RCTs: first, rare diseases, for which there are logistical and feasibility issues

related to conducting RCTs, and, second, the belief that outcomes in control arms are so well known that RCTs would not be necessary.

Discussion

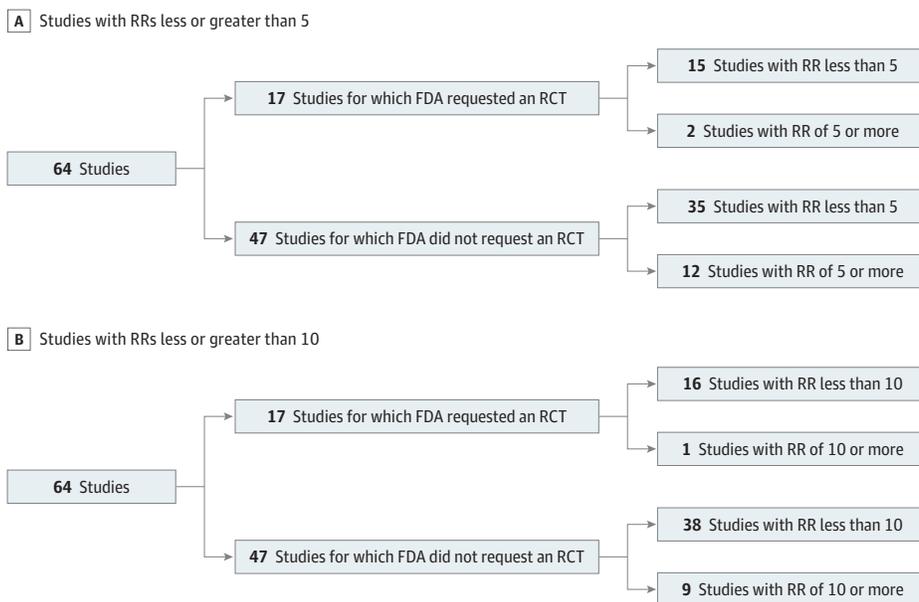
A system of RCTs is widely considered the most reliable vehicle for advances in therapeutics that result in the development of slightly more than 50% of new treatments that are superior to standard treatments.¹⁴ Paradoxically, it is this very randomization that is increasingly seen as a barrier to more efficient drug development.¹⁵ Therefore, increasing calls for using observational, real-world data in drug development and the practice of medicine are being made.¹⁶ However, even though previous studies comparing the reliability of the use of observational studies with RCTs for drug development and use found that the results of observational research were often similar to those of experimental studies, they also concluded that the results can frequently differ to the extent that it is impossible to predict with confidence the amount and direction of bias in observational research.² The only other

Figure 2. Forest Plots Summarizing Estimated Effects



Estimated effect and SE refer to the natural logarithm of the odds ratio (OR) and the SE of the natural logarithm of the OR, respectively. Box sizes indicate effect size, number of participants, and size of SE; RCT indicates randomized clinical trial.

Figure 3. Request for Randomized Clinical Trials (RCTs) According to 2 Criteria of Dramatic Effects



A, US Food and Drug Administration (FDA) requests for RCTs by studies, stratified by definition of dramatic effect as a risk ratio (RR) of 5 or greater.⁸ B, US Food and Drug Administration requests for RCTs, stratified by definition of dramatic effects as RR of 10 or greater.⁹ The denominator is 64 because 4 studies had continuous data as their primary outcome, thereby preventing the computation of RRs.

way to replace randomization and reliably assess if new treatments are better than old ones is to develop drugs displaying very large (ie, dramatic) effects.⁹ Such dramatic effects are believed to be reliable because the sheer magnitude of the treatment differences is deemed sufficient to override the combined effects of allocation bias and random error (for which randomization so effectively controls).⁹ As a result, drug development has been increasingly characterized by searching for treatments with dramatic effects. It is thought that the application of rational drug design and the exploitation of omics-based technologies to develop precision medicine and targeted therapies could replace the RCT system; hence, testing in observational studies (ie, non-RCTs) could suffice.

But how large is large, and when should effects be considered sufficiently dramatic that further testing in RCTs is not needed? What is the probability of developing treatments that will have dramatic effects when tested in patients? Having such empirical benchmarks may arguably speed up the drug discovery process.

One way to answer these questions is to evaluate how often and with what size treatment effects regulatory agencies such as the FDA approve drugs for clinical use without requiring testing in RCTs. When enrollment in RCTs is difficult owing to practical limitations, a strong estimated effect may be sufficient to convince regulators that such testing is unnecessary.¹⁷ In fact, the FDA and EMA have introduced regulatory mechanisms to support the approval of drugs demonstrating substantial improvement over existing therapies based on testing in non-RCTs.⁷ However, empirical evidence concerning the magnitude of estimated treatment effects sufficient to approve novel medical treatments, especially if the proposed intervention is based on observational evidence, is limited.

In our previous work, we found that the EMA approved 7% of drugs based on data from nonrandomized studies alone. Of those, between 2% and 4% of EMA approvals based on nonrandomized data exhibited so-called dramatic effects; we also found that larger effect sizes in nonrandomized studies were associated with higher rates of EMA licensing approval.¹⁰

In the EMA study, estimates of dramatic effects based on observational data were seen in 32% of interventions using the criteria of an RR of 10 or greater and 40% of interventions using the criteria of an RR of 5 or greater.^{8,9} In the current study, the FDA approved 10.0% (68 of 677) of drugs or devices based on observational studies. The rate was very similar to the EMA rate when drugs were considered, while it was much higher for devices in FDA studies. Depending on the definition, dramatic effects were observed in between 1.3% (9 of 677) and 1.7% (12 of 677) of FDA approvals. This should be contrasted with the 50% success rate if the drug is brought forward to evaluation in RCTs.¹⁴

In our EMA study,¹⁰ the average OR beyond which the agency would not require further testing in RCTs was 12; however, in the FDA data, the respective OR was substantially smaller (6.3). The current study confirms the observations we made evaluating EMA approvals,¹⁰ indicating that when larger magnitudes of effects are seen in non-RCTs, there is less desire from the regulatory agency to require subsequent confirmatory RCTs. However, as with our earlier study,¹⁰ we found no clear evidence of a threshold above which FDA would approve drugs or devices based solely on the magnitude of estimated effects. This implies that other factors may play significant roles in FDA decisions regarding approvals based on non-RCT studies.

Limitations

Our study has several limitations. Although it was based on all studies of drugs approved using non-RCTs within the FDA BTB and device programs, our sample size remained small.

Second, the availability of data on comparators was inadequate. As the evaluation of the effects of treatments always entails a comparison, this is very concerning; as in the case of our EMA analysis,¹⁰ the available data did not frequently have clear descriptions of numerators and denominators in experimental and control arms. As a result, some of our treatment effect estimates were necessarily approximations. In fact, explicit numbers for control arms were available in only about one-third of instances. A number of comparisons of primary outcomes relied on historical data, with few explanations of the rationale of the chosen comparators given. If reliable data on controls

are not available, how did the FDA make a judgement that is transparent to the public? In our attempt to associate magnitudes of effects with FDA decisions, we also noticed that the FDA assumed very low event rates for favorable outcomes in the control arm based on few observations, resulting in some estimates of improbably large magnitudes of effects. Thus, as with the EMA cohort, we judged that the quality of the underlying evidence used to make approval decisions was very low and unaccompanied by transparent justifications. Because of this, we deemed it unnecessary to assess the quality of studies on other dimensions recommended by Downs and Black,¹⁸ which we used in our EMA study. However, even under these best case scenarios, the average magnitude of estimated effects observed for most of the breakthrough interventions was relatively modest. This should give us a pause regarding the current excitement surrounding the development of drugs using a precision medicine approach¹⁹ apparently aimed to detect treatments with very large effects. Similar results observed in the analysis of the EMA and FDA databases appear to reinforce a more cautious and realistic approach to drug development and drug licensing.

Although larger estimates of effects were noticed in the group that required RCTs than in the group that did not, we did not detect any threshold dramatic effect. In the future, as more studies become available, identifying a cutoff point may be feasible and useful. Indeed, discovering such a threshold may serve as an important benchmark²⁰ for the development of drugs and devices based on observational studies. We believe that discovering threshold effect size and defining how large a treatment effect should be to obviate RCTs are currently among the most pressing clinical research questions. The answers could be informative for clinical trialists (to decide if they should pursue RCTs vs non-RCTs), practitioners (to decide if the treatment effect of a particular drug is unaffected by bias), and drug developers (to provide them with benchmarks in terms of likelihood of developing a treatment with dramatic effects).

Our findings of no clear evidence of a dramatic effect threshold raised the issue of the importance of other factors that may have influenced the FDA approval decisions, such as treatments for rare diseases or current treatments that are thought to be so well known that comparison with novel treatments may appear unnecessary. Better understanding of FDA decisions might be gleaned by further in-depth, formal qualitative analyses.

Conclusions

In this study, we explored the association of magnitude of estimated effects with FDA requests for further testing in RCTs. We concluded that the FDA has not requested subsequent RCTs when a treatment has displayed a larger magnitude of estimated treatment effects. However, the quality of data on which the FDA and EMA⁵ presumptively make these estimates and eventually decisions for approval is very low. As inferences about treatment effects always rely on comparisons (direct or counterfactual), greater transparency about the basis for choosing (historical) comparators is needed to instill confidence that decisions about drug and device approval based on observational data are appropriate.

ARTICLE INFORMATION

Accepted for Publication: July 24, 2019.

Published: September 11, 2019. doi:10.1001/jamanetworkopen.2019.11111

Open Access: This is an open access article distributed under the terms of the [CC-BY-NC-ND License](#). © 2019 Razavi M et al. *JAMA Network Open*.

Corresponding Author: Benjamin Djulbegovic, MD, PhD, Department of Supportive Care Medicine and Hematology, City of Hope, 1500 E Duarte Rd, Duarte, CA 91010 (bdjulbegovic@coh.org).

Author Affiliations: Department of Supportive Care Medicine, City of Hope, Duarte, California (Razavi, Djulbegovic); Bond University, Gold Coast, Queensland, Australia (Glasziou); Program for Comparative

Effectiveness Research, University of South Florida, Tampa (Klocksieben); Stanford Prevention Research Center, Departments of Medicine, Health Research and Policy, and Biomedical Data Science, Stanford University School of Medicine, Stanford, California (Ioannidis); Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California (Ioannidis); Meta-Research Innovation Center at Stanford, Stanford University, Stanford, California (Ioannidis); James Lind Library, Oxford, United Kingdom (Chalmers); Department of Hematology, City of Hope, Duarte, California (Djulbegovic).

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

Concept and design: Glasziou, Ioannidis, Chalmers, Djulbegovic.

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

Drafting of the manuscript: All authors.

Critical revision of the manuscript for important intellectual content: Glasziou, Ioannidis, Chalmers, Djulbegovic.

Statistical analysis: Razavi, Klocksieben, Djulbegovic.

Obtained funding: Chalmers, Djulbegovic.

Administrative, technical, or material support: Razavi, Djulbegovic.

Supervision: Djulbegovic.

Conflict of Interest Disclosures: Drs Razavi and Djulbegovic reported grants from GlaxoSmithKline during the conduct of the study. No other disclosures were reported.

Funding/Support: This study was supported by grant 616210515 from GlaxoSmithKline.

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

Additional Contributions: Patrick Vallance, MD (GlaxoSmithKline), provided valuable input regarding our findings and promoted funding for this project. He was not compensated for his time.

REFERENCES

1. Frieden TR. Evidence for health decision making: beyond randomized, controlled trials. *N Engl J Med*. 2017;377(5):465-475. doi:10.1056/NEJMra1614394
2. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998;317(7167):1185-1190. doi:10.1136/bmj.317.7167.1185
3. Anglemeyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014;4(4):MR000034. doi:10.1002/14651858.MR000034.pub2
4. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ*. 2016;352:i493. doi:10.1136/bmj.i493
5. Rossouw JE, Anderson GL, Prentice RL, et al; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321-333. doi:10.1001/jama.288.3.321
6. Moore TJ. *Deadly Medicine*. New York, NY: Simon & Schuster; 1995.
7. Sherman RELJ, Li J, Shapley S, Robb M, Woodcock J. Expediting drug development—the FDA's new "breakthrough therapy" designation. *N Engl J Med*. 2013;369(20):1877-1880.
8. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol*. 2011;64(12):1277-1282.
9. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? picking signal from noise. *BMJ*. 2007;334(7589):349-351. doi:10.1136/bmj.39070.527986.68
10. Djulbegovic B, Glasziou P, Klocksieben FA, et al. Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. *J Clin Epidemiol*. 2018;98:24-32. doi:10.1016/j.jclinepi.2018.01.011
11. US Food and Drug Administration. Breakthrough Devices Program. <https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/ucm441467.htm#s1>. Accessed August 2, 2019.
12. Straus S, Richardson W, Glasziou P, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. London, United Kingdom: Churchill Livingstone; 2005.

13. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0*. Oxford, United Kingdom: Wiley-Blackwell; 2011.
14. Djulbegovic B, Kumar A, Glasziou P, Miladinovic B, Chalmers I. Medical research: trial unpredictability yields predictable therapy gains. *Nature*. 2013;500(7463):395-396. doi:10.1038/500395a
15. Freedman DH. Clinical trials have the best medicine but do not enroll the patients who need it. *Scientific American*. January 2019. <https://www.scientificamerican.com/article/clinical-trials-have-the-best-medicine-but-do-not-enroll-the-patients-who-need-it/>. Accessed August 6, 2019.
16. National Academies of Sciences, Engineering, and Medicine. *Examining the Impact of Real-World Evidence on Medical Product Development: Proceedings of a Workshop Series*. Washington, DC: National Academies Press; 2019.
17. Behera M, Kumar A, Soares HP, Sokol L, Djulbegovic B. Evidence-based medicine for rare diseases: implications for data interpretation and clinical trial design. *Cancer Control*. 2007;14(2):160-166. doi:10.1177/107327480701400209
18. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377-384. doi:10.1136/jech.52.6.377
19. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med*. 2011;364(12):1144-1153. doi:10.1056/NEJMra1010600
20. Miladinovic B, Kumar A, Mhaskar R, Djulbegovic B. Benchmarks for detecting 'breakthroughs' in clinical trials: empirical assessment of the probability of large treatment effects using kernel density estimation. *BMJ Open*. 2014;4(10):e005249. doi:10.1136/bmjopen-2014-005249

SUPPLEMENT.

eFigure. Distribution of Magnitude of Treatment Effects as a Function of FDA Preference for Testing of Approved Treatments in Further Randomized Controlled Trials (RCTs)

eTable 1. Characteristics of Included Drugs

eTable 2. Characteristics of Included Devices