Selective reporting in trials of high risk cardiovascular devices: cross sectional comparison between premarket approval summaries and published reports

Lee Chang,¹ Sanket S Dhruva,² Janet Chu,³ Lisa A Bero,⁴ Rita F Redberg⁵

¹Department of Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA

²Division of Cardiovascular Medicine, University of California, Davis, Sacramento, CA 95817, USA

³University of California, San Francisco School of Medicine, San Francisco, CA 94143, USA

⁴Charles Perkins Centre, University of Sydney, Sydney, Australia

⁵Division of Cardiology, Suite M-1180, 505 Parnassus Avenue, University of California-San Francisco, San Francisco, CA 94143, USA

Correspondence to R F Redberg
redberg@medicine.ucsf.edu

Cite this as: *BMJ* 2015;350:h2613
doi: 10.1136/bmj.h2613

Accepted: 20 April 2015

ABSTRACT

OBJECTIVE

To investigate characteristics of clinical trials and results on safety and effectiveness reported in US Food and Drug Administration (FDA) documents for recently approved high risk cardiovascular devices compared with the characteristics and results reported in peer reviewed publications.

DESIGN

A search of the publicly available FDA database was performed for all cardiovascular devices that received premarket approval from 1 January 2000 to 31 December 2010. For each study listed in the premarket approval documents, a Medline search was conducted to obtain the corresponding publication.

MAIN OUTCOME MEASURES

Clinical trial characteristics, primary endpoints, and safety and efficacy results in the FDA documents and corresponding publications.

RESULTS

106 cardiovascular devices received premarket approval from 1 January 2000 to 31 December 2010. FDA premarket approval documents for these devices contained 177 studies, of which 86 (49%) had been published by 1 January 2013. These 86 publications corresponded to 60 distinct devices. The mean time from FDA approval to publication in a peer reviewed journal was 6.5 months (range -4.8-7.5 years). In 22 (26%) of the 86 compared studies the number of participants enrolled in the study differed in the FDA summary and the corresponding publications. Of 152 primary endpoints identified in the FDA documents, in the corresponding publications three (2%) were labeled as secondary, 43 (28%) were unlabeled, and 15 (10%) were not found. Among the primary results, 69 (45%) were identical, 35 (23%) were similar, 17 (11%) were substantially different, and 31 (20%) could not be compared.

CONCLUSIONS

Many clinical trials for high risk cardiovascular devices approved by the FDA remain unpublished. Even when trials are published, the study population, primary endpoints, and results can differ substantially from data submitted to the FDA.

Introduction

As medical devices play an important role in health-care, the quality of clinical trial evidence supporting their use is critical to patients' health and safety. In the United States between 1990 and 2002, over 17 000 pacemakers and implantable cardioverter defibrillators were removed and over 60 deaths were attributed to confirmed device malfunctions.¹ Manufacturers of new high risk devices, including many used to treat cardiovascular disease, are often required to submit clinical trial data to the US Food and Drug Administration (FDA) for evaluation of safety and effectiveness via the premarket approval process.² A review of recently approved high risk cardiovascular devices showed that they are often approved on the basis of a single study, most of which are not randomized and lack prospective controls.^{3,4}

FDA documents containing the evidence to support approval of devices are available but are not easy to access. For informed clinical decision making, the trial data that the FDA reviews should be accessible to clinicians in peer reviewed publications. Previous work has found that when clinical trial information in FDA new drug applications is compared with the corresponding publications, there was significant selective reporting. Results favoring the new drug over the comparator were significantly more likely to be reported in the literature than unfavorable results.⁵ The selective publication of favorable results for drugs includes several types of reporting bias, including failure to publish entire studies, failure to publish unfavorable outcomes, and publication of only selected prespecified outcome analyses.

The extent of selective reporting for medical devices has not been examined. To assess selective reporting for trials of high risk cardiovascular devices, we examined clinical trial data found in FDA summaries compared with corresponding peer reviewed publications. We compared characteristics of clinical trials and results on safety and effectiveness reported in FDA premarket approval documents for recently approved high risk cardiovascular devices (such as coronary stents and implantable cardioverter defibrillators) with the trial characteristics and results reported in journal publications.

WHAT IS ALREADY KNOWN ON THIS TOPIC

There is reporting bias of clinical data in new drug applications when they are published

The extent of selective reporting for medical device clinical trials is unknown

WHAT THIS STUDY ADDS

Most clinical studies used to justify FDA approval of high risk cardiovascular devices are not published

When these studies are published, there are often clinically relevant discrepancies between FDA documents and corresponding publications, including changed primary endpoints

Methods

Trials in FDA summaries of safety and effectiveness data for premarket approval process

We searched the publicly available FDA premarket approval database (www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm) on 1 August 2012 for all devices meeting the following parameters: “Cardiovascular” under advisory committee and “Originals Only” under supplement type, with an FDA approval date between from 1 January 2000 to 31 December 2010. The end date was selected to allow at least two years from approval of the device to publication of the clinical trial.⁶ For each device’s premarket approval located using this approach, we examined the “summary of clinical studies” section to identify clinical studies. For all the identified studies, we abstracted study characteristics and data.

Published trials

For each FDA summary, we searched Medline on 15 January 2013 for a corresponding publication with a publication date from 1 January 1990 to 1 January 2013. As the summaries do not include study authors or names of principal investigators, we conducted the search primarily using clinical trial titles and product names. Publication matches were confirmed by comparing methods, number of study centers, enrollment number, primary endpoint, primary results, and study sponsor. There were no instances in which there was ambiguity between multiple publications.

If this process failed to identify a matching publication, we emailed the device manufacturer to inquire about publication status. If the manufacturer was unable to provide a reference for publication or did not respond after three separate contact attempts one month apart, we considered the clinical trial unpublished.

Data coding and comparing summaries with publications

Data were abstracted from FDA summaries and publications and compiled in a database to compare the study characteristics and results from each respective source. One author (SD) abstracted and coded the summary, which was verified by another author (LC). Two authors (LC and JC) coded the data from publications. Disagreements were resolved by consensus of all authors.

A study was classified as pivotal if it was the only one included in the summary, a multicenter randomized controlled trial, or explicitly noted in the summary as being a pivotal study. If there were multiple multicenter randomized controlled trials for a given device, we considered all of them as pivotal studies. Otherwise, the study was classified as feasibility, early, or supportive in accordance with what was stated in the summary.

For each publication, we recorded time to publication (defined as the earliest date of publication, either online or in print), journal of publication, source of funding (industry or public) disclosed in the publication, and any author disclosure of conflict of interest.

After abstraction of data from both sources, we compared the following study characteristics between the summary and publications: randomization (yes/no), blinding (single, double, or none as reported in the summary or publications), number of sites, number of patients enrolled, patient demographics (age, sex, race), and whether results from multiple studies were pooled into one publication.

Primary endpoint analysis

We identified all explicitly stated primary endpoints in FDA summaries and publications for each study. When not specified in the summary or publication, all endpoints were designated as primary if there were three or fewer endpoints. If there were more than three endpoints and none was designated as primary, the study was considered as having no primary endpoint.² All summary primary endpoints were compared with matching endpoints in the corresponding publication (if one could be found), even if they were not designated as primary in the publication.

Primary endpoint features compared between summaries and publications were as follows: type of primary endpoint (equivalence, non-inferiority, objective performance criteria), type of controls, and number of patients analyzed for primary endpoint. An objective performance criterion is a surrogate benchmark for an outcome that serves as a control group for some studies.

We compared the results for each primary endpoint between the summaries and publications and classified them as identical, similar, different, or unknown. To be classified as identical, the results had to match in both numerical value and significance (same P value and confidence interval). Results were designated as similar if the publication value differed by less than 5% from the original summary values and significance was unchanged. Otherwise, the results were considered as different. They were classified as unknown if no corresponding primary endpoint result could be found in the publication.

Data analysis

We first compared characteristics of published and unpublished studies. We then further examined the characteristics of the publications. Finally, we examined discrepancies between study characteristics, primary endpoints, and results as reported in the summaries and publications. Summary statistics were calculated for each comparison outlined above and presented as numbers, percentages, means, standard deviations, and ranges when applicable. Significance was assessed with a generalized linear mixed effects model with a canonical logit link function for a binomial model and a random effect placed on the intercept to account for heterogeneity across devices. We then computed odds ratios, P values, and 95% confidence intervals. For the primary endpoint comparison of number analyzed, we used a general linear mixed effects model with random effect placed on the intercept to account for device clustering. All tests were two sided with $P < 0.05$ as the criterion for significance. We performed a subgroup analysis on all pivotal studies.

Table 1 | Characteristics of published trials on high risk cardiovascular devices. Figures are numbers (percentage) unless stated otherwise

	All studies (n=86)	Pivotal studies* (n=66)
Mean (range) time to publication (months) from FDA premarket approval	6.5 (–58–90)	7.9 (–58–90)
Industry funded	62 (71)	50 (76)
Source of funding not specified	24 (28)	16 (24)
Author conflicts disclosed	56 (65)	43 (65)
Multiple trials presented in pooled fashion	6 (7)	5 (8)

*Study was classified as pivotal if it was only one included in summary, multicenter randomized controlled trial, or explicitly noted in summary as being pivotal study.

Results

From 2000 to 2010 there were summaries for 106 cardiovascular devices. We identified 177 studies (mean 1.7 per device) in these summaries, of which 86 (49%) were published. The 86 published studies corresponded to 60 distinct devices (mean 1.4 published studies per device). The pivotal studies also corresponded to the same 60 distinct devices (mean 1.1 published pivotal studies per device). We contacted 23 manufacturers to request publication references, and eight (35%) responded, confirming that the trials of interest had not been published. One manufacturer could not be contacted because of lack of any contact information. Subgroup analysis restricted to the pivotal studies showed that of 112 pivotal studies, there were 66 corresponding publications (59%).

Published versus unpublished trials

Published trials were more often randomized, blinded, and based exclusively in the US than unpublished trials. Of the 86 published trials, 34 (40%) were randomized, compared with 26 (29%) of the 91 unpublished trials (95% confidence interval for difference 3.7% to 25.1%). Similarly, 23/86 (27%) of published trials were blinded compared with 6/91 (7%) of unpublished trials (8.5% to 31.6%). Although nearly the same proportion of published and unpublished trials were conducted at multiple centers—85% and 87%, respectively—more published trials than unpublished trial (43% v 23%) were conducted exclusively in the US (5.3% to 33.6%).

Table 2 | Comparison of enrollment and demographics reported in FDA premarket approval summaries with corresponding publications and those published studies. Figures are numbers (percentage) unless stated otherwise

	All studies (n=86)	Pivotal studies* (n=66)
Discrepancy in enrollment	22 (26)	18 (27)
Discrepancy >5%	14 (16)	11 (17)
Discrepancy <5%	8 (9)	7 (11)
Mean (range) difference in No of patients enrolled (range)	–1.4 (–104–181)	–1.1 (–104–181)
Mean (range) % difference in No of patients enrolled	–1.2 (–39–39)	–0.7 (–39–39)
Absolute difference in No of patients enrolled (range)	12.8 (0–181)	16.7 (0–181)
Absolute mean difference in % enrolled (range)	4 (0–39)	4 (0–39)
Discrepancy in mean age	9 (11)	8 (12)
Discrepancy in % male	14 (16)	12 (18)

*Study was classified as pivotal if it was only one included in summary, multicenter randomized controlled trial, or explicitly noted in summary as being pivotal study.

Overall, randomized and blinded studies were more likely to be published. Of the randomized studies, 35/58 (60%) were published compared with 51/119 (43%) of non-randomized studies (95% confidence interval for difference 0.9% to 32.7%). Among blinded studies, 79% were published compared with 43% of non-blinded studies (15.5% to 51.2%).

Characteristics of published trials

The average time from FDA approval to publication was 6.5 months, with a range of –4.8 years to 7.5 years (table 1). For the 66 pivotal studies, the average time to publication was 7.9 months, and 22 (33%) were published before FDA approval. Of publications that specified a funding source, all were industry funded. Most publication authors disclosed additional conflicts of interest. Six studies, five of them pivotal, were presented as pooled data and not individual studies when published.

Comparison of demographics between summaries and publications

In 22 (26%) of the 86 published studies, the number of patients enrolled in study differed between the summary and the publication (table 2). When we used the summary data as the reference, the mean difference in the stated total number of patients enrolled was 12.8 (4% difference), with a maximum difference of 181 patients. Sixteen publications reported fewer patients enrolled and five reported more patients enrolled than the corresponding summary.

Demographic information also differed between several FDA summaries and published studies. In nine (11%) of the 86 studies, the average age differed by more than one year, and in 14 (16%) the breakdown by sex differed by more than 1% in absolute terms (table 2).

Comparison of study characteristics between summaries and publications

Overall, 35 (41%) of the published studies reported a randomized design (table 3). Twenty three (27%) were blinded, 13 (57%) of which were double blinded. Characteristics of randomization and blinding were identical between the summaries and publications. Pivotal studies, compared with feasibility, early, and supportive studies, were similar in terms of likelihood of being randomized: 28/66 (42%) were randomized and 16/66 were blinded (24%), of which 10/16 (63%) were double blinded.

Comparison of primary endpoints characteristics between summaries and publications

For the 86 summary studies for which we found a corresponding publication, 139 total primary endpoints were characterized. Thirteen summary studies had no identifiable primary endpoint, but for purposes of analysis, as the summary data were considered the standard, these studies were considered to have one unknown primary endpoint and added to the 139 to comprise the denominator of 152 possible primary endpoints from the summaries. The primary endpoints were heterogeneous

Table 3 | Comparison of study characteristics reported in FDA premarket approval summaries with corresponding publications and those published studies. Figures are numbers (percentage) unless stated otherwise

	All (n=86)			Pivotal (n=66)*		
	Summary	Published	Odds ratio (95% CI), P value†	Summary	Published	Odds ratio (95% CI), P value†
Randomized	34 (40)	34 (40)	1.00 (0.43 to 2.33), 1.0	28 (42)	28 (42)	1.00 (0.38 to 2.62), 1.0
Blinded studies	23 (27)	23 (27)	1.00 (0.37 to 2.67), 1.0	16 (24)	16 (24)	1.00 (0.33 to 3.06), 1.0
Double blind	13 (15)	13 (15)	1.00 (0.35 to 2.85), 1.0	10 (15)	10 (15)	1.00 (0.32 to 3.11), 1.0
Single blind	10 (12)	10 (12)	1.00 (0.32 to 3.16), 1.0	6 (9)	6 (9)	1.00 (0.25 to 3.95), 1.0
Multicenter	73 (85)	76 (89)	1.45 (0.54 to 3.89), 0.46	64 (97)	65 (99)	—‡
Single center	8 (9)	7 (8)	—‡	1 (2)	1 (2)	—‡
No of centers not specified	5 (6)	3 (4)	0.58 (0.13 to 2.53), 0.47	1 (2)	0 (0)	—‡
Exclusively US sites	37 (43)	43 (50)	1.42 (0.73 to 2.77), 0.31	35 (53)	36 (55)	1.09 (0.49 to 2.42), 0.84
Location not specified	24 (28)	14 (16)	0.49 (0.23 to 1.04), 0.07	12 (18)	10 (15)	0.80 (0.31 to 2.03), 0.64

*Study was classified as pivotal if it was only one included in summary, multicenter randomized controlled trial, or explicitly noted in summary as being pivotal study.

†As calculated by generalized linear mixed effects model with canonical logit link and random effect placed on the intercept to account for heterogeneity across devices.

‡Too few events or counter-events for fitting of random effects model.

given the many classes of devices studied, but the most common included success of the device implant, functional improvement such as in a six minute walk test, and a composite of major cardiovascular events (including a combination of myocardial infarction, stroke, revascularization, and death). Thirty four (40%) studies differed in the number of primary endpoints between summaries and publications. Among the 66 pivotal studies in the summaries, 120 primary endpoints (mean 1.8 per pivotal study) were found, while none could be identified in seven (11%) studies.

When we tried to locate the primary endpoints identified in the summaries in the corresponding publications,

three (2%) were labeled as secondary endpoints, 43 (28%) were unlabeled, and 15 (10%) could not be found (table 4). Summary primary endpoints were compared with any matching endpoints in the corresponding publication even if they were not designated as primary. Of the summary primary endpoints, 110 (72%) were identical in definition and nine (6%) were similar to their corresponding publication endpoints. The findings were similar among the pivotal studies.

With regard to primary endpoint type, 47 (31%) in the summaries were measured against an objective performance criterion compared with 16 (11%) in the publications (odds ratio 0.15, 95% confidence interval 0.07 to 0.32) (table 5). Most primary endpoints in either the summary or publication did not specify the type: 77 (51%) versus 115 (76%), respectively (4.01, 2.30 to 6.99). Sixty three (41%) of the primary endpoints were compared with prospective control groups for both the summaries and publications. There were, however, 13 (9%) instances when the reported controls differed between summary and publication (table 4).

In 46 (30%) cases, the number of patients used for the primary endpoint analysis differed between the summaries and publications. Overall, however, the difference in number analyzed did not seem to be significant (table 5).

Comparison of primary endpoint results between summaries and publications

Analysis of the overall primary endpoint results showed that 69 (45%) were identical, 35 (23%) were similar, 17 (11%) were different, and 31 (20%) could not be compared (table 4). With regard to significance, 94 (62%) of the 152 summary primary results favored the device, eight (5%) favored the control (these eight included studies for which the objective performance criteria fell within the confidence interval or non-inferiority was not met), 12 (8%) favored neither device nor control, and 38 (25%) did not have a control or objective performance criteria by which to compare (table 5). In comparison, 65 (43%) of the published primary results significantly favored the device, three (2%) favored the control, 14 (9%) favored neither, and 70 (46%) could not be compared with a control or with

Table 4 | Findings in published reports of studies of high risk cardiovascular devices compared with FDA premarket approval summaries with corresponding publications. Figures are numbers (percentage) unless stated otherwise

	All studies (n=152)	Pivotal studies* (n=127)
Primary labeled secondary in publication	3 (2)	3 (2)
Primary not labeled in publication	43 (28)	36 (28)
Primary not found in publication	15 (10)	14 (11)
No primary in either SSED or publication	4 (3)	2 (2)
Endpoint definition:		
Identical	110 (72)	96 (76)
Similar	9 (6)	9 (7)
Unknown	33 (22)	22 (17)
Discrepancy in control type	13 (9)	13 (10)
No of patients analyzed:		
Mean difference (range)	8.9 (−81-272)	10.6 (−81-272)
Mean % difference (range)	4.3 (−48-131)	5.3 (−48-131)
Absolute mean difference (range)	15.9 (0-272)	17.8 (0-272)
Absolute % difference (range)	8.2 (0-131)	9 (0-131)
Discrepancy in No analyzed	46 (30)	43 (34)
Endpoint results:		
Content:		
Identical	69 (45)	55 (43)
Similar	35 (23)	33 (26)
Different	17 (11)	16 (13)
Unknown	31 (20)	23 (18)
Discrepancies in direction of favorability:		
Discrepancy	6 (4)	6 (5)
No discrepancy	75 (49)	61 (48)
Unknown if discrepancy	71 (47)	60 (47)

SSED=summary of safety and effectiveness data.

*Study was classified as pivotal if it was only one included in summary, multicenter randomized controlled trial, or explicitly noted in summary as being pivotal study.

Table 5 | Comparison of primary endpoints reported in FDA summaries with corresponding publications and those published studies. Figures are numbers (percentage) unless stated otherwise

	All studies (n=152)			Pivotal studies* (n=127)		
	Summary	Published	Odds ratio (95% CI), P value†	Summary	Published	Odds ratio (95% CI), P value†
Type of endpoint analysis:						
Non-inferiority	11 (7)	15 (10)	1.65 (0.61 to 4.44), 0.32	9 (7)	13 (10)	1.84 (0.61 to 5.55), 0.28
Equivalence	4 (3)	3 (2)	—‡	4 (3)	3 (2)	0.70 (0.14 to 3.67), 0.68
Historical control	13 (9)	3 (2)	0.10 (0.02 to 0.48), 0.004	13 (10)	3 (2)	0.10 (0.02 to 0.48), 0.004
Objective performance criteria	47 (31)	16 (11)	0.15 (0.07 to 0.32), <0.001	43 (34)	15 (12)	0.15 (0.07 to 0.34), <0.001
Unknown	77 (51)	115 (76)	4.01 (2.30 to 6.99), <0.001	58 (46)	93 (73)	4.30 (2.37 to 7.78), <0.001
Controls:						
Prospective	63 (41)	63 (41)	1.00 (0.50 to 1.99), 1.0	55 (43)	55 (43)	1.00 (0.43 to 2.30), 1.0
Retrospective	20 (13)	15 (10)	0.64 (0.27 to 1.47), 0.29	19 (15)	13 (10)	0.54 (0.22 to 1.33), 0.18
None	67 (44)	68 (45)	1.05 (0.57 to 1.94), 0.88	53 (42)	54 (43)	1.06 (0.54 to 2.11), 0.86
Unknown	2 (1)	6 (4)	3.34 (0.63 to 17.66), 0.16	0 (0)	5 (4)	—‡
Mean No of patients analyzed	325.6	334.4	10.32 (−57.18 to 77.83), 0.77§	338.1	348.6	12.22 (−23.64 to 48.08), 0.51§
Direction of favorability of endpoint results:						
Device	94 (62)	65 (43)	0.44 (0.27 to 0.70), <0.001	86 (68)	58 (46)	0.37 (0.22 to 0.63), <0.001
Control	8 (5)	3 (2)	0.35 (0.09 to 1.37), 0.13	8 (6)	3 (2)	0.35 (0.09 to 1.37), 0.13
Neither	12 (8)	14 (9)	1.33 (0.56 to 3.15), 0.51	10 (8)	11 (9)	1.26 (0.49 to 3.26), 0.63
Unknown	38 (25)	70 (46)	2.77 (1.65 to 4.64), <0.001	23 (18)	55 (43)	3.84 (2.10 to 7.05), <0.001

*Study was classified as pivotal if it was only one included in summary, multicenter randomized controlled trial, or explicitly noted in summary as being pivotal study.

†Calculated by generalized linear mixed effects model with canonical logit link and random effect placed on intercept to account for heterogeneity across devices.

‡Too few events or counter-events for fitting of random effects model.

§Mean difference, (95% CI), P value calculated by general linear mixed effects model with random effect placed on intercept to account for device clustering.

objective performance criteria. The pivotal primary results showed the same trends in significance. In one pivotal study, a disparity in reported significance substantially changed the interpretation of the result such that the device seemed more favorable in the summary than in the publication.^{7 8}

Discussion

Principal findings

About half (51%) of clinical studies of novel high risk cardiovascular devices remain unpublished over two years after FDA approval. Even when these studies are published, the results can differ substantially from FDA summaries in their presentation of primary endpoints. This finding remained true even when we restricted our analysis to pivotal studies. Reassuringly, a comparison of clinical design features including randomization, blinding, and number of centers as reported in summaries and publications showed that they were nearly identical. A quarter of trials, however, had discrepancies between summaries and publications in the numbers of enrolled patients, suggesting that the composition of the patient population had changed substantially and introducing a possible source of bias in reporting. This is further supported by the finding that sex and age demographics also differed between summary and publication in over 10% of the studies.

Clinicians might not be aware of the FDA device summaries and so might not critically examine these data. Thus for many high risk devices, clinical trial evidence might never be made readily available to the medical community or might be made available only after a long delay. While nearly all studies with disclosed funding were funded by industry, lengthy delays of publication have also been noted in NHLBI funded cardiovascular trials and in other NIH supported trials.⁵⁻¹¹

Summaries and published studies differed in how their primary endpoints were presented: some endpoints labeled as primary in summaries were labeled as secondary in published studies and nearly 10% were entirely missing. These alterations mean that primary endpoints were rebranded to modify the emphasis of their findings in the literature. For instance, the summary for the “Xience V Rapid Exchange Everolimus-Eluting Stent System” (Abbott Vascular, IL, US) designates “co-primary endpoints” of “in-segment late loss at 240 days and target vessel failure at 270 days.”¹² In the corresponding publication, however, the primary endpoint is “in-segment late loss at 240 days,” while the “secondary end point was ischemia-driven target vessel failure at 270 days.”¹³ In this case, the primary endpoint in the publication showed superiority of the tested device, whereas the rebranded secondary endpoint showed only non-inferiority. Others have reported similar findings.¹⁴ In a Cochrane review, 4-50% of randomized controlled trials exhibited discrepancies between primary outcomes registered and those published.¹⁵

Comparisons of results for primary endpoints between summaries and published studies showed that less than half of such results were identical, and 11% were substantially different. In addition, while many primary endpoints (31%) in the summaries were measured against objective performance criteria, far fewer endpoints (12%) were held to these criteria in the published literature. The significantly decreased number of objective performance criteria in the corresponding publications was the primary reason why devices seemed more favorable in the summaries than in the literature. Nonetheless, in several of these cases, the disappearance of the objective performance criteria in the literature was associated with failure of the primary

endpoint result to meet target criteria in the summary. For instance, two of the three endpoints for the “7F Freezor Cardiac Cryoablation Catheter” (Medtronic Cryocath, QC, Canada) failed to meet the prespecified objective performance criteria.¹⁶ The publication for the same clinical trial, however, includes no mention of any objective performance criteria for any of its endpoints but simply reports a procedural success rate of 83% and major complication rate of 4.2%, and concludes that catheter cryoablation is a safe and clinically effective alternative to the standard of care.¹⁷

In rare cases, the endpoint reporting in the summary seemed more favorable than in the publication. For example, the summary for the “Talent Thoracic Stent Graft System” (Medtronic Vascular, MN, US) states a primary safety endpoint of all cause mortality at 12 months, to be compared with both an objective performance criterion of <29.8% and a retrospective control group. The summary reported 12 month all cause mortality of 16.1%, which both meets the objective performance criteria and is significantly lower than the retrospective control group rate of 20.6%.⁷ In contrast, the publication fails to mention any objective performance criteria, but reports a non-significant P value of 0.29 for the comparison with the retrospective control.⁸

Study limitations

A potential limitation of this study is that there is no systematic listing in the summaries of the trial principal investigators and clinical trial registries. In addition, we performed the Medline search on 15 January 2013, for publications up to 1 January 2013, and delays in indexing publications on Medline can exceed two weeks. To ensure that we did not inadvertently overlook a publication, we directly contacted the research divisions of device manufacturers and asked about any publications not found through our search algorithm. All of the manufacturers who responded confirmed that there was no publication for the respective study of interest. Another limitation is that our study focuses on selective outcome reporting and reporting of selected analyses such as follow-up intervals, and individual outcomes of composite endpoints were not explored.

Comparison with other studies

This study is the first to document selective reporting of device trials at both the study level and the outcome level, but it underscores a broader problem in the way clinical trial evidence is presented. Similar discrepancies have been identified for drug trials by comparing trial data reported in FDA New Drug Applications with the data in corresponding publications. A comparison of 128 drug trials as reported in FDA documents compared with the corresponding journal publications found differences in primary outcomes reported, types of data, results, P values, confidence intervals, and adverse event tables.⁵ Several types of selective outcome reporting for trials of off-label use of gabapentin have also been described, including discrepancies in

numbers of patients randomized, definitions of primary outcomes, significance of reported outcomes, additions and deletions of outcomes, and differences in analyses.^{18 19} These discrepancies in the reporting led to more favorable efficacy data being presented in publications. These and our findings point to the importance of mandatory registration on a public clinical trials platform. Clinicaltrials.gov is an important step in this direction, but recent data show that published trials often have discrepant findings between clinicaltrials.gov and publications.²⁰

Conclusions and policy implications

Our findings have potential international implications. In the European Union, independent organizations called notified bodies authorize device marketing. The evidence reviewed by notified bodies is not mandated to be publicly available, which makes it challenging to directly compare published data with data reviewed by notified bodies.²¹ As devices generally receive CE mark before FDA approval, however, it is less likely that clinical studies have been published in the medical literature at the time of CE mark.²²

To protect the integrity of clinical trial research and, ultimately, safety of patients, the data from FDA reviews of devices should be readily accessible through peer reviewed publications. Although systematic reviews increasingly serve as a basis for evidence based clinical practice guidelines, documents from regulatory agencies are seldom included. Systematic reviewers might be unaware of the data that can be obtained from regulatory agencies.^{23 24} Thus, clinical practice guidelines might not be based on complete and accurate information about drugs or devices. Recent efforts, such as the intent of the European Medicines Agency (EMA) to make all clinical trial data for drug applications publicly available, would help enormously to fill these gaps by increasing the transparency and availability of important clinical trial data.^{25 26} In response to pushback, including from 33 drug companies or pharmaceutical industry associations, however, the EMA has imposed severe restrictions on the use of these data along with confidentiality agreements while making the data available only to registered users in a “view on screen only” mode.²⁷ A recent study shows that FDA documents are more challenging to navigate than EMA reports, and the accessibility and user friendliness of FDA reports needs to be considerably improved.²⁸ As clinicians can use devices immediately after FDA approval, it is in the public interest that all of the data be available to clinicians at that time.

We thank Jeppe Schroll for reviewing this manuscript. This work was also conducted with statistical support from Harvard Catalyst, Harvard Clinical and Translational Science Center, and the Massachusetts General Hospital Department of Biostatistics.

Contributors: LC designed the study, acquired the data, performed the statistical analyses, and drafted the paper. SSD designed the study, acquired the data, and revised the draft paper. JC acquired the data. LAB and RFR designed the study and revised the draft paper. RFR is guarantor.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: RFR is a member of the FDA Circulatory System Devices Panel; no support from any organisation for the submitted work; no financial relationships with any organizations 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.

Ethical approval: Not required.

Transparency: The lead author (the manuscript's guarantor) affirms that this 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 (and, if relevant, registered) have been explained.

Data sharing: Additional data from the study can be obtained for the corresponding author.

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/>.

- Maisel WH, Moynahan M, Zuckerman BD, et al. Pacemaker and ICD generator malfunctions: analysis of Food and Drug Administration annual reports. *JAMA* 2006;295:1901-6.
- Premarket Approval. US Food and Drug Administration. www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/default.htm
- Dhruva SS, Bero LA, Redberg RF. Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices. *JAMA* 2009;302:2679-85.
- Chen CE, Dhruva SS, Redberg RF. Inclusion of comparative effectiveness data in high-risk cardiovascular device studies at the time of premarket approval. *JAMA* 2012;308:1740-2.
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 2008;5:1561-70.
- Ross JS, Mocanu M, Lampropoulos JF, Tse T, Krumholz HM. Time to publication among completed clinical trials. *JAMA Intern Med* 2013;173:825-8.
- "Talent Thoracic Stent Graft System" summary of safety and effectiveness data. US Food and Drug Administration. www.accessdata.fda.gov/cdrh_docs/pdf7/P070007b.pdf
- Fairman RM, Criado F, Farber M, et al. Pivotal results of the Medtronic Vascular Talent Thoracic Stent Graft System: the VALOR trial. *J Vasc Surg* 2008;48:546-54.
- Gordon D, Taddei-Peters W, Mascette A, et al. Publication of trials funded by the National Heart, Lung, and Blood Institute. *N Engl J Med* 2013;369:1926-34.
- Ross JS, Tse T, Zarin DA, et al. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ* 2012;344:d7292.
- 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;358:252-260.
- "XIENCE V Rapid Exchange (RX) Everolimus Eluting Coronary Stent System" summary of safety and effectiveness data. US Food and Drug Administration. www.accessdata.fda.gov/cdrh_docs/pdf7/P070015b.pdf
- Stone GW, Midei M, Newman W, et al. Comparison of an everolimus-eluting stent and a paclitaxel-eluting stent in patients with coronary artery disease: a randomized trial. *JAMA* 2008;299:1903-13.
- Mathieu S, Boutron I, Moher D, Altman DG, Ravaut P. Comparison of registered and published primary outcomes in randomized clinical trials. *JAMA* 2009;302:977-84.
- Dwan K, Altman DG, Cresswell L, et al. Comparison of protocols and registry entries to published reports for randomized controlled trials. *Cochrane Database Syst Rev* 2011;1:MR000031.
- "7F Freezor Cardiac Cryoablation Catheter and CCT.2 CryoConsole System" summary of safety and effectiveness data. US Food and Drug Administration. www.accessdata.fda.gov/cdrh_docs/pdf2/P020045b.pdf
- Friedman PL, Dubuc M, Green MS, et al. Catheter cryoablation of supraventricular tachycardia: results of the multicenter prospective "frosty" trial. *Heart Rhythm* 2004;1:129-38.
- Vedula S, Bero L, Scherer R, Dickersin K. Outcome reporting in industry-sponsored trials of Gabapentin for off-label use. *N Engl J Med* 2009;361:1963-71.
- Vedula S, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of Gabapentin. *PLoS Med* 2013;10:e1001378.
- Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA* 2014;311:1063-5.
- Kramer DB, Xu S, Kesselheim AS. Regulation of medical devices in the United States and European Union. *N Engl J Med* 2012;366:848-55.
- Basu S, Hassenplug JC. Patient access to medical devices—a comparison of U.S. and European review processes. *N Engl J Med* 2012;367:485-8.
- Wolfe N, Gotzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. *Syst Rev* 2013;2:31.
- Schroll JB, Bero L, Gotzsche P. Searching for unpublished data for Cochrane reviews: cross sectional study. *BMJ* 2013;346:f2231.
- Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS Med* 2012;9:e1001201.
- Eichler H-G, Abadie E, Breckenridge A, Leufkens H, Rasi G. Open clinical trial data for all? A view from regulators. *PLoS Med* 2012;9:e1001202.
- Torjeson I. European drug agency backtracks on plan to give researchers access to clinical trial reports. *BMJ* 2014;2014:348:g3432.
- Schroll JB, Abdel-Sattar M, Bero L. The Food and Drug Administration reports provided more data but were more difficult to use than the European Medicines Agency reports. *J Clin Epidemiol* 2015;68:102-7.