Journal of Clinical Epidemiology 114 (2019) 49–59

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Single pivotal trials with few corroborating characteristics were used for FDA approval of cancer therapies

Aviv Ladanie[a,b], Benjamin Speich[a,c], Matthias Briel[a,d], Francesco Sclafani[e], Heiner C. Bucher[a], Arnav Agarwal[d,f], John P.A. Ioannidis[g,h,i,j,k], Tiago V. Pereira[l], Benjamin Kasenda[a,m,n], Lars G. Hemkens[a,*]

[a]*Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital and University of Basel, Basel, Switzerland*

[b]*Swiss Tropical and Public Health Institute (Swiss TPH), Basel, Switzerland*

[c]*Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK*

[d]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada*

[e]*Gastrointestinal Unit, Department of Medical Oncology, Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium*

[f]*Department of Medicine, University of Toronto, Toronto, Ontario, Canada*

[g]*Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA*

[h]*Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA*

[i]*Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA*

[j]*Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA USA*

[k]*Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA*

[l]*Health Technology Assessment Unit, Institute of Education and Health Sciences, Oswaldo Cruz German Hospital, São Paulo, Brazil*

[m]*Department of Medical Oncology, University Hospital and University of Basel, Basel, Switzerland*

[n]*Department of Haematology/Oncology and Palliative Care, Klinikum Stuttgart, Stuttgart, Germany*

Accepted 28 May 2019; Published online 31 May 2019

Abstract

Background and objective: Novel cancer therapies are often approved with evidence from a single pivotal trial alone. There are concerns about the credibility of this evidence. Higher validity may be indicated by five methodological and statistical characteristics of pivotal trial evidence that were described by the U.S. Food and Drug Administration (FDA), which may corroborate the reliance on a single trial alone for approval decisions.

Study design: We did a metaepidemiologic evaluation of all single pivotal trials supporting FDA approval of novel drugs and therapeutic biologicals for cancers between 2000 and 2016. For each trial, we determined the presence of these five characteristics, which we operationalized as (1) large and multicenter trial (≥200 patients; more than one center); consistent treatment benefits across (2) multiple patient subgroups (in view of FDA reviewers), (3) multiple endpoints (including overall survival, progression-free survival, response rate, health related quality of life), and (4) multiple treatment comparisons (e.g., multi-arm studies); and (5) ''statistically very persuasive'' results (P-values <0.00125).

Results: Thirty-five of 100 approvals were based on evidence from a single pivotal trial without any further supporting evidence on beneficial effects (20 randomized controlled trials and 15 single-arm trials). The number increased substantially from one approval before 2006 to 23 after 2011. Sixty-six percent (23/35) of the trials were large multicenter trials (median 301 patients and 63 centers). Consistent effects were demonstrated across subgroups in 66% (23/35), across endpoints in 43% (15/35), and across multiple comparisons in 3% (1/35). Very low P-values for the primary endpoint were seen in 34% (12/35). At least one of the corroborating characteristics was present in

Funding: This project was supported by the Swiss Cancer League (Grant No KLS-3587-02-2015). The Basel Institute for Clinical Epidemiology and Biostatistics is supported by Stiftung Institut für klinische Epidemiologie.

Conflict of interest: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf. Heiner C. Bucher has received grants, support for traveling, consultancy fees, and honorarium from Gilead, BMS, Viiv Healthcare, and Roche that were not related to this project in the 36 months before the submission of this manuscript. He serves as the president of the association contre le HIV et autres infections transmissibles. In this function, he has received support for the Swiss HIV Cohort Study from ViiV Healthcare, Gilead, BMS, MSD, and AbbVie. Benjamin Kasenda has received support for traveling and consultancy fees from Roche that were not related to this project in the 36 months before the resubmission of this manuscript. Aviv Ladanie is employed by Novartis Pharma AG, Basel, Switzerland. All other authors declare no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

* Corresponding author: Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital and University of Basel, Basel, Switzerland. Tel.: +41-(0)61-55-65100; fax: +41-(0)61-26-53109.

E-mail address: lars.hemkens@usb.ch (L.G. Hemkens).

https://doi.org/10.1016/j.jclinepi.2019.05.033

94% (33/35) of all approvals, two or more were present in 54% (19/35), and none had all characteristics.

Conclusions: Single pivotal trials typically have some of the corroborating characteristics, but often only one or two. These characteristics need to be better operationalized, defined, and reported and whether single trials with such characteristics provide similar evidence about benefits and harms of novel treatments as multiple trials would do needs to be shown.

Keywords: U.S. food and drug administration; Drug approval; Cancer; Confirmatory evidence; Single pivotal trial; FDAMA section 115a; Metaresearch

1. Introduction

Benefits and harms of novel treatments are commonly associated with uncertainty, and early evidence often provides only limited guidance to assess the true merits of such therapies [1–3]. The Federal Food, Drug, and Cosmetic Act (FFDCA) in 1962 established the ''effectiveness requirement'' for drug approvals by the U.S. Food and Drug Administration (FDA) and stipulates ''substantial evidence'' from ''adequate and well-controlled investigations'' [4]. Although this has traditionally been interpreted as general rule that at least two pivotal studies are needed for drug approval, under particular circumstances, the FFDCA allows that the FDA grants approval based on early evidence from only a single pivotal trial [5]. These circumstances and their legal and scientific basis are outlined in an FDA ''Guidance for Industry'' document [6]. It summarizes several general prerequisites for such situations, for example, assuming that ''the single study has been appropriately designed, that the possibility of bias due to baseline imbalance, unblinding, post hoc changes in analysis, or other factors is judged to be minimal, and that the results reflect a clear prior hypothesis documented in the protocol.'' It also differentiates approval situations with a single pivotal study that is accompanied by additional ''supporting evidence'' (from ''related adequate and well-controlled studies'', for example, in other populations, disease stages, or closely related diseases) and situations where there is only a single pivotal study alone. For the latter situation, it is further required that the pivotal trial ''has demonstrated a clinically meaningful effect on mortality, irreversible morbidity, or prevention of a disease with potentially serious outcome and confirmation of the result in a second trial would be practically or ethically impossible.''

Beyond supporting data from other studies, clinical relevance, ethical implications of the results, and possible legal circumstances, there are also some specific methodological factors of the study design and statistical results listed in the guidance document as examples of trial characteristics that may corroborate the reliance on a single pivotal trial. The corroborating characteristics specifically mentioned are (1) a large trial size and multicenter design (with no center having disproportionately large influence on the number of enrolled patients or the overall treatment effect); treatment effects that are consistent across (2) patient subgroups, (3) multiple endpoints (''involving different events'' or representing different beneficial effects), and (4) multiple comparisons (in factorial designs or multiarm studies); and (5) ''statistically very persuasive'' results.

In oncology, approval based on evidence from a single pivotal study alone has become standard. Metaepidemiological surveys show that 80 to 85% of oncology drug approvals by the FDA and the European Medicines Agency are based on a single pivotal study alone but often with accompanying ''supportive efficacy evidence'' (Table 1). This draws attention to the abovementioned circumstances. Although supporting data from other studies, clinical relevance, and ethical implications are complex and subject to numerous debates often on a case by case basis, we expected that such methodological characteristics can be objectively assessed and compared across approvals.

We aimed to focus on these situations at approval of novel cancer therapies where there is only a single pivotal study alone, without accompanying ''supportive efficacy evidence.'' We took for granted that the general prerequisites are fulfilled, that is, all these single studies provide clinically meaningful effects and are adequate and well-controlled investigations. In such situations, the characteristics would be most important and a better understanding of their prevalence would provide valuable insights in the assessment of early treatment effects and their surrounding uncertainty.

2. Methods

2.1. Identification of approval evidence

We identified all 92 cancer drugs and therapeutic biologics that received a marketing authorization by the FDA for the first time (so-called new molecular entities and therapeutic biologic products, herein referred to as ''drugs'') between 2000 and 2016 in a related project [7]. The methods and search strategies have been described previously [7,8]. Briefly, for each eligible drug, we obtained FDA approval packages containing reviews of the data submitted by the drug manufacturer to support their claims. These provide information about the pivotal studies and are publicly available (for approved drug products) from the drugs@FDA database [9]. We reused the information about brand and generic name of the cancer drugs, FDA-approved indication, approval date, orphan status, and whether a drug was granted accelerated approval,

What is new?

Key findings

- Regulatory approval of modern cancer treatments is increasingly based on only a single pivotal trial, which is often not randomized and often without any further supporting evidence on beneficial effects.

What this adds to what was known?

- Over almost two decades, single pivotal trials providing the only evidence for the establishment of drug efficacy always involved many centers with a median of 301 patients, but rarely more than 500 patients.
- Such trials typically have some of the corroborating characteristics that were defined by the FDA, but many of them have only one of these characteristics.
- The reporting and description of these corroborating characteristics and their impact on regulatory decision-making is often not transparent and may not be consistent.

What is the implication and what should change now?

- Corroborating characteristics need to be better operationalized, defined, and reported.
- Whether single pivotal trials with such characteristics provide similar evidence about benefits and harms of novel treatments as multiple trials would do remains to be shown.

breakthrough therapy status, or priority review from the related project [7]. We obtained information on fast track designations from the FDA website [10,11]. We obtained from clinicaltrials.gov additional information on the included randomized trials for cross-checks of our extractions (last search: 25 July 2018).

Approval packages for each of the 92 drugs were carefully examined by one reviewer (A.L.). We included all approvals with a clear statement in the summary review that only one study provided the efficacy information used for the approval decision. When there was no summary review available, no statement or an unclear statement such as that the efficacy evidence used for the approval came ''mainly'' or ''primarily'' from only one study, we searched in the medical and statistical reviews for studies described as supporting the efficacy claims. We excluded all approvals that included two or more studies described as ''pivotal'' or where we found statements that more than one study provided efficacy evidence that supported the approval decision. A second reviewer (either L.G.H. or B.K.) verified the eligibility. There was almost perfect agreement (99%).

2.2. Characteristics of single pivotal studies

For approvals based on a single pivotal study alone, one reviewer (A.L.) identified the pivotal study, extracted the study name as well as the number of study arms, and categorized the study as either ''randomized controlled trial (RCT)'' or ''single-arm trial (SAT)''; we use the term ''trial'' because all pivotal studies were experimental clinical trials. The SAT category encompasses both trials without concurrent controls as well as dose-comparison trials in which patients in all groups are receiving the experimental drug. The description of the five characteristics and the way we operationalized them is described in Table 2.

All extractions and categorizations were conducted by one reviewer (A.L.). The categorizations of the first reviewer were confirmed by another reviewer (B.S.), and any disagreements were resolved by a third reviewer (L.G.H.). Information on fast track designations was collected by one reviewer (L.G.H.) and confirmed by a second one (A.H.).

2.3. Cross-check with clinicaltrials.gov

One reviewer (M.B.) perused the information in clinicaltrials.gov (as of 25 July 2018) and determined whether overall survival (OS), progression-free survival (PFS), response rate (RR), or health-related quality of life were listed as primary or secondary endpoints of the trials (which was with only very few exceptions always the case). The endpoint that we determined as primary endpoint for regulatory decision-making was in all cases the primary endpoint of the trial (with one possible exception for ALIMTA®/pemetrexed disodium where the entry in clinicaltrials.gov was unclear).

2.4. Data analysis

We report the results descriptively using summary statistics (minimum, median, interquartile range [IQR], and maximum). We separately assessed the subset of approvals with orphan status. We conducted no statistical hypothesis tests.

3. Results

3.1. Characteristics of approvals based on a single pivotal trial

A total of 92 drugs received FDA approval between 2000 and 2016 for 100 cancer indications. Thirty-three novel drugs were approved based on evidence from 34 single pivotal trials alone for 35 cancer indications (two drugs

Table 1. Studies quantifying the prevalence of single pivotal studies supporting drug approvals

Study reference	Study objective(s)	Health authority analyzed	Time period analyzed[a]	Disease and/or drug characteristics	Prevalence of single pivotal trials supporting FDA approval, % (*n*/*N*) [analysis level]	Distinction of single trial approvals with and without supporting evidence[b]
Downing et al. (2014) [12]	To characterize pivotal efficacy trials for newly approved novel therapeutic agents.	FDA	2005 to 2012	Cancer (not otherwise specified)	80% (33/41) (indication)	No
Sridhara et al. (2010) [13]	To conduct an overview of products that were reviewed by the FDA's Office of Hematology and Oncology Products for marketing approval and the regulatory actions taken during July 2005 to December 2007.	FDA	2005 to 2007	Solid tumors and hematologic malignancies	83% (44/53[c]) (indication)	No
Martell et al. (2013) [14]	To describe approval trends and characteristics.	FDA	2006 to 2011	Solid tumors and hematologic malignancies	83% (NR/NR[c]) (indication)	No
Morant et al. (2017) [15]	To analyze the clinical efficacy evidence submitted in support of the initial marketing authorizations of new active substances approved between 2012 and 2016, with focus on approvals based on a single pivotal clinical trial.	EMA	2012 to 2016	Oncology products (according to the Anatomical Therapeutic Chemical Classification System)	84% (43/51) (drug)	No
Gentry (2015) [16]	To generate insight into the design of single pivotal studies and how the trial features were applied to provide adequate data for approval.	FDA	2005 to June 2015	Cancer (not otherwise specified, but probably both solid tumors and hematologic malignancies)	85% (55/65) (indication)	No
Tibau et al. (2017) [17]	To derive the clinically meaningful benefit for FDA-approved drugs using the "European Society for Medical Oncology—Magnitude of Clinical Benefit Scale".	FDA	2006 to 2016	Solid tumors	82% (97/118[c]) (indication)	No

Abbreviations: EMA, European Medicines Agency; FDA, Food and Drug Administration; NR, not reported.

[a] Covering the time period from 1 January to 31 December if not otherwise specified.

[b] Distinction between approvals that are genuinely based on a single trial and approvals that are based on evidence from a single pivotal trial plus supporting evidence from other studies that are unrelated to the efficacy assessment.

[c] Including both original (i.e., the first-ever FDA-approved drug use) and supplemental approvals (i.e., changes to the original drug use approved by the FDA, such as adding a new indication).

were approved for 2 or 3 indications, and two indications were approved based on evidence from the same trial). The number of approvals based on a single pivotal trial alone increased since 2000, with approximately two-thirds of indications being approved since 2012 (Table 3). Of the 35 approvals, 23 (66%) had an orphan status and 14 (40%) underwent accelerated approval. Twenty (56%) of single pivotal trials were RCTs and 15 (43%) were SATs (Table 3).

3.2. Proportion of single pivotal trials considered large and multicenter

The median number of patients enrolled across the 35 single pivotal trials was 301 (IQR 125, 571). All trials had multicenter designs with 16 to 184 centers (Table 3). No trial was dominated by a single center, and 22 (63%) of 35 trials enrolled more than 200 patients (9 trials [26%] had more than 500 and 4 [11%] more than

Table 2. Corroborating characteristics of single pivotal trials

Large multicenter study	
''In a large multicenter study in which (1) no single study site provided an unusually large fraction of the patients and (2) no single investigator or site was disproportionately responsible for the favorable effect seen, the study's internal consistency lessens concerns about lack of generalizability of the finding or an inexplicable result attributable only to the practice of a single investigator.'' [6]	We extracted for all single pivotal trials the total number of patients enrolled, the number of participating centers, and any statement about whether a single study site provided a large fraction of the patients. The latter aspect was usually reported in the chapter ''Ethics and good clinical practices'' of the medical review and is mainly derived from on-site inspection reports by the FDA's Office of Scientific Investigations and financial disclosure information from participating investigators. Because there is no established definition of what constitutes a large trial, we defined a trial with at least 200 enrolled patients (across all trial arms) to be large for our main analysis. We acknowledge that this threshold may be arbitrary, but this cutoff is commensurate with recent analyses distinguishing between small and large cancer trials [19]. Trials with more than one enrolling center were considered to have multicenter designs.
Consistency of treatment effects across patient subgroups	
''Frequently, large trials have relatively broad entry criteria and the study populations may be diverse with regard to important covariates such as concomitant or prior therapy, disease stage, age, gender, or race. Analysis of the results of such trials for consistency across key patient subsets addresses concerns about generalizability of findings to various populations in a manner that may not be possible with smaller trials or trials with more narrow entry criteria.'' [6]	For all trials, the primary outcome used for regulatory decision-making was determined on the basis of the most important endpoint used in the approval. If the primary outcome was not clearly specified, we used the order of the outcomes listed in the medical review (previous empirical studies on FDA reports indicated that primary outcomes are typically discussed first and in more detail in the ''review of efficacy'' sections [7, 8]). Then, we scrutinized all FDA documents to identify a discussion of subgroup effects on this primary endpoint, extracted FDA's statements about the consistency across demographic and disease subgroups (for example, across age groups, sex, race, geographic region, or disease stage), and classified the statements as indicating consistency or not for this endpoint.
Consistency of treatment effects across multiple endpoints	
''In some cases, a single study will include several important, prospectively identified primary or secondary endpoints, each of which represents a beneficial, but different, effect. Where a study shows statistically persuasive evidence of an effect on more than one of such endpoints, the internal weight of evidence of the study is enhanced.'' [6]	We determined whether the experimental treatment group had more favorable effects compared to the control arm (i.e., statistically significant effects or effects crossing a prespecified efficacy threshold) for all endpoints of overall survival (OS), progression-free survival (PFS), and objective response rate (RR), and for all primary endpoints used for regulatory decision-making (in cases where they would not be OS, PFS, or RR). Because this information is usually only available for comparative treatment effects, we specifically assessed this characteristic only for comparative trials (i.e., RCTs, not SATs), although we deemed the other trials as not corroborated by this characteristic per se. We extracted statements about statistical significance, *P*-values, alpha-levels, or thresholds (where applicable) and treatment effects. For *P*-values reported only as being below a cutoff (e.g., $P < 0.0001$), we used this cutoff value and we doubled *P*-values from one-sided hypothesis tests to consistently obtain two-sided *P*-values [42]. We classified treatment effects as ''statistically significant'' whenever explicitly declared by the FDA. In cases without an explicit statement, we deemed *P*-values below an explicitly defined alpha level ''statistically significant'' or when it was below 0.05 (in cases without a defined alpha level). Similarly, estimates not accompanied by *P*-values were considered statistically significant when the confidence interval excluded the null. When only one of these endpoints (i.e., the primary, OS, PFS, RR) had statistically significant favorable effects, we evaluated health-related quality of life (HRQoL) as one further patient-important outcome and perused the approval documents for any statements about statistically significant favorable effects on HRQoL. HRQoL was captured as patient-reported outcomes on the perceived effect of the treatment on any domain (e.g., physical, social, and mental health, well-being, symptoms), typically measured globally or within specific domains through QoL scales such as EQ-5D and EORTC QLQ. We deemed this characteristic met when at least two endpoints (i.e., the primary, OS, PFS, RR or HRQoL) had more favorable effects.

(*Continued*)

Table 2. Continued

Consistency of treatment effects across treatment comparisons	
''Properly designed factorial studies may be analyzed as a series of pairwise comparisons, representing, within a single study, separate demonstrations of activity of a drug [...]'' [6]	We evaluated whether a second treatment comparison within the same trial (in the FDA guidance document described as ''multiple studies in a single study'') confirmed a beneficial effect on the primary endpoint. Only multiarm or factorial design RCTs were assessed for this characteristic.
Statistically very persuasive findings	
''In a multicenter study, a very low *P*-value indicates that the result is highly inconsistent with the null hypothesis of no treatment effect [...]'' [6]	We categorized findings on the primary endpoint used for regulatory decision-making as ''very persuasive'' when they favored the experimental treatment with a *P*-value of 0.00125 or less. A *P*-value of 0.00125 is frequently discussed in the context of single pivotal trials [20–23] and is statistically equivalent to running two independent trials where each trial demonstrates a statistically significant effect in favor of the experimental treatment at the conventional 0.05 level [20–23]. Because it required comparative treatment effects, we also assessed this only for comparative trials (RCTs) and deemed the other trials as not corroborated by this characteristic per se.
Alternative operationalization in sensitivity analyses	
Because there is no clear operationalization for these characteristics, we varied our approach in sensitivity analyses We used other thresholds (i.e., 500 and 1,000 participants) for the ''large multicenter'' criterion. We also considered PFS (and event-free survival, the primary endpoint in the trial used for approval of UNITUXIN®/dinutuximab) not being sufficiently different to OS because survival is an integral part of the outcome definition of PFS. When then only one of the remaining endpoints (i.e., the primary, OS, RR) had statistically significant favorable effects, we also searched for potential benefits on HRQoL. We also searched for any *P*-value < 0.05 reported in relationship to effects on the primary outcome, OS, PFS, RR, or HRQoL in reviews of approvals based on SATs and counted such studies as fulfilling the ''Consistency of treatment effects across multiple endpoints'' and ''statistically very persuasive'' criterion.	

Abbreviations: FDA, Food and Drug Administration; RCT, randomized controlled trial; SAT, single-arm trial; OS, overall survival; PFS, progression-free survival; RR, response rate; HRQoL, health-related quality of life.

1,000 participants). The ''large multicenter study'' characteristic was present for 66% of all approvals (23/35) (Table 4).

3.3. Presence of statistical significance

In all RCTs, statistically significant benefits for the primary endpoint used for regulatory decision-making were found, which was PFS in 10 (50%) and OS in seven of the 20 RCTs (35%). The other 3 RCTs showed beneficial effects on RR and PFS as coprimary endpoints (TREANDA®/bendamustine hydrochloride), medical castration rate in prostate cancer (FIRMAGON®/degarelix acetate, where efficacy was established without *P*-value), and event-free survival (UNITUXIN®/dinutuximab where the pivotal trial was stopped early after the 7th interim analysis indicated a *P*-value that was close to the prespecified stopping boundary but did not formally cross it).

3.4. Consistency of treatment effects across subgroups

Treatment benefits were consistent across subgroups in 66% (23/35) of the trials (Table 4).

3.5. Consistency of treatment effects across endpoints

Treatment benefits were consistent across multiple endpoints (i.e., OS, PFS, RR; and event-free survival as different primary endpoint in one case) in 15 of the 20 RCTs. Thus, of all 35 approvals based on single pivotal trial evidence (RCT or SATs), 43% (15/35) were supported by trial evidence with this corroborating characteristic (Table 4).

3.6. Consistency of treatment effects across multiple comparisons

Only one trial (3%) of 35 had a multiarm or factorial design and evaluated the efficacy of two different maintenance doses against an active treatment control. Both

Table 3. Characteristics of cancer indications approved based on single pivotal trials

Characteristic	Overall	RCTs	SATs
Indications (*n*, %)	35 (100)	20 (56)	15 (43)
Orphan indication (*n*, %)	23 (66)	10 (50)	13 (87)
Accelerated approval (*n*, %)	14 (40)	3 (15)	11 (73)
Breakthrough program (*n*, %)	7 (20)	4 (20)	3 (20)
Priority review (*n*, %)	27 (77)	18 (90)	9 (60)
Fast track program (*n*, %)	16 (46)	10 (50)	6 (40)
Approval period (*n*, %)			
<2006 (6 yr)	1 (3)	1 (5)	0 (0)
2006 - 2011 (6 yr)	11 (31)	6 (30)	5 (31)
2012 - 2016 (5 yr)	23 (66)	13 (65)	10 (67)
Single pivotal trial			
Median number of enrolled patients (IQR); [range]	301 (IQR 125, 571); [58 to 1,226]	460 (316, 758) [133 to 1,226]	119 (104, 230) [58 to 571]
Median number of participating centers (IQR); [range][a]	63 (IQR 31, 113); [16 to 184]	90 (80, 147) [16 to 184]	31 (25, 41) [17 to 79]

Abbreviations: RCTs, randomized controlled trials; SAT, single-arm trials; IQR, interquartile range.
[a] Information about the number of trial sites was not available for one single pivotal trial.

treatment comparisons were deemed to be effective by the FDA, thus demonstrating consistency across treatment comparisons (Table 4).

3.7. Proportion of single pivotal trials with "statistically very persuasive" results

For almost all RCTs, small *P*-values were reported for the primary endpoint used for regulatory decision-making with a median two-sided *P*-value of 0.0001 (IQR: 0.0001, 0.0108; $n = 19$, for one trial, we identified no *P*-value). For 10 RCTs, *P*-values were reported as "<0.0001." Overall, the *P*-values of 12 RCTs were smaller than 0.00125 (13 RCTs < 0.005, 18 RCTs < 0.05 (Table 4). Of all 35 approvals based on single pivotal trial evidence (RCT or SATs), 34% (12/35) were supported by trial evidence having this corroborating characteristic (Table 4).

3.8. Corroborating characteristics in single pivotal trials

Overall, 33 (94%) of the 35 approvals based on a single pivotal trial had at least one corroborating characteristic (Table 4). Only one characteristic was found in 14 trials (40%), two in five trials (16%), three in six trials (17%), and four characteristics were present in eight trials (23%). When there was only one characteristic, this was typically the "consistent across subgroups" or the "large multicenter" criterion (8 and 5 trials, respectively). There was no trial with all characteristics but two trials had none (6%). The median number of corroborating characteristics per trial was 2 (IQR 1; 3).

3.9. Sensitivity and subset analyses

Using an alternative approach to define the characteristics necessary for a large trial (i.e., 1,000 patients) and considering PFS and event-free survival as not sufficiently different to OS, 77% (27/35) approvals were supported by a single pivotal trial with at least one characteristic. Using a cutoff of 500 patients, results were similar (81%, 29/35). Considering SATs for the "statistically persuasive" or "consistent across endpoints" criterion when *P*-values related to efficacy outcomes were below 0.05 (found in 4/15 SATs) did also not change the overall interpretation (Appendix). Approvals with orphan status had more often only a single characteristic and not multiple ones (61%, 14/23), but the overall interpretation was not different (Appendix).

4. Discussion

Our evaluation of FDA approvals over almost 2 decades indicates that approximately one in three novel drugs for cancer is based on evidence from only a single pivotal trial without any further "supporting evidence" from related trials. About half of the single pivotal trials were nonrandomized, constantly over time. Most approvals were supported by trials having at least one of the corroborating characteristics, but many of them by only one alone, in particular, when they had orphan status. The overall number of approvals based on single pivotal trial evidence increased substantially over time and likewise the need for a better understanding of the associated uncertainty, challenges, and limitations.

Single pivotal trials providing the only evidence for the establishment of drug efficacy always involved multiple centers with a median of 301 patients, but rarely more than 500 patients. Consistent effects across patient subgroups were present in most trials according to FDA reviewers. Fewer approvals were supported by trials with consistent effects across endpoints or by statistically very persuasive findings.

Table 4. Corroborating characteristics of single pivotal trial evidence

Proprietary drug name (generic drug name): approved disease	Design	Large, multicenter (No of patients)	Consistency across subgroups	Consistency across endpoints: Beneficial effects for OS	PFS	RR	Multiple EP	Consistency across comparisons	Stat. very persuasive findings	Total no of characteristics
ADCETRIS® (*brentuximab vedotin*): anaplastic large cell lymphoma	SAT	No (58)	Yes	NA	NA	NA	NA	NA	NA	1
ADCETRIS® (*brentuximab vedotin*): hodgkin lymphoma	SAT	No (102)	Yes	NA	NA	NA	NA	NA	NA	1
AFINITOR® (*everolimus*): advanced renal cell carcinoma	RCT	Yes (416)	Yes	No	Yes[p]	No	No[2]	NA	Yes	3
ALIMTA® (*pemetrexed disodium*): malignant pleural mesothelioma, combined with cisplatin	RCT	Yes (456)	No[4]	Yes[p]	NR	Yes	Yes	NA	No	2
BOSULIF® (*bosutinib monohydrate*): relapsed or refractory chronic, accelerated or blast phase philadelphia chromosome-pos. CML	SAT	Yes (571)	No[4]	NA	NA	NA	NA	NA	NA	1
COMETRIQ® (*cabozantinib s-malate*): progressive, metastatic medullary thyroid cancer	RCT	Yes (330)	Yes	No	Yes[p]	Yes	Yes	NA	Yes	4
COTELLIC® (*cobimetinib*): unresectable or metastatic melanoma with BRAF V600E or V600K mutation, combined with vemurafenib	RCT	Yes (495)	Yes	No	Yes[p]	Yes	Yes	NA	Yes	4
ERIVEDGE® (*vismodegib*): locally advanced or metastatic basal cell carcinoma	SAT	No (104)	No[4]	NA	NA	NA	NA	NA	NA	0
ERWINAZE® (*asparaginase Erwinia chrysanthemi*): acute lymphoblastic leukemia, part of multi-agent chemotherapy	SAT	No (59)	NR	NA	NA	NA	NA	NA	NA	0
FIRMAGON® (*degarelix acetate*): advanced prostate cancer	RCT	Yes (620)	NR	No	No	Yes[p]	Yes[3]	Yes	No	3
FOLOTYN® (*pralatrexate*): relapsed or refractory peripheral T-cell lymphoma	SAT	No (109)	Yes	NA	NA	NA	NA	NA	NA	1
GAZYVA® (*obinutuzumab*): CD20+ CLL, combined with chlorambucil	RCT	Yes (356)	Yes	No	Yes[p]	NR	No[2]	NA	Yes	3
GILOTRIF® (*afatinib*): EGFR mutation (exon 19 deletion or L858R)-positive metastatic non-small cell lung cancer	RCT	Yes (345)	No[4]	No	Yes[p]	NR	No[2]	NA	Yes	2
IBRANCE® (*palbociclib*): ER+/HER2- advanced breast cancer, combined with letrozole	RCT	No (165)	Yes	No	Yes[p]	NR	No[2]	NA	Yes	2
ICLUSIG® (*ponatinib* hydrochloride): chronic phase, accelerated phase, or blast phase CML	SAT	Yes (444)	NR	NA	NA	NA	NA	NA	NA	1
IMBRUVICA® (*ibrutinib*): mantle cell lymphoma	SAT	No (111)	Yes	NA	NA	NA	NA	NA	NA	1
JEVTANA KIT® (*cabazitaxel*): hormone-refractory metastatic prostate cancer	RCT	Yes (755)	Yes	Yes[p]	Yes	Yes	Yes	NA	Yes	4
KEYTRUDA® (*pembrolizumab*): unresectable or metastatic melanoma	SAT	No (173)	Yes	NA	NA	NA	NA	NA	NA	1
KYPROLIS® (*carfilzomib*): multiple myeloma	SAT	Yes (266)	NR	NA	NA	NA	NA	NA	NA	1
LARTRUVO® (*olaratumab*): soft tissue sarcoma with histologic subtype for which an anthracycline-containing regimen is appropriate	RCT	No (133)	No[4]	Yes	Yes[p]	NR	Yes	NA	No	1
NINLARO® (*ixazomib*): multiple myeloma, combined with lenalidomide and dexamethasone	RCT	Yes (722)	No[4]	No	Yes[p]	NR	No[2]	NA	No	1
ODOMZO® (*sonidegib*): locally advanced basal cell carcinoma	SAT	Yes (230)	Yes	NA	NA	NA	NA	NA	NA	2
PORTRAZZA® (*necitumumab*): metastatic squamous non-small cell lung cancer, combined with gemcitabine and cisplatin	RCT	Yes (1093)	No[4]	Yes[p]	Yes	NR	Yes	NA	No	2
STIVARGA® (*regorafenib*): metastatic colorectal cancer	RCT	Yes (760)	Yes	Yes[p]	Yes	No	Yes	NA	No	3
TASIGNA® (*nilotinib hydrochloride monohydrate*): chronic phase and accelerated phase philadelphia chromosome-positive CML	SAT	Yes (385)	No[4]	NA	NA	NA	NA	NA	NA	1
TREANDA® (*bendamustine hydrochloride*): CLL	RCT	Yes (301)	Yes	No	Yes[p]	Yes[p]	Yes	NA	Yes	4
UNITUXIN® (*dinutuximab*): pediatric patients with high-risk neuroblastoma, combined with GM-CSF, IL-2, and 13-cis-retinoic acid	RCT	Yes (251)	Yes	No	Yes	NR	Yes[3]	NA	No	3
VECTIBIX® (*panitumumab*): EGFR-expressing, metastatic colorectal carcinoma	RCT	Yes (463)	Yes	No	Yes[p]	Yes	Yes	NA	Yes	4
VENCLEXTA® (*venetoclax*): 17p-deletion CLL	SAT	No (107)	Yes	NA	NA	NA	NA	NA	NA	1
XTANDI® (*enzalutamide*): metastatic castration-resistant prostate cancer	RCT	Yes (1199)	Yes	Yes[p]	Yes	NR	Yes	NA	Yes	4
ZALTRAP® (*ziv-aflibercept*): metastatic colorectal cancer, combined with 5-fluorouracil, leucovorin, and irinotecan	RCT	Yes (1226)	Yes	Yes[p]	Yes	Yes	Yes	NA	No	3
ZYDELIG® (*idelalisib*): relapsed CLL, combined with rituximab	RCT	Yes (220)	Yes	No	Yes[p]	Yes	Yes	NA	Yes	4
ZYDELIG® (*idelalisib*): relapsed follicular B-cell non-Hodgkin lymphoma[1]	SAT	No (125)	Yes	NA	NA	NA	NA	NA	NA	1
ZYDELIG® (*idelalisib*): relapsed small lymphocytic lymphoma[1]	SAT	No (125)	Yes	NA	NA	NA	NA	NA	NA	1
ZYTIGA® (*abiraterone acetate*): metastatic castration-resistant prostate cancer	RCT	Yes (1195)	Yes	Yes[p]	Yes	Yes	Yes	NA	Yes	4
Total with characteristic n (%)		23 (66)	23 (66)	8 (23)	18 (51)	10 (29)	15 (43)	1 (3)	12 (34)	None: 2 (6%) One: 14 (40%) Two: 5 (16%) Three: 6 (17%) Four: 8 (23%) Five: 0 (0%)

(1) These two indications for two cancers are based on efficacy data from the same single pivotal trial.
(2) No statements indicating benefits on quality of life found
(3) Characteristic met because beneficial effects also shown for (co-)primary endpoint.
(4) FDA statement not conclusive or interpreted by us not to be consistent across subgroups.

Abbreviations: CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; EGFR, Epidermal Growth Factor Receptor; EP, endpoint; FDA, Food and Drug Administration; GM-CSF, granulocyte-macrophage colony–stimulating factor; IL, interleukin; NA, not applicable due to study design; NR, not reported; OS, overall survival; PFS, progression-free survival; p, primary endpoint used for regulatory decision-making; RR, response rate; RCT, randomized controlled trial; SAT, single-arm trial.

Surprisingly, for two approvals (ERIVEDGE®/vismodegib and ERWINAZE®/asparaginase Erwinia chrysanthemi), we found none of the corroborating characteristics that are specifically listed in the FDA guidance document. ERIVEDGE®/vismodegib was approved for a serious and rare skin cancer without an available FDA-approved treatment and the approval history of ERWINAZE®/Asparaginase Erwinia chrysanthemi is rather complex and began in 1968. This may reflect various special approval circumstances. Of note, the FDA guidance authors clearly emphasize that it is not "a complete listing of the circumstances in which existing efficacy data may provide independent substantiation of related claims."

On the other hand, beyond these five characteristics, we identified no further specific criteria mentioned in the perused approval documents. Actually, there were only two approval packages in which we found a clear summary referring and discussing all five characteristics (Box 1).

We report, to our knowledge for the first time, the prevalence of these five characteristics in single pivotal trials submitted to the FDA in support for the approval of novel drugs for cancer indications over 17 years. The number of oncological approvals supported by single trials has been explored in several other metaepidemiological surveys before [12–17], but none specifically focused on cases where there is no supportive evidence.

Several limitations of this study merit closer attention. First, the approval packages are large documents and very complex, and despite our efforts to standardize and double-check our extractions, we cannot rule out possible

Box 1 Example of explicit and transparent summarizing statement

"The primary issue considered during the review of this application was whether the results of a single adequate and well-controlled trial were sufficient to support approval. FDA guidance (Reference) identified characteristics that can contribute to the conclusion that results from a single study can support an efficacy claim. The characteristics identified were (a) large multicenter study; (b) consistency across study subsets; (c) multiple studies in a single study; (d) multiple endpoints involving different events; and (e) statistically very persuasive findings. Results of the VELOUR trial submitted in support of this Biologics License Application (BLA) satisfied all of these characteristics except (c)".

Cross Discipline Team Leader Review; ZALTRAP®/ziv-aflibercept https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/125418Orig1s000CrossR.pdf).

extraction errors. We cross-checked the pivotal trials used in this analysis with the database of our related project [7]. Any residual extraction errors would have minor impact, and the overall interpretation would be unaffected.

Second, the approval documents over almost 2 decades had different structures and levels of transparency, and sometimes documents (such as summary reviews) were not available in the drugs@FDA database. It was not always clear to us which evidence on efficacy was exactly used for the regulatory decision, for example, when overview tables of evaluated studies were redacted or when statements such as "the main study for efficacy assessment was…" left it open if there was other pertinent evidence. However, in some cases, we found more than a single study provided by the manufacturer but the FDA classified them as "nonsupportive." Then we kept these cases in our sample to reflect the regulatory decision. We aimed to focus on the approval situations where no supporting evidence on the drug's efficacy came from other studies and thus the five characteristics would probably have the highest impact on regulatory decisions. We had expected that such methodological characteristics would be more transparently reported, similar to risk of bias or study quality ratings in systematic reviews and health technology assessments. Although we carefully perused all the documents, we cannot rule out that we missed such supporting evidence. Again, we believe the overall interpretation would be unaffected.

Third, the five characteristics and their description in the guidance documents are rather vague, and we could not rely on a clear definition. Therefore, several of our operationalizations remain to some extent arbitrary. There is no established cutoff to distinguish between small and large trials [18], so we used a threshold of 200 patients, which was used previously to estimate the robustness of effect estimates depending on the size of oncology trials [19] and selected further thresholds in sensitivity analyses. For the "multiple endpoints involving different events" characteristic, we evaluated the primary endpoints used for regulatory decision-making, systematically perused OS, PFS, and RR as the most frequently used and probably most important outcomes for approval of cancer drugs, together with quality of life, which we also considered. Almost all of them were primary or secondary endpoints in the trials (according to clinicaltrials.gov). However, a beneficial effect should be demonstrated across independent endpoints, and it is frequently argued in discussions about outcome surrogacy that OS, PFS, and RR correlate well and that the FDA often accepts, for example, PFS as a surrogate outcome of OS. We addressed this issue in a sensitivity analysis, which did not alter the main interpretation. For the "statistically very persuasive finding" characteristic, the FDA does not provide a threshold of what they consider to constitute a "very persuasive" effect. Our cutoff of 0.00125 is often proposed as the appropriate level of statistical evidence to reach for regulatory decision-making in single pivotal trials [20–23]. However, in the two rare cases where the FDA reviewer described the corroborating characteristics explicitly, the *P*-values were 0.0032 (ZALTRAP®/ziv-aflibercept) and 0.01 (STIVARGA®/regorafenib). We deemed this to be special situations because FDA authors stated that they "ordinarily have said that a value in the neighborhood of 0.001 is good enough for a single trial" [24]. Because *P*-values were often reported imprecisely as "smaller than a cutoff" and we did not know the true, smaller, *P*-value, our analyses may underestimate the median *P*-value. Because half of the endpoints were reported as "<0.0001" and the two largest ones were 0.06 and 0.02, even when we had used a stricter cutoff, this would not have changed our interpretation.

Fourth, we focused on selected features of pivotal studies and evaluated only a fraction of factors that are relevant for approval decisions and benefit assessments of novel treatments. Because we did not evaluate treatment effects, clinical impact of findings, validity of comparators or outcomes, other types of bias, or further relevant aspects, our findings do not provide information about the benefits and harms of these drugs that would be beyond the scope of this project.

Finally, we have not compared the prevalence of these characteristics with a control group of other studies. We do not know about applications where single pivotal trials did not suffice for licensing because no successful applications are represented in the published database. We also do not know how often a second pivotal trial was initiated in reaction of insufficient evidence from a single pivotal trial alone.

Overall, we found that most of the single pivotal studies have at least one of the corroborating characteristics. They often have many centers per trial but typically have less than 500 or 1,000 patients what other investigators would consider ''large'' [25−29]. Multicenter trials provide on average smaller effects than single-center trials [30], but smaller trials exaggerate treatment effects compared with larger trials [18,31−35]. They may be underpowered, more prone to publication bias or have lower methodological quality, which may explain some of the exaggeration [32,33,35,36]. One might argue that in regulatory settings publication biases play no role, methodological quality may be better, and therefore this characteristic might be of less relevance. Furthermore, the rationale for this characteristic in the guidance focuses less on sample size and more on the number of centers and on the generalizability to different care settings. Our findings may indeed indicate a better generalizability of pivotal trials to different care settings than sometimes assumed, albeit we have not assessed the characteristics of the trial populations. The characteristic of consistent effects across subgroups is problematic because detection of subgroup effects requires sufficient statistical power [37] and absence of evidence for subgroup effects does not provide proof of consistent effects. Subgroup effects in trials are often not credible and spurious [38], but they may be interpreted as indications for inconsistent effects in the framework of these characteristics. For example, sex-based subgroup differences are often discussed in approval documents, but subgroup findings from individual randomized trials are rarely corroborated in meta-analyses [39]. Thus, this characteristic may falsely corroborate or falsely weaken the approval evidence. The consistency of effects across outcomes is difficult to interpret as discussed previously, and the corroboration of the result by consistent findings in multiple comparisons is rare.

The fact that these single trials represent the first evidence on benefits also requires further consideration because even when the studies truly show benefits, this observed benefit may be substantially inflated because of regression to the mean and related effects [32,40,41].

5. Conclusion

Approval of several modern cancer treatments is increasingly based on only a single pivotal trial and often without any further supporting evidence on beneficial effects. Such trials have typically one of five characteristics described by the FDA before to corroborate the approval evidence, but a clearer operationalization and definition of these characteristics along with a more structured and transparent reporting of their use would be helpful. Whether single trials with such characteristics provide similar evidence about benefits and harms of novel treatments as multiple trials would do remains to be shown.

Acknowledgments

The authors thank Amanda Herbrand, University Hospital Basel, for her support with the data extraction.

None of the funders/sponsors had a role in the design and conduct of the project and preparation, review, approval of the manuscript and decision to submit the manuscript for publication.

No further data available. The study data set is available from the corresponding author.

Authors' contribution: A.L. conceived the study with input from L.G.H. A.L., B.S., M.B., A.A., T.V.P., B.K., and L.G.H. collected and verified the data. A.L. and L.G.H. conducted the analyses. All authors interpreted the results. A.L. wrote the first draft, and all authors made critical revisions on to the manuscript. All authors read and approved the final version of the paper.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2019.05.033.

References

[1] Krist AH. "Needs more research"-Implications of the Proteus effect for researchers and evidence adopters. Mayo Clin Proc 2018;93(3):273−5.

[2] Naci H, Wouters OJ, Gupta R, Ioannidis JPA. Timing and characteristics of cumulative evidence available on novel therapeutic agents receiving Food and Drug Administration accelerated approval. Milbank Q 2017;95(2):261−90.

[3] Mullins CD, Montgomery R, Tunis S. Uncertainty in assessing value of oncology treatments. Oncologist 2010;15(Suppl 1):58−64.

[4] Kefauver-Harris Amendments of 1962. Amendment to section 505(d) of the FD&C Act (21 USC 355(d)).

[5] Section 115 of the FDA modernization act of 1997. Amendment to section 505(d) of the FD&C act (21 USC 355(d)).

[6] US Food and Drug Administration. Guidance for industry: providing clinical evidence of effectiveness for human drug and biological products. 1998. https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072008. Accessed June 21, 2019.

[7] Ladanie A, Speich B, Naudet F, Agarwal A, Pereira TV, Sclafani F, et al. The Comparative Effectiveness of Innovative Treatments for Cancer (CEIT-Cancer) project: rationale and design of the database and the collection of evidence available at approval of novel drugs. Trials 2018;19(1):505.

[8] Ladanie A, Ewald H, Kasenda B, Hemkens LG. How to use FDA drug approval documents for evidence syntheses. BMJ 2018;362:k2815.

[9] US Food and Drug Administration. Drugs@FDA: FDA approved drug products. https://www.fda.gov/drugsatfda.

[10] US Food and Drug Administration. New Drugs at FDA: CDER's new molecular entities and new therapeutic biological products. 2018. https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/default.htm. Accessed June 21, 2019.

[11] US Food and Drug Administration. CDER fast track products Approved Since 1998 through June 1, 2010. https://wayback.archive-it.org/7993/20170404174211/https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedand

Approved/DrugandBiologicApprovalReports/UCM216527.pdf. Accessed June 21, 2019; 2010.
[12] Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. JAMA 2014;311:368−77.
[13] Sridhara R, Johnson JR, Justice R, Keegan P, Chakravarty A, Pazdur R. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between july 2005 and December 2007. J Natl Cancer Inst 2010;102:230−43.
[14] Martell RE, Sermer D, Getz K, Kaitin KI. Oncology drug development and approval of systemic anticancer therapy by the U.S. Food and Drug Administration. Oncologist 2013;18(1):104−11.
[15] Morant AV, Vestergaard HT. European marketing authorizations granted based on a single pivotal clinical trial: the rule or the exception? Clin Pharmacol Ther 2017;104:169−77.
[16] Gentry L. One and done: are single pivotal studies the new norm in cancer therapeutics? 2015. https://www.ask-cato.com/one-and-done-are-single-pivotal-studies-the-new-norm-in-cancer-therapeutics/. Accessed June 21, 2019.
[17] Tibau A, Molto C, Ocana A, Templeton AJ, Del Carpio LP, Del Paggio JC, et al. Magnitude of clinical benefit of cancer drugs approved by the US Food and Drug Administration. J Natl Cancer Inst 2017;110:486−92.
[18] Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. JAMA 1998;279:1089−93.
[19] Cheng J, Zhang H, Tang S, Sridhara R. Inference based on small randomized oncology clinical trials: is the observed treatment effect true? Int J Clin Biostat Biom 2017;3:010.
[20] Deng CQ. Significant level of 0.00125. http://onbiostatistics.blogspot.ch/2010/01/significant-level-of-000125.html. Accessed June 21, 2019.
[21] Fisher L. One large, well-designed, multicenter study as an alternative to the usual FDA paradigm. Drug Inf J 1999;33(1):265−71.
[22] Senn S. Statistical issues in drug development. Chapter 12.2.8.: The two-trials rule. 2nd ed. Chichester: John Wiley; 2007.
[23] Shun Z, Chi E, Durrleman S, Fisher L. Statistical consideration of the strategy for demonstrating clinical evidence of effectiveness-one larger vs two smaller pivotal studies. Stat Med 2005;24:1619−37.
[24] Temple R. How FDA currently makes decisions on clinical studies. Clin Trials 2005;2:276−81.
[25] Sivakumar H, Peyton PJ. Poor agreement in significant findings between meta-analyses and subsequent large randomized trials in perioperative medicine. Br J Anaesth 2016;117:431−41.
[26] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001;135:982−9.
[27] LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997;337:536−42.
[28] Jones CW, Handler L, Crowell KE, Keil LG, Weaver MA, Platts-Mills TF. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ 2013;347:f6104.
[29] Myles PS. Why we need large randomized studies in anaesthesia. Br J Anaesth 1999;83:833−4.
[30] Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med 2011;155:39−51.
[31] Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.
[32] Ioannidis JP. Why most discovered true associations are inflated. Epidemiology 2008;19:640−8.
[33] Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. JAMA 2012;308:1676−84.
[34] Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
[35] Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One 2013;8:e59202.
[36] Gluud LL, Thorlund K, Gluud C, Woods L, Harris R, Sterne JA. Correction: reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2008;149:219.
[37] Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57:229−36.
[38] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ 2012;344:e1553.
[39] Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW, Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. BMJ 2016;355:i5826.
[40] Alahdab F, Farah W, Almasri J, Barrionuevo P, Zaiem F, Benkhadra R, et al. Treatment effect in earlier trials of patients with chronic medical conditions: a meta-epidemiologic study. Mayo Clin Proc 2018;93(3):278−83.
[41] Lord EM, Weir IR, Trinquart L. Design analysis indicates potential overestimation of treatment effects in randomized controlled trials supporting Food and Drug Administration cancer drug approvals. J Clin Epidemiol 2018;103:1−9.
[42] Senn S. Statistical issues in drug development. Chapter 12.2.7: Should the two-sided p-value always be twice the one-sided value? 2nd ed. Chichester: John Wiley; 2007.