Characteristics of efficacy evidence supporting approval of supplemental indications for prescription drugs in United States, 2005-14: systematic review

Bo Wang, Aaron S Kesselheim

EDITORIAL
by Light and Lexchin

RESEARCH, p 12

Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, 1620 Tremont St, Boston, MA 02120, USA

Correspondence to: A S Kesselheim
akesselheim@partners.org

Cite this as: *BMJ* 2015;351:h4679
doi: 10.1136/bmj.h4679

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h4679

thebmj.com

- Feature: No correction, no retraction, no apology, no comment: paroxetine trial reanalysis raises questions about institutional responsibility (*BMJ* 2015;351:h4629)
- Feature: FDA drug summaries: a simplification too far (*BMJ* 2015;350:h3135)
- Editorial: When drugs don't make it (*BMJ* 2015;350:h2852)

STUDY QUESTION

What are the characteristics of clinical efficacy trials used to support the approval of supplemental indications for prescription drugs approved by the US Food and Drug Administration (FDA)?

SUMMARY ANSWER

Approvals of supplemental indications were supported by low rates of clinical efficacy trials using active study comparators or study endpoints directly related to patients' function or mortality, especially among supplements that expanded the drugs' approved patient population.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

New prescription drugs are approved by regulatory agencies such as the FDA on the basis of pivotal clinical trials that vary in some of their essential features, including type of comparator and study endpoint. Similar variations in characteristics of clinical trials underlie the approval of these agents' subsequently approved supplemental indications.

Selection criteria for studies

We used the publicly accessible Drugs@FDA database to gather supplemental indications approved between 2005 and 2014. We classified each supplemental indication as a new indication, a modification of an already approved indication, or an expansion of the patient population. We then accessed FDA approved prescription drug labels from the same database to determine the use of study comparators (active, placebo, historical, none) and study endpoints (clinical outcomes, clinical scales, surrogate) used in the main efficacy trials supporting each supplemental indication, as well as in these drugs' originally approved indications. In addition, we determined the orphan drug status (granted to treatments for diseases affecting fewer than 200 000 people in the United States annually) of each supplemental and originally approved indication.

Characteristics of clinical efficacy trials supporting approval of supplemental drug indications

Supplement category*	Active comparator—No (%)	Clinical outcomes—No (%)
New indication	41/136 (30)	44/137 (32)
Modified indication	47/93 (51)	28/93 (30)
Expanded population	7/65 (11)	14/65 (22)
P value	<0.001	0.29

*Total n=295 supplemental indications for 164 novel therapeutic agents. No study comparator information for trials supporting the supplemental indication approval for celecoxib (Celebrex, Pfizer, New York City, NY, USA) in July 2005.

Primary outcome(s)

The primary outcomes were the study comparators (active, placebo, historical, none) and study endpoints (clinical outcomes, clinical scales, surrogate) used in the main efficacy trials.

Main results and the role of chance

Our study sample included 295 supplemental indications. Thirty per cent (41/136) of approvals for new indications were supported by efficacy trials with active comparators, compared with 51% (47/93) of modified use approvals and 11% (7/65) of approvals expanding the patient population (P<0.001), almost all of which related to paediatric patients (61/65; 94%). Trials using clinical outcome endpoints led to approval for 32% (44/137) of supplemental approvals for new indications, 30% (28/93) of modified indication approvals, and 22% (14/65) of expanded population approvals (P=0.29). Orphan drugs had supplemental approvals for 40 non-orphan indications, which were supported by similar proportions of trials using active comparators (28% (11/40) for non-orphan supplemental indications versus 24% (10/42) for original orphan indications; P=0.70) and clinical outcome endpoints (25% (10/40) versus 31% (13/42); P=0.55).

Bias, confounding, and other reasons for caution

We included only supplemental indication approvals for drugs originally approved as novel therapeutic agents (that is, new molecular entities and original therapeutic biological agents). In addition, we assessed the clinical trial evidence supporting supplemental and original indication approvals by using FDA approved drug labels rather than FDA medical reviews (which contain more detailed clinical trial information) owing to inaccessibility of the medical reviews for 80% of the supplemental indications in our study. However, the distribution of study comparators and clinical endpoints for the originally approved indications in our study is consistent with previous research.

Study funding/potential competing interests

ASK's work is supported by the Greenwall faculty scholar programme in bioethics and the Harvard programme in therapeutic science.

Trends in utilisation of FDA expedited drug development and approval programmes, 1987-2014: cohort study

Aaron S Kesselheim, Bo Wang, Jessica M Franklin, Jonathan J Darrow

EDITORIAL
by Light and Lexchin

RESEARCH, p 11

Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston 02120, MA, USA

Correspondence to: A S Kesselheim
akesselheim@partners.org

Cite this as: *BMJ* 2015;351:h4633
doi: 10.1136/bmj.h4633

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h4633

STUDY QUESTION

How has the US Food and Drug Administration's approval of new drugs through its special expedited development and review programmes changed over the past 25 years?

SUMMARY ANSWER

In the past two decades, drugs newly approved by the FDA have been associated with an increasing number of expedited development or review programmes, a trend that is being driven by non-first in class drugs that are potentially less innovative.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

The FDA offers four primary pathways to accelerate the development and review of drugs intended for serious diseases or unmet medical need, but approving investigational drugs on expedited time frames and more limited data increases risks for patients. We found that these "exceptions" to the basic drug development paradigm are becoming substantially more common, particularly in the context of non-first in class drugs.

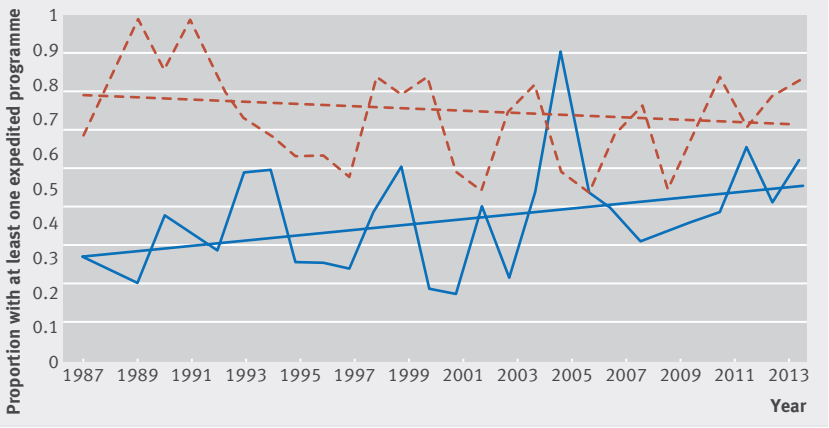
Participants and setting

Database comprising 774 drugs approved by the FDA from 1987 to 2014.

Design, size, and duration

Data obtained on the drugs in the cohort include innovativeness (first in class versus not first in class based on an FDA determined framework), World Health Organization Anatomic Therapeutic Classification, and which (if any) of the FDA's four primary expedited development and review programmes or designations were associated with each drug: orphan drug, fast track, accelerated approval, and priority review. Logistic regression models evaluated trends in the proportion of drugs associated with each of the four expedited development and review programmes. Poisson models were employed, with the number of programmes as the dependent variable and a linear term for year of approval to evaluate the number of programmes associated with each approved drug over time. We compared the difference in trends between drugs that were first in class and those that were not.

Time trend analyses comparing all expedited programmes associated with first in class and follow-on therapeutic approved by FDA showing proportion of newly approved first in class (red dotted line) and non-first in class (blue line) prescription drugs, 1987-2014, that were granted at least one of the four programmes



Main results and the role of chance

Among the cohort of drugs one third represented first in class agents. There was a significant increase of 2.6% each year in the number of expedited review and approval programmes granted to each newly approved agent (incidence rate ratio 1.026, 95% confidence interval 1.017 to 1.035, P<0.001) and a 2.4% increase in the proportion of drugs associated with at least one such programme (odds ratio 1.024, 95% confidence interval 1.006 to 1.043, P=0.009). Driving this trend was an increase in the proportion of approved, non-first in class drugs associated with at least one programme (P=0.03 for interaction).

Bias, confounding, and other reasons for caution

The programmes we analysed have some varying characteristics; for example, accelerated approval and fast track formally change the nature of the evidence considered sufficient for approval, whereas the orphan drug and priority review designations do not. In addition, some of the programmes are correlated, for example, in that a drug granted accelerated approval designation is more likely to receive priority review status. But in a sensitivity analysis, we excluded the orphan drug designation from the analysis and the trends we identified did not change.

Generalisability to other populations

Other drug regulatory agencies around the world have expedited review and development programmes, but may implement their programmes differently or have different requirements for qualification.

Study funding/potential competing interests

ASK is supported by the Greenwall faculty scholars programme in bioethics and the Harvard programme in therapeutic science. We have no competing interests.

Defining safe criteria to diagnose miscarriage: prospective observational multicentre study

Jessica Preisler,[1] Julia Kopeika,[2] Laure Ismail,[1 3] Veluppillai Vathanan,[4] Jessica Farren,[1] Yazan Abdallah,[1] Parijat Battacharjee,[5] Caroline Van Holsbeke,[6] Cecilia Bottomley,[4] Deborah Gould,[3] Susanne Johnson,[7] Catriona Stalder,[1] Ben Van Calster,[8] Judith Hamilton,[2] Dirk Timmerman,[6 8] Tom Bourne[1 6 8]

EDITORIAL
by McCarthy and Tong

[1]Queen Charlotte's and Chelsea Hospital, Imperial College, London W12 0HS, UK
[2]Early pregnancy and acute gynaecology unit, St Thomas' Hospital, London, UK
[3]St Mary's Hospital, Imperial College NHS Trust, London, UK
[4]Chelsea and Westminster Hospital, London, UK
[5]Northwick Park Hospital, Harrow, Middlesex, UK
[6]Department of Obstetrics and Gynaecology, University Hospitals Leuven, Leuven, Belgium
[7]Princess Anne Hospital, Southampton, Hampshire, UK
[8]Department of Development and Regeneration, KU Leuven, Belgium
Correspondence to: T Bourne
tbourne@imperial.ac.uk

Cite this as: *BMJ* 2015;351:h4579
doi: 10.1136/bmj.h4579

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h4579

STUDY QUESTION

Are criteria currently used to diagnose miscarriage safe?

SUMMARY ANSWER

Recently changed cut-offs of gestational sac and embryo size defining miscarriage are appropriate, but guidance on intervals between scanning and expected findings on repeat scans are too liberal.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Mean gestational sac diameter (MSD) and crown-rump length (CRL) measurements are used to diagnose miscarriage. Criteria for 100% specificity for miscarriage were an empty gestational sac of mean diameter ≥25 mm on initial scan or an embryo with a CRL ≥7 mm with no heart activity; on follow-up scans an embryo with no heart activity on an initial and repeat scan ≥7 days later, no embryo and MSD <12 mm initially if sac size had not doubled after ≥14 days, and no embryo and MSD ≥12 mm initially with no embryo heart activity after ≥7 days; finally, beyond 70 days' gestation, an MSD ≥18 mm with no embryo or an embryo with CRL ≥3 mm with no heart activity.

Participants and setting

Seven hospital based early pregnancy assessment units in the United Kingdom.

Design, size, and duration

Prospective multicentre observational study. 2845 women with intrauterine pregnancies of unknown viability were recruited consecutively from September 2010 to March 2011 and then from August 2011 to May 2013. The reference standard was viability at 11-14 weeks.

Main results and the role of chance

We present specificity data, as the priority in the analysis was for a minimal chance of a false positive test result for miscarriage. Specificity relates to the percentage of viable pregnancies classified as not being a miscarriage. The following indicated a miscarriage at initial scan: MSD ≥25 mm with an empty sac (364/364 specificity: 100%, 95% confidence interval 99.0% to 100%), embryo with CRL ≥7 mm without visible heart activity (110/110 specificity: 100%, 96.7% to 100%), MSD ≥18 mm for gestational sacs without an embryo after 70 days' gestation (907/907 specificity: 100%, 99.6% to 100%), and embryo with CRL ≥3 mm without visible heart activity after 70 days' gestation (87/87 specificity: 100%, 95.8% to 100%). The following were indicative of miscarriage at a repeat scan: initial scan and repeat scan after ≥7 days showing an embryo without visible heart activity (103/103 specificity: 100%, 96.5% to 100%), pregnancies without an embryo and MSD <12 mm where the mean diameter has not doubled after ≥14 days (478/478 specificity: 100%, 99.2% to 100%), pregnancies without an embryo and MSD ≥12 mm showing no embryo heart activity after ≥7 days (150/150 specificity: 100%, 97.6% to 100%).

Bias, confounding, and other reasons for caution

Despite the large number of cases, numbers around some decision boundaries are too small to be definitive. Around 10% of women were lost to follow-up.

Generalisability to other populations

The multicentre design of the study in different types of hospitals with diverse populations should make our findings generalisable.

Study funding/potential competing interests

TB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. DT is fundamental clinical researcher of the Research Foundation–Flanders. We have no competing interests.

Overview of criteria to define miscarriage associated with no false positive test results according to data in study

Initial scan group	Empty sac	Gestational+yolk sacs	Embryo (no heart activity)
Initial scan criteria for miscarriage	MSD ≥25 mm	MSD ≥18 mm at GA ≥70 days	CRL ≥7 mm
	MSD ≥18 mm at GA ≥70 days		CRL ≥3 mm at GA ≥70 days
Repeat scan criteria for miscarriage	MSD <12 mm on initial scan, not doubled and no embryo heart activity after ≥14 days	MSD <12 mm on initial scan, not doubled and no embryo heart activity after ≥14 days	No embryo heart activity after ≥7 days
	MSD ≥12 mm on initial scan, no embryo heart activity after ≥7 days	MSD ≥12 mm on initial scan, no embryo heart activity after ≥7 days	

MSD=mean gestational sac diameter, GA=gestational age, CRL=crown-rump length

Frequency of discrepancies in retracted clinical trial reports versus unretracted reports: blinded case-control study

Graham D Cole, Alexandra N Nowbar, Michael Mielewczik, Matthew J Shun-Shin, Darrel P Francis

EDITORIAL by Freemantle and Rait

International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, W2 1LA, UK
Correspondence to: G D Cole
g.cole@imperial.ac.uk

Cite this as: *BMJ* 2015;351:h4708
doi: 10.1136/bmj.h4708

This is a summary of a paper that was published on thebmj.com as *BMJ* 2015;351:h4708

STUDY QUESTION

Are discrepancies (defined as mathematically or logically contradictory statements) in clinical trial reports associated with retraction?

SUMMARY ANSWER

Discrepancies were more common in retracted trial reports than in paired unretracted trial reports.

WHAT IS KNOWN AND WHAT THIS PAPER ADDS

Although retractions are increasing, they remain far less common than erroneous research, implying that the literature could be burdened by findings that are insecure but unrecognised as such. This study provides evidence that if there are numerous discrepancies they should not be ignored as they may signal unreliability.

Participants and setting

Fifty randomly selected manuscripts classified as retracted clinical trials on PubMed (from inception to December 2012), paired with 50 controls. Controls were the preceding unretracted clinical trial published in the same journal.

Design, size, and duration

Retracted manuscripts had markings of retraction removed and were randomly intermingled with unretracted control manuscripts. Three scientists, blinded to retraction status and with no specialist skill in the field, scrutinised the 100 manuscripts and recorded any discrepancies. Proposed discrepancies were subsequently pooled and accepted as valid only if agreed by all three and a fourth scientist. After this, the study was unblinded for analysis.

Primary outcome, risks, and exposures

The measured outcome was total number of discrepancies in a clinical trial report.

Main results and the role of chance

Of 479 discrepancies found in the 100 trial reports, 348 were in the 50 retracted reports and 131 in the 50 unretracted reports. Discrepancy counts were 2.7-fold higher in retracted trial reports than in paired unretracted trial reports (median 4 (interquartile range 2-8.75) *v* 0 (0-5); P<0.001). Trial reports with a discrepancy were significantly more likely to be retracted than those without a discrepancy (odds ratio 5.7 (95% confidence interval 2.2 to 14.5); P<0.001). Results from a retrospective analysis indicated that citations and journal impact factor were unlikely to affect the result.

Bias, confounding, and other reasons for caution

By selecting the preceding trial report in the same journal as the control, we addressed many likely confounders: editorial processes, impact factor, readership, and post-publication policy. We could not control for other possible confounders, such as medical specialty, without risking introducing bias through selection. The sample size was constrained by resources because of the time taken to identify, verify, and collate discrepancies. We also could not establish the mechanism for the discrepancies (which might range from innocent administrative error to intentional fabrication), and it remains uncertain whether specific types of discrepancy are particularly strong markers of trial report unreliability.

Generalisability to other populations

We selected clinical trial reports because the results could directly affect patient care, and are more expensive and time consuming to replicate than other research. Since it would be impractical to repeat entire trials, readers need other pointers to reliability.

Study funding/potential competing interests

This study was not funded by any external organisation. All authors are associated with Imperial College London. GDC and MJS-S are clinical research training fellows at the British Heart Foundation (FS/12/12/29294 and FS/14/27/30752, respectively). DPF is a senior clinical research fellow at the British Heart Foundation (FS/10/038). Neither the institution nor any funder had any role in devising, conducting, analysing, or reporting on this study. We declare no competing interests.

Paired comparison of discrepancy counts between retracted cases and unretracted controls