



OPEN ACCESS

Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review

Alison M Pease,¹ Harlan M Krumholz,^{2,3,4,5} Nicholas S Downing,⁶ Jenerius A Aminawung,⁷ Nilay D Shah,⁸ Joseph S Ross^{3,4,5,7}

For numbered affiliations see end of article.

Correspondence to: J S Ross joseph.ross@yale.edu.

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2017;357:j1680 <http://dx.doi.org/10.1136/bmj.j1680>

Accepted: 23 March 2017

ABSTRACT

OBJECTIVE

To characterise the prospective controlled clinical studies for all novel drugs that were initially approved by the Food and Drug Administration on the basis of limited evidence.

DESIGN

Systematic review.

DATA SOURCES

Drugs@FDA database and PubMed.

STUDY INCLUSION

All prospective controlled clinical studies published after approval for all novel drugs initially approved by the FDA between 2005 and 2012 on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease as primary endpoints, or both.

RESULTS

Between 2005 and 2012 the FDA approved 117 novel drugs for 123 indications on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease, or both (single surrogate trials). We identified 758 published controlled studies over a median of 5.5 years (interquartile range 3.4-8.2) after approval, most of which (554 of 758; 73.1%) were studies for indications approved on the basis of surrogate markers of disease. Most postapproval studies used active comparators—67 of 77 (87.0%) indications approved on the basis of single pivotal trials, 365 of 554 (65.9%) approvals based on

surrogate marker trials, and 100 of 127 (78.7%) approvals based on single surrogate trials—and examined surrogate markers of efficacy as primary endpoints—51 of 77 (66.2%), 512 of 554 (92.4%), and 110 of 127 (86.6%), respectively. Overall, no postapproval studies were identified for 43 of the 123 (35.0%) approved indications. The median total number of postapproval studies identified was 1 (interquartile range 0-2) for indications approved on the basis of a single pivotal trial, 3 (1-8) for indications approved on the basis of pivotal trials that used surrogate markers of disease as primary endpoints, and 1 (0-2) for single surrogate trial approvals, and the median aggregate number of patients enrolled in postapproval studies was 90 (0-509), 533 (122-3633), and 38 (0-666), respectively. The proportion of approved indications with one or more randomized, controlled, double blind study using a clinical outcome for the primary endpoint that was published after approval and showed superior efficacy was 18.2% (6 of 33), 2.0% (1 of 49), and 4.9% (2 of 41), respectively.

CONCLUSIONS

The quantity and quality of postapproval clinical evidence varied substantially for novel drugs first approved by the FDA on the basis of limited evidence, with few controlled studies published after approval that confirmed efficacy using clinical outcomes for the original FDA approved indication.

Introduction

In the US, the Food and Drug Administration determines whether a new drug is sufficiently safe and effective to be made available to doctors for use by patients.¹ To do this, it must find a balance between requiring sufficient high quality clinical evidence from premarket evaluation and allowing promising new drugs to enter the marketplace quickly with continued evaluation after approval. The FDA maintains a “usual requirement” of “more than one” well controlled clinical trial that independently proves a drug’s efficacy.² However, it also describes several situations in which fewer trials or studies with non-clinical outcomes, such as surrogate markers of disease, might suffice for premarket evaluation.² Thus, FDA approval is binary, but the clinical trial evidence that forms the basis of the FDA’s decision varies widely.

From 2005 to 2012 the FDA approved 188 new drugs for 206 indications, more than a third of which were approved on the basis of a single pivotal trial and 44% of which were approved on the basis of pivotal trials that used surrogate markers of disease instead of clinical outcomes for primary endpoints (41 approvals met both criteria).³

WHAT IS ALREADY KNOWN ON THIS SUBJECT

The US Food and Drug Administration frequently approves new drugs on the basis of either a single pivotal trial or pivotal trials that used surrogate markers of disease instead of clinical outcomes for primary endpoints

These drugs are used widely after approval, even if subsequent clinical studies are not conducted to confirm the expected benefits

WHAT THIS STUDY ADDS

For 117 novel drugs approved by the FDA for 123 indications on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease, or both, the quantity and quality of postapproval clinical evidence varied substantially

After a median period of 5.5 years after approval, the median total number of postapproval clinical studies of the same indication for which the drug was first approved by the FDA was 1 (interquartile range 0-2), 3 (1-8), or 1 (0-2) for drugs approved on the basis of a single pivotal trial, surrogate markers of disease, or both, respectively

Only 18.2% (six of 33), 2.0% (one of 49), and 4.9% (two of 41) indications approved by the FDA on the basis of a single pivotal trial, surrogate markers, and both, respectively, had ≥1 published postapproval, randomized controlled, double blind study that used a clinical outcome for the primary endpoint and showed superior efficacy

Box 1: Lifecycle evaluation

Emphasizes the importance of continuing to collect data on the effectiveness and safety of medical products after approval for a given indication over the entire span of their use by patients

Involves ongoing review of the published literature, adverse event reporting systems, manufacturer safety reports, and drug use databases

Specific methods or intensity of postapproval evaluation can vary based on what is known about the benefits and risks of each drug

Contrasts with historical concept of regulators as completing product evaluation exclusively (or near exclusively) by the time of initial regulatory approval

Regulation has historically occurred primarily in the period before a drug is approved and becomes widely available. Over the past decade, however, the FDA has adopted “lifecycle evaluation” (box 1).^{4,5} This approach was originally proposed by the Institute of Medicine in its 2007 report, *The Future of Drug Safety*, which was commissioned by the FDA after a series of well publicized postmarket drug safety concerns raised questions about drug safety in the US.^{4,6} In addition to monitoring drug safety, the lifecycle evaluation approach enables regulators to approve drugs based on clinical evidence that is less robust, with the understanding that drugs will continue to be evaluated after approval.

As the FDA continues to use lifecycle evaluation,⁶⁻⁹ we need a better understanding of the accumulation of clinical evidence after approval. Potential sources of postapproval clinical information could include prospective clinical studies, such as randomized controlled trials; observational studies of existing data, such as administrative claims; and adverse event reporting through programs such as the Sentinel System.¹⁰ The FDA also has formal postmarketing requirements (statutory requirements used when the FDA determines that an investigation is required to assess a known or potential serious safety risk) and postmarketing commitments (any other research agreed upon by manufacturers and the FDA that is not an investigation of a potential or known serious safety risk—such as any investigation that primarily evaluates efficacy—and mandates an annual report to the FDA), which frequently include postapproval clinical research studies.¹¹

Despite the FDA’s increased emphasis on generating postapproval evidence, whether prospective controlled trials for the same indication for which the drug was first approved are performed after approval, and whether those that are done confirm the benefits expected on the basis of the premarket pivotal trials, is unknown. For example, one study found that 31 of 36 cancer drugs approved on the basis of surrogate markers from 2008 to 2012 had either unknown effects on overall survival or failed to show a positive effect on survival after a median follow-up of over four years.¹² Moreover, even the postapproval studies required by the FDA are not always carried out.¹³⁻¹⁵

We sought to better understand how efficacy evidence from prospective controlled clinical trials accumulates in the years after approval for drugs initially approved on the basis of less robust evidence.

Methods**Study design and sample**

We used the Drugs@FDA database to identify all novel drugs approved by the FDA between 1 January 2005 and 31 December 2012 on the basis of either a single pivotal trial or pivotal trials focused on surrogate markers of disease, as described in previous work.³ Our sample included all new drug and biologic licensing applications for small molecule and biologic drugs, excluding generic drugs, reformulations, combination treatments of non-novel agents, and non-treatment agents such as diagnostic and contrast agents. Pivotal trials were those labeled as such in the FDA medical review. If the FDA review did not specify, we identified trials that were described as essential to approval, were otherwise prioritized in the review, or were new efficacy based trials provided as part of a resubmitted application for approval.

We also used the Drugs@FDA database to obtain the year of approval for each novel drug and whether it was considered a pharmacologic (small molecule) or biologic product. We used FDA approval letters, which are hyperlinked in the Drugs@FDA database, to identify orphan drugs, drugs approved through the accelerated approval pathway, and the indication for which all novel drugs were initially approved. Orphan drugs were those classified as such by the FDA, meaning drugs for rare diseases that were given extended market exclusivity. The Drugs@FDA database only indicates orphan status for biologics approved after 2010.

Indications were categorized by expected length of treatment using a previously developed framework:³ expected length of use of less than one month was acute, between one month and two years was intermediate, and more than two years was chronic. We used the World Health Organization’s anatomic therapeutic classification system, contextualized for clinical relevance, to categorize each indication into therapeutic areas,¹⁶ which we collapsed into four categories: cancer, cardiovascular disease and diabetes mellitus, infectious disease, and other.

Systematic review of postapproval trials

Two investigators (AMP, JAA) systematically searched all fields of Medline for the international non-proprietary name of each drug to identify all postapproval prospective studies in humans that used an active or placebo control and examined efficacy for the same indication for which the drug was first approved by the FDA. We searched for studies published between 1 January of the year before the drug was approved and 31 December 2014. This timeframe was chosen to find studies that started enrolling before FDA approval but did not produce data soon enough to be included in the FDA submission.

Although Medline is not a complete repository of all publications in all biomedical journals, it is the largest database of biomedical journal articles that can be searched freely using the PubMed system. Nearly all doctors and policy makers depend on it to learn about and obtain access to the findings of clinical trials. As such, it is the most important dissemination source

of clinical trial publications to inform and influence real world clinical practice.

We included all English language publications and put no exclusions on types of funder or investigator, including developers of the drug or its competitors, government agencies, and independent investigators.

Two investigators reviewed the titles and abstracts (AMP, JAA) and one (AMP) reviewed the full text to exclude the following: conference abstracts; case reports or series; studies performed on healthy people; pharmacokinetic or drug interaction studies; studies evaluating an indication substantially different from that for which the drug was initially approved; studies without an arm for placebo or active control (multi-arm studies of different dosages of the same drug were considered non-controlled); studies without efficacy endpoints described in the abstract (unless novel or unfamiliar surrogate markers in the abstract could be verified as examining efficacy in the full text); studies that were submitted to the FDA as part of the initial new drug application or were completed before FDA submission; and extension studies or new analyses of data from previously published studies.

We defined a substantially different indication narrowly, meaning that to exclude a publication it had to evaluate an entirely different disease. Investigations of merely a different population (children versus adults) or disease subtype (clear cell versus non-clear cell renal cell carcinoma) were included. Publications for which inclusion was questionable for any reason were reviewed by JSR, and conflicts were resolved by consensus.

Trial data abstraction

One investigator (AMP) abstracted the following information from each publication: electronic publication date (if not available, print publication date); national clinical trial number, if provided; whether funding was from industry or non-industry sources; whether authors were employees of the drug manufacturer; use of randomization; whether subject allocation was double blind, single blind, or open label; study design (superiority or non-inferiority); whether the comparator was placebo, active, or

both, with results for studies in the “both” category recorded for the comparison against the active agent; number of patients in the overall and drug arm intention to treat populations; study duration; whether the primary endpoint was a clinical outcome, a clinical scale, or a surrogate marker of disease; the primary endpoint result—that is, whether the drug was found to be statistically significantly better than (positive), no different from (neutral), or worse than (negative) the comparator in studies using a superiority design or whether non-inferiority was established (positive) or not established (negative) in studies using a non-inferiority design; whether secondary endpoints described in the abstract were clinical outcomes, clinical scales, surrogate markers of disease, or mixed; and the results of the secondary endpoints (using the same approach as for primary endpoints). One investigator (JAA) subsequently validated the characterization of studies in a 5% random sample, resolving conflicts by consensus. Any uncertainty was reviewed with JSR.

Primary trial endpoints were classified as clinical outcomes, clinical scales, or surrogate markers based on an established framework and a report from the Institute of Medicine (box 2).^{3,17,18} Clinical outcomes, such as death or admission to hospital, measure patient survival or function. Clinical scales, such as the Crohn’s disease activity index or the visual analog scale for pain, represent rubrics for the quantification of subjective patient reported symptoms. Surrogate markers, such as levels of glycosylated hemoglobin or hepatitis C ribonucleic acid, represent biomarkers expected to predict clinical benefit.

For studies with multiple primary endpoints, we categorized study results on the basis of clinical endpoints over surrogate markers, if both were reported. For studies with multiple primary endpoints that were all clinical or all surrogate marker and for studies with multiple secondary endpoints of any type, we categorized studies as statistically significantly better or worse based on whether they featured some non-significantly different endpoints and some significantly better or worse endpoints, respectively. The statistical significance of the result for each endpoint was abstracted directly from the publication. If statistical significance or lack thereof was not reported in the text and a significance threshold was not defined in the paper, we considered P values of less than 0.05 or completely non-overlapping confidence intervals to show a statistically significant difference. Two investigators (AMP, JSR) reviewed studies with mixed results—some significantly better and some significantly worse (or non-inferiority established and non-inferiority not established)—and decided the most appropriate categorization based on the relative clinical importance of the endpoints or, if equally important, the result supported by the preponderance of endpoints. We identified only two studies with multiple primary endpoints that were all surrogate or all clinical with mixed results.

Statistical analysis

All analyses were stratified based on whether the novel drug was first approved by the FDA on the basis of a single pivotal trial, multiple pivotal trials that focused on

Box 2: Brief description of trial endpoint classifications

Clinical outcome

Endpoint that reflects or characterizes a patient’s symptoms, overall mental state, or the effects of a disease or condition on how the patient functions

Examples: overall survival, admission to hospital, incidence of splenectomy, functional status

Clinical scale

Endpoint that reflects or characterizes a patient’s symptoms, overall mental state, or the effects of a disease or condition on how the patient functions using a rubric that quantifies subjective patient reported symptoms

Examples: Crohn’s disease activity index, visual analog scale for pain, blepharospasm disability index

Surrogate marker of disease

Endpoint that measures a biological characteristic that is expected to consistently and accurately predict a clinical outcome

Examples: glycosylated hemoglobin, hepatitis C ribonucleic acid, intraocular pressure

surrogate markers of disease for the primary efficacy endpoint, or both (a single pivotal trial that focused on surrogate markers). We used descriptive statistics to summarize the characteristics, design features, and findings of each study, as well as sample characteristics, aggregate number of trials, number of patients enrolled, and patient years of observation for each indication. We used χ^2 , Wilcoxon's rank sum, and Kruskal-Wallis tests to examine differences across the three categories of approval. We used χ^2 and Fisher's test of exact probability to examine differences in study findings within categories, comparing studies using surrogate markers of disease and clinical outcomes as primary endpoints. Analyses were repeated, stratifying by year of indication approval, therapeutic area, and expected length of treatment. All analyses were performed using Microsoft Excel 2013 and JMP 7.0.1 (SAS Institute; Cary, NC). All statistical tests were two tailed and used a type I error rate of 0.005 to account for the 10 comparisons made across categories of indication approval.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results to study participants or the relevant patient community.

Results

Sample characteristics

Of the 188 novel drugs approved by the FDA for 206 indications between 2005 and 2012, 30 novel drugs for

33 indications were approved on the basis of a single pivotal trial, 48 novel drugs for 49 indications on the basis of trials focused on surrogate markers of disease, and 40 novel drugs for 41 indications on the basis of a single pivotal trial that focused on surrogate markers. Our study sample comprised 117 unique novel drugs approved for 123 indications (one drug had two approvals in different categories). Most approvals, across all three categories, were for pharmacologic drugs for the treatment of chronic disease; a few were granted orphan status or approved through the accelerated pathway (table 1). The most common therapeutic area was cancer, which represented a third (41 of 123) of the overall sample. The proportion of pharmacologic drugs, accelerated approvals, therapeutic areas, and expected lengths of treatment (P values <0.02) showed statistically significant differences between categories, but orphan status and approval year did not. Median follow-up duration from the date of FDA approval to 31 December 2014 was 5.5 years (interquartile range 3.4-8.2).

Indications approved on the basis of a single pivotal trial

We performed systematic reviews of the literature for all 33 indications for the 30 novel drugs that were approved by the FDA on the basis of a single pivotal trial. We identified a total of 77 published postapproval controlled studies using an active or placebo control and examining efficacy for the same indication for which the drug was first approved (table 2). Of these, 81.8% (n=63) were randomized, 45.5% (n=35) used open label allocation, 66.2% (n=51) used primary efficacy endpoints that were surrogate markers of disease, 87.0% (n=67) used an active drug as control, 94.8% (n=73) used a superiority design, and 33.8% (n=26) were funded by industry. The median overall intention to treat population was 96 (interquartile range 56-260), and the median study duration was four weeks (interquartile range 0.5-26). Among the 60 superiority studies using active comparators, one of 13 (7.7%) studies examining clinical primary endpoints reported positive results, whereas 30 of 47 (63.8%) studies examining surrogate markers reported positive results (table 3). Among the 10 superiority studies using placebo comparators, 62.5% (five of eight) of clinical endpoint studies and 50.0% (one of two) of surrogate marker studies reported positive results.

We aggregated studies by indication and found that the median number of postapproval studies for each indication was 1 (interquartile range 0-2) (see supplementary appendix figure 1a) and the median number of studies that were randomized and double blind was 0 (0-1) (table 4). The median number of patients enrolled was 90 (0-509), and the median number of patients in the intervention arm was 52 (0-250) (see supplementary appendix figure 2a). The median number of patient years was 30.0 (0-352.4), and the median number of intervention patient years was 15.0 (0-212.9) (see supplementary appendix figure 3a). No postapproval studies were identified for 15 of 33 (45.5%) approved indications. Seven indications (21.2%) had one or more

Table 1 | Characteristics of novel drugs approved by the FDA between 2005 and 2012. Values are numbers (percentages) unless stated otherwise

Characteristics	Single pivotal trial	Surrogate markers	Both	P value
Total indications	33 (100.0)	49 (100.0)	41 (100.0)	
Agent type:				
Pharmacologic	23 (69.7)	45 (91.8)	31 (75.6)	0.02
Biologic	10 (30.3)	4 (8.2)	10 (24.4)	
Orphan status:				
Yes	6 (18.2)	10 (20.4)	10 (24.4)	0.85
No	27 (81.8)	39 (79.6)	31 (75.6)	
Approval pathway:				
Accelerated	0 (0.0)	12 (24.5)	9 (22.0)	0.002
Regular	33 (100.0)	37 (75.5)	32 (78.0)	
Therapeutic area:				
Cancer	9 (27.3)	8 (16.3)	24 (58.5)	0.004
Cardiovascular disease, diabetes mellitus	5 (15.2)	13 (26.5)	2 (4.9)	
Infectious disease	3 (9.1)	11 (22.4)	3 (7.3)	
Other	16 (48.5)	17 (34.7)	12 (29.3)	
Expected length of treatment:				
Acute	4 (12.1)	6 (12.2)	4 (9.8)	0.002
Intermediate	12 (36.4)	9 (18.4)	24 (58.5)	
Chronic	17 (51.5)	34 (69.4)	13 (31.7)	
Year of approval:				
2005-07	11 (33.3)	19 (38.8)	14 (34.1)	0.93
2008-10	10 (30.3)	15 (30.6)	11 (26.8)	
2011-12	12 (36.4)	15 (30.6)	16 (39.0)	

Table 2 | Characteristics of prospective controlled postapproval studies of novel drugs of the same indication for which the drug was first approved by the FDA between 2005 and 2012 on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease as primary endpoints, or both. Values are numbers (percentages) unless stated otherwise

Characteristics	Single pivotal trial	Surrogate markers	Both	P value
Total studies	77 (100.0)	554 (100.0)	127 (100.0)	
Randomized	63 (81.8)	493 (89.0)	112 (88.2)	0.18
Allocation:				
Double blind	18 (23.4)	223 (40.3)	22 (17.3)	
Single blind	24 (31.2)	44 (7.9)	13 (10.2)	<0.001
Open label	35 (45.5)	287 (51.8)	92 (72.4)	
Primary endpoint:				
Clinical outcome or scale	23 (29.9)	14 (2.5)	9 (7.1)	
Surrogate marker	51 (66.2)	512 (92.4)	110 (86.6)	<0.001
No primary efficacy endpoint	3 (3.9)	28 (5.1)	8 (6.3)	
Comparator:				
Active	67 (87.0)	365 (65.9)	100 (78.7)	
Placebo	10 (13.0)	148 (26.7)	24 (18.9)	<0.001
Active and placebo	0 (0.0)	41 (7.4)	3 (2.4)	
Trial design:				
Superiority	73 (94.8)	492 (88.8)	99 (78.0)	<0.001
Non-inferiority	4 (5.2)	62 (11.2)	28 (22.0)	
Intention to treat population*:				
Entire trial	96 (56-260)	127 (49-429)	207 (63-453)	0.01
Intervention group	50 (30-119)	68 (26-201)	111 (36-230)	0.01
Study duration:				
Duration in weeks*	4 (0.5-26)	24 (8-48)	26 (12-52)	<0.001
Duration ≥24 weeks	21 (27.3)	288 (52.0)	83 (65.4)	<0.001
Funder type:				
Industry or mixed funding that includes industry	26 (33.8)	337 (60.8)	98 (77.2)	<0.001
All others (non-profit, government, mixed non-industry, none, not specified)	51 (66.2)	217 (39.2)	29 (22.8)	

*Values are medians (interquartile ranges).

postapproval, randomized controlled, double blind positive study showing superior efficacy, six of which (18.2%) used a clinical outcome for the primary endpoint.

Indications approved on the basis of surrogate markers

We performed systematic reviews of the literature for all 49 indications for the 48 drugs that were approved by the FDA on the basis of multiple pivotal trials evaluating surrogate markers of disease. We identified a total of 554 published postapproval controlled studies using an active or placebo control and examining efficacy for the same therapeutic indication for which the drug was first approved (table 2). Of these, 89.0% (n=493) were randomized, 51.8% (n=287) used open label allocation, 92.4% (n=512) used primary efficacy endpoints that were surrogate markers of disease, 65.9% (n=365) used an active drug as control, 88.8% (n=492) used a superiority design, and 60.8% (n=337) were funded by industry. The median overall intention to treat population was 127 (interquartile range 49-429), and median study duration was 24 weeks (8-48). Among the 325 superiority studies using active comparators, two of eight (25.0%) studies examining clinical efficacy endpoints reported positive results, whereas 143 of 317 (45.1%) of those examining surrogate markers reported positive results (table 3). Among the 139 superiority studies using placebo comparators, 50.0% (two of four) of

clinical endpoint studies and 80.0% (108 of 135) of surrogate marker studies reported positive results.

We aggregated studies by indication and found that the median number of postapproval studies for each indication was 3 (interquartile range 1-8) (see supplementary appendix figure 1b), and the median number of randomized and double blind studies was 1 (0-3.5) (table 4). The median number of patients enrolled was 533 (122-3633), and the median number of patients in the intervention arm was 352 (104-2080) (see supplementary appendix figure 2b). The median number of patient years was 448.5 (23.1-2952.0), and the median number of intervention patient years was 350.0 (12.2-1412.8) (see supplementary appendix figure 3b). No postapproval studies were identified for eight of 49 (16.3%) approved indications. Twenty eight indications (57.1%) had one or more postapproval, randomized controlled, double blind positive study showing superior efficacy, one of which (2.0%) used a clinical outcome for the primary endpoint.

Indications approved on the basis of a single pivotal trial evaluating surrogate markers

We performed systematic reviews of the literature for all 41 indications for the 40 drugs that were approved by the FDA on the basis of a single pivotal trial evaluating surrogate markers of disease. We identified a total of 127 published postapproval controlled studies using an active or placebo control and examining efficacy for the

Table 3 | Findings from prospective controlled postapproval studies with primary efficacy endpoints of novel drugs of the same indication for which the drug was first approved by the FDA between 2005 and 2012 on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease as primary endpoints, or both. Values are numbers (percentages)

Study findings	Single pivotal trial	Surrogate markers	Both
Superiority studies	70	464	91
Active comparators*:			
Clinical primary endpoint (n=25):	13 (100)	8 (100)	4 (100)
Positive	1 (7.7)	2 (25.0)	1 (25.0)
Neutral	9 (69.2)	6 (75.0)	2 (50.0)
Negative	3 (23.1)	0 (0.0)	1 (25.0)
Surrogate marker of disease (n=429):	47 (100)	317 (100)	65 (100)
Positive	30 (63.8)	143 (45.1)	27 (41.5)
Neutral	8 (17.0)	129 (40.7)	29 (44.6)
Negative	9 (19.1)	45 (14.2)	9 (13.8)
Placebo comparator:			
Clinical primary endpoint:	8 (100)	4 (100)	3 (100)
Positive	5 (62.5)	2 (50.0)	2 (66.7)
Neutral	2 (25.0)	2 (50.0)	1 (33.3)
Negative	1 (12.5)	0 (0.0)	0 (0.0)
Surrogate marker of disease:	2 (100)	135 (100)	19 (100)
Positive	1 (50.0)	108 (80.0)	13 (68.4)
Neutral	1 (50.0)	27 (20.0)	5 (26.3)
Negative	0 (0.0)	0 (0.0)	1 (5.3)
Non-inferiority studies			
Clinical primary endpoint:	4	62	28†
Positive	2 (100)	2 (100)	2 (100)
Negative	2 (100.0)	2 (100.0)	1 (50.0)
Surrogate marker of disease:	2 (100)	60 (100)	25 (100)
Positive	2 (100.0)	59 (98.3)	23 (92.0)

*Includes studies with both active and placebo comparators, recorded with respect to active comparator.

†One trial was unable to demonstrate non-inferiority or inferiority and is not included.

same therapeutic indication for which the drug was first approved (table 2). Of these, 88.2% (n=112) were randomized, 72.4% (n=92) used open label allocation, 86.6% (n=110) used primary efficacy endpoints that were surrogate markers of disease, 78.7% (n=100) used an active drug as control, 78.0% (n=99) used a superiority design, and 77.2% (n=98) were funded by industry. The median overall intention to treat population was 207 (interquartile range 63-453), and the median study duration was 26 weeks (12-52). Among the 69 superiority studies using active comparators, one of four (25.0%) studies examining clinical efficacy endpoints reported positive results, and 27 of 65 (41.5%) of those examining surrogate markers reported positive results (table 3). Among the 22 superiority studies using placebo comparators, 66.7% (two of three) of clinical endpoint studies and 68.4% (13 of 19) of surrogate marker studies reported positive results.

We aggregated studies by indication and found that the median number of postapproval studies for each indication was 1 (interquartile range 0-2) (see supplementary appendix figure 1c), and the median number of randomized and double blind studies was 0 (0-0) (table 4). The median number of patients enrolled was 38 (0-666), and the median number of patients enrolled in the intervention arm was 19 (0-378) (see supplementary appendix figure 2c). The median number of patient years was 4.7 (0-998.7), and the median number of intervention patient years was 3.8 (0-693.9) (see

supplementary appendix figure 3c). No postapproval studies were identified for 20 of 41 (48.8%) approved indications. Six indications (14.6%) had one or more postapproval, randomized controlled, double blind positive study showing superior efficacy, two of which (4.9%) used a clinical outcome for the primary endpoint.

Subgroup analyses

Medians and rates were largely consistent when examined by year of indication approval, by therapeutic area, or by expected length of treatment, with the exception of medians by year of indication of approval among approvals based on a single pivotal trial using surrogate markers (see supplementary appendix table 1) and medians by therapeutic area among approvals based on multiple trials using surrogate markers (see supplementary appendix table 2); no consistent differences were found between expected lengths of treatment (see supplementary appendix table 3).

Discussion

We found substantial variation in the quantity and quality of studies of novel drugs published after they were approved by the FDA on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease, or both. We found few published randomized controlled, double blind studies showing superior efficacy based on clinical outcomes that examined the same indication for which the drug was first approved by the FDA after a median follow-up of 5.5 years. These findings have important implications for clinical care. Both doctors and patients have high expectations for the safety and efficacy of a drug approved by the FDA.¹⁹⁻²³ But less than one third of new drug indications approved by the FDA on the basis of a single pivotal trial had at least one postapproval trial showing superior efficacy; even fewer used clinical outcomes. Similarly, less than one 10th of new drug indications approved by the FDA on the basis of surrogate markers of disease had at least one postapproval trial validating the use of the surrogate marker by showing superior efficacy using clinical outcomes. Our work corresponds with a previous study demonstrating that highly cited studies are infrequently followed by subsequent clinical studies that confirm original effects²⁴ and is consistent with a study showing that cancer drugs approved on the basis of surrogate markers are infrequently followed by clinical studies demonstrating improved overall survival.¹²

As well as finding few robustly designed, confirmatory studies, we also found noticeable variability in the degree to which novel drugs were studied in the postapproval period, both within and between the categories of approval. Although a small number of approved indications had dozens of qualifying postapproval studies, more than one third had no published postapproval controlled studies investigating efficacy. Furthermore, across approval categories, the studies showed statistically significant differences in all features of trial design, except randomization, as well as in aggregated median numbers of studies and in patient enrollment.

Table 4 | Aggregated number and patients enrolled in prospective controlled postapproval studies of novel drugs of the same indication for which the drug was first approved by the FDA between 2005 and 2012 on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease as primary endpoints, or both. Values are medians (interquartile ranges) unless stated otherwise

Characteristics	Single pivotal trial (n=33)	Surrogate markers (n=49)	Both (n=41)	P value
Studies per indication	1 (0-2)	3 (1-8)	1 (0-2)	<0.001
Randomized and double blind studies	0 (0-1)	1 (0-3.5)	0 (0-0)	<0.001
Total patients enrolled	90 (0-509)	533 (122-3633)	38 (0-666)	<0.001
Intervention patients enrolled	52 (0-250)	352 (104-2080)	19 (0-378)	<0.001
Total patient years of exposure	30.0 (0-352.4)	448.5 (23.1-2952.0)	4.7 (0-998.7)	0.003
Intervention patient years of exposure	15.0 (0-212.9)	350.0 (12.2-1412.8)	3.8 (0-693.9)	0.003
Indications with (No (%)):				
≥1 study	18 (54.5)	41 (83.7)	21 (51.2)	0.001
≥1 randomized and double blind study	10 (30.3)	31 (63.3)	9 (22.0)	<0.001
≥1 randomized and double blind study with clinical primary endpoint	10 (30.3)	3 (6.1)	3 (7.3)	0.005
≥1 study with positive results	9 (27.3)	37 (75.5)	15 (36.6)	<0.001
≥1 randomized and double blind study with positive results	7 (21.2)	28 (57.1)	6 (14.6)	<0.001
≥1 randomized and double blind study with clinical primary endpoint and positive results	6 (18.2)	1 (2.0)	2 (4.9)	0.02

For many novel drugs, the problem is not that postapproval studies are poorly designed or have negative efficacy results, but rather that they are not being published or performed at all. Our findings are consistent with previous studies showing that even postapproval studies required by the FDA to be conducted by manufacturers are not always performed in a timely manner¹³⁻¹⁵ and with previous work demonstrating the variability in the number and completion of postmarketing studies of high risk medical devices.²⁵

The majority of the postapproval studies that we identified focused on surrogate markers of disease. Using surrogate markers instead of clinical outcomes for trial endpoints has become increasingly controversial,²⁶ with several approved drugs ultimately failing to confirm any clinical benefit²⁷⁻²⁹ and an analysis of oncology surrogate markers finding variable correlation with overall survival.³⁰ Surrogate markers have been used for pivotal clinical studies to facilitate more rapid regulatory evaluation, with the expectation that subsequent clinical studies will focus on clinical outcomes. We found that this occurs infrequently, as approximately 90% of postapproval studies of drugs for indications approved on the basis of surrogate markers also used surrogate markers of disease for trial endpoints.

Recent proposals for FDA reform include creating a comprehensive benefit and risk assessment and management plan to be updated at regular intervals and with any shift in a drug's benefit to risk profile⁷; increasing reliance on surrogate markers, smaller and shorter trials, and evidence derived from registries and observational studies³¹; and switching to a "consumer reports" approach of grades for efficacy, safety, and degree of evidence.³² Our findings show that caution would be needed for these approaches—high quality postapproval evidence does not necessarily accumulate and may require additional regulatory requirements. To strengthen lifecycle evaluation of recently approved drugs, requirements for postapproval studies might need to be heightened or might need to specify study design characteristics or trial endpoints in detail to ensure that these studies provide high quality evidence to further inform clinical practice. Of course, it must be

remembered that publishing one short term, small postapproval study might be inadequate for a hypertension drug but entirely reasonable for one treating a rare disease. A customized, specific plan for postapproval studies will be key to generating an adequate amount of postapproval evidence on comparative efficacy, which should take into consideration disease prevalence and severity, the drug's anticipated length of use, and the availability of other treatment options.

Limitations of this study

Our study has several limitations. Firstly, we restricted our analysis to studies that had been published in the biomedical literature and indexed in Medline and did not search other databases, such as Embase, Scopus, research conference proceedings (abstracts), or ClinicalTrials.gov. Because our study comprised 123 unique searches to determine whether new drugs approved on the basis of limited evidence continued to be studied in the postmarketing period, we considered this to be a common sense strategy to identify all clinically relevant trial papers. Moreover, our Medline search for publications was extensive, using a systematic query method to capture all clinical trial results that were most likely to inform and influence clinical practice. We might have missed studies that had been conducted but not indexed in Medline, but without widespread dissemination any such studies are unlikely to have informed clinical practice. During project planning, we performed a non-systematic pilot search of ClinicalTrials.gov and found that nearly every entry either did not report study results or was already published in a journal indexed on Medline and had been identified through our systematic search, leading us to conclude that searching this database for additional trials would not be worth the considerable effort. Moreover, we are likely to have identified all controlled studies that confirmed or validated the efficacy of novel drugs given the ample literature suggesting that positive, completed studies are most likely to be published.^{33 34} To ensure that all relevant studies were identified, we performed a robustness check of our search strategy, searching both Scopus and ClinicalTrials.gov for all indications for

which we had found no postapproval studies (47 of 123). We found no additional studies that had been published in a journal not indexed by Medline, four unpublished studies reporting results with statistical analyses affecting three indications on ClinicalTrials.gov, and five additional studies affecting four indications that had been published in a journal indexed on Medline that were not identified on our original search. We included these published studies in our results; only three offered evidence on drug efficacy, two of which showed superiority of the drug and one found inferiority. Thus, we do not expect our results or interpretation to be affected by adopting a wider search strategy that includes Scopus and ClinicalTrials.gov.

Secondly, we limited our analysis to prospective studies that examined efficacy for the first approved indication and used a comparator as a control, as we think that this generates the most informative, high quality data to understand drug efficacy. However, non-comparative or retrospective analyses might also have been published and provide important insights, as might studies for other indications, although our classification of indication was broad to include as many publications as could possibly be relevant. Thirdly, we included only English language publications. Fourthly, we completed our systematic reviews on 31 December 2014, and relevant clinical studies might have been published since then. However, the median follow-up period after approval exceeded five years, and the value of evidence to inform and direct practice diminishes as time passes. Finally, we only examined novel drugs approved on the basis of limited evidence. Although these approvals accounted for more than half of the FDA's approvals from 2005 to 2012, our findings may not be applicable to drugs approved on the basis of multiple pivotal trials or trials focused on clinical outcomes.

Conclusion

The quantity and quality of postapproval clinical evidence varied substantially for novel drugs approved by the FDA on the basis of a single pivotal trial, pivotal trials that used surrogate markers of disease, or both. Fewer than 10% of approved indications had one or more published randomized controlled, double blind study showing superior efficacy based on clinical outcomes that examined the same indication for which the drug was first approved by the FDA after a median of 5.5 years after approval. These findings should inform both clinical decision making and regulatory policy regarding requirements before and after approval of novel drugs.

AUTHOR AFFILIATIONS

¹State University of New York Downstate College of Medicine, Brooklyn, NY, USA

²Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

³Robert Wood Johnson Foundation Clinical Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

⁴Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

⁵Center for Outcomes Research and Evaluation, Yale-New Haven Health, New Haven, CT, USA

⁶Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁷Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

⁸Division of Health Care Policy and Research and Kern Center for the Science of Health Care Delivery, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Contributors: AMP and JSR conceived and designed this study, conducted the statistical analysis, and drafted the manuscript. AMP and JAA acquired the data. JSR provided supervision. All authors participated in the analysis and interpretation of the data and critically revised the manuscript for important intellectual content. AMP and JSR had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. AMP and JSR are guarantors.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: HMK and JSR receive support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Centers of Medicare and Medicaid Services to develop and maintain performance measures that are used for public reporting, from Medtronic and the FDA to develop methods for postmarket surveillance of medical devices, from the Blue Cross Blue Shield Association to better understand medical technology evaluation, and from the Laura and John Arnold Foundation to support the Collaboration on Research Integrity and Transparency at Yale. HMK reports that he chairs a scientific advisory board for UnitedHealthcare. NDS and JSR receive support from the FDA to establish the Yale-Mayo Clinical Center for Excellence in Regulatory Science and Innovation. AMP, NSD, and JAA declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Funding: This project was supported in part by the Robert E Leet and Clara Guthrie Patterson Trust Awards Program in Clinical Research, Bank of America, NA, Trustee. The sponsor had no role in the design of the study, analysis or interpretation of findings, or drafting the manuscript and did not review or approve the manuscript before submission. The authors assume full responsibility for the accuracy and completeness of the ideas presented. AMP was supported in part by the SUNY Downstate College of Medicine Alumni Fund Summer Research Fellowship. Ethical approval: The Yale University Institutional Review Board approved this study and considered it non-human subjects research.

Transparency: AMP and JSR, on behalf of the authors, affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing: Requests for collected data and statistical code can be made to the corresponding author at joseph.ross@yale.edu.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 US Food and Drug Administration (FDA). About FDA product approval. <https://www.fda.gov/NewsEvents/ProductsApprovals/ucm106288.htm>.
- 2 US Food and Drug Administration (FDA). Guidance for industry: providing clinical evidence of effectiveness for human drugs and biological products. <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm078749.pdf>.
- 3 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. doi:10.1001/jama.2013.282034.
- 4 Institute of Medicine. *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. The National Academies Press, 2007.

- 5 US Food And Drug Administration (FDA). Postmarket drug and biologic safety evaluations. <https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/ucm204091.htm>.
- 6 Psaty BM, Meslin EM, Breckenridge A. A lifecycle approach to the evaluation of FDA approval methods and regulatory actions: opportunities provided by a new IOM report. *JAMA* 2012;307:2491-2. doi:10.1001/jama.2012.5545.
- 7 Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*. Washington (DC): National Academies Press, 2012.
- 8 Shuren J, Califf RM. Need for a national evaluation system for health technology. *JAMA* 2016;316:1153-4. doi:10.1001/jama.2016.8708.
- 9 US Food and Drug Administration (FDA). Statement from the Center for Drug Evaluation and Research on the IOM report: ethical and scientific issues in studying the safety of approved drugs, released May 1, 2012. <http://www.fda.gov/Drugs/DrugSafety/ucm302485.htm>.
- 10 US Food and Drug Administration (FDA). FDA's Sentinel Initiative—background. <https://www.fda.gov/Safety/FDAsSentinelInitiative/ucm149340.htm>.
- 11 US Food and Drug Administration (FDA). Guidance for industry: postmarketing studies and clinical trials. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf>.
- 12 Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of us food and drug administration approvals. *JAMA Intern Med* 2015;175:1992-4. doi:10.1001/jamainternmed.2015.5868.
- 13 US Government Accountability Office. FDA needs to enhance its oversight of drugs approved on the basis of surrogate endpoints. <http://www.gao.gov/products/GAO-09-866>.
- 14 Bognet RA, Chiang DS, Katz KA. Dermatology-related postmarketing study commitments to the Food and Drug Administration. *J Am Acad Dermatol* 2006;55:721-3. doi:10.1016/j.jaad.2006.06.025.
- 15 Fain K, Daubresse M, Alexander GC. The food and drug administration amendments act and postmarketing commitments. *JAMA* 2013;310:202-4. doi:10.1001/jama.2013.7900.
- 16 World Health Organization (WHO) Collaborating Center for Drug Statistics Methodology. ATC/DDD index 2017. https://www.whocc.no/atc_ddd_index/.
- 17 Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 2009;302:1437-43. doi:10.1001/jama.2009.1409.
- 18 Institute of Medicine. *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. National Academies Press, 2010.
- 19 Schwartz LM, Woloshin S. Communicating uncertainties about prescription drugs to the public: a national randomized trial. *Arch Intern Med* 2011;171:1463-8. doi:10.1001/archinternmed.2011.396.
- 20 Weeks JC, Catalano PJ, Cronin A, et al. Patients' expectations about effects of chemotherapy for advanced cancer. *N Engl J Med* 2012;367:1616-25. doi:10.1056/NEJMoa1204410.
- 21 Anderson GM, Juurlink D, Detsky AS. Newly approved does not always mean new and improved. *JAMA* 2008;299:1598-600. doi:10.1001/jama.299.13.1598.
- 22 Chen DT, Wynia MK, Moloney RM, Alexander GC. US physician knowledge of the FDA-approved indications and evidence base for commonly prescribed drugs: results of a national survey. *Pharmacoepidemiol Drug Saf* 2009;18:1094-100. doi:10.1002/pds.1825.
- 23 Healy D. *Pharmageddon*. University of California Press, 2012.
- 24 Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218-28. doi:10.1001/jama.294.2.218.
- 25 Rathi VK, Krumholz HM, Masoudi FA, Ross JS. Characteristics of clinical studies conducted over the total product life cycle of high-risk therapeutic medical devices receiving fda premarket approval in 2010 and 2011. *JAMA* 2015;314:604-12. doi:10.1001/jama.2015.8761.
- 26 Ciani O, Buysse M, Drummond M, et al Time to review the role of surrogate end points in health policy: state of the art and the way forward. *Value Health* 2017;20:487-95.
- 27 Carpenter D, Kesselheim AS, Joffe S. Reputation and precedent in the bevacizumab decision. *N Engl J Med* 2011;365:e3. doi:10.1056/NEJMp1107201.
- 28 Krumholz HM. Biomarkers, risk factors, and risk: clarifying the controversy about surrogate end points and clinical outcomes. *Circ Cardiovasc Qual Outcomes* 2015;8:457-9. doi:10.1161/CIRCOUTCOMES.115.002245.
- 29 Krumholz HM. Niacin: Time to Believe Outcomes Over Surrogate Outcomes: If Not Now, When? *Circ Cardiovasc Qual Outcomes* 2016;9:343-4. doi:10.1161/CIRCOUTCOMES.116.003094.
- 30 Prasad V, Kim C, Burotto M, Vandross A. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. *JAMA Intern Med* 2015;175:1389-98. doi:10.1001/jamainternmed.2015.2829.
- 31 Avorn J, Kesselheim AS. The 21st century cures act—will it take us back in time? *N Engl J Med* 2015;372:2473-5. doi:10.1056/NEJMp1506964.
- 32 Sotos J. It's time to radically change how the FDA approves drugs. *Wall Street Journal* 2016 June 19. <http://blogs.wsj.com/experts/2016/06/29/its-time-to-radically-change-how-the-fda-approves-drugs/>
- 33 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60. doi:10.1056/NEJMsa065779.
- 34 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454-65. doi:10.1001/jama.289.4.454.

Appendix: supplementary material