RESEARCH PAPER

Late, never or non-existent: the inaccessibility of preclinical evidence for new drugs

C A Federico[1], B Carlisle[1], J Kimmelman[1] and D A Fergusson[2]

[1]*Study of Translation, Ethics, and Medicine (STREAM), Biomedical Ethics Unit, McGill University, Montreal, QC, Canada, and* [2]*Centre for Practice-Changing Research, Ottawa Hospital Research Institute, University of Ottawa, Ottawa, ON, Canada*

BACKGROUND AND PURPOSE
Animal studies establish much of the evidence used to support clinical development of new drugs. Recent studies suggest that many preclinical investigations are withheld from publication, leading to exaggerated estimates of clinical utility. We sought to estimate the volume and properties of all published animal efficacy studies for a cohort of novel drugs.

EXPERIMENTAL APPROACH
We searched biomedical databases to identify 47 novel drugs whose first trials were reported between 2000 and 2003, inclusive. Next, we searched for all published animal studies testing the same drug, regardless of publication date. We then extracted items from titles and abstracts of eligible studies.

KEY RESULTS
We identified 2462 efficacy studies, representing an average of 52 studies per drug. No published efficacy studies were available for three drugs in our sample. The volume of efficacy studies was related to how far the drug had progressed in clinical development (Spearman's correlation coefficient = 0.66, $P < 0.0001$). Most (87%) accessible animal efficacy studies were reported after publication of the first trial, and for 17% of the drugs in our sample, no efficacy studies were published before the first trial report. Disease indications used in trials often did not match those modelled in efficacy studies; for 35% of indications tested in trials, we were unable to identify any published efficacy studies in models of the same indication.

CONCLUSIONS AND IMPLICATIONS
The volume of published efficacy studies is large, although numerous gaps reflect non-publication, publication delay or non-performance of efficacy studies supporting trials.

Abbreviations
IND, Investigational New Drug; KRAS, Kirsten rat sarcoma viral oncogene homolog; MeSH, Medical Subject Headings; PRISMA, Preferred Reporting Items of Systematic Reviews and Meta-Analyses

Introduction

Animal studies establish much of the evidence used to support the clinical development of new drugs. In particular, studies aimed at demonstrating a drug's ability to intervene in a pathophysiological process provide the biological rationale for early phase trials, and help secure the moral basis for exposing volunteers to unproven drugs (for brevity, we call these 'efficacy studies').

Several recent studies suggest that many preclinical investigations are withheld from publication. In one meta-analysis, non-publication of preclinical stroke studies was

purported to explain an exaggerated effect size of 30% (Sena *et al.*, 2010). Other reports describe problems replicating preclinical studies (Prinz *et al.*, 2011; Steward *et al.*, 2012) – a phenomenon that some attribute to selective reporting. Such non-reporting is likely to reflect that private drug developers have little incentive to publish preclinical studies. However, it potentially deprives investigators, ethics committees and other decision makers of complete evidence for making risk/ benefit judgments. It also does not redeem the killing of animals used in investigations and frustrates the search for explanations when drug trials fail to reproduce the promise shown in preclinical studies (Kimmelman and Anderson, 2012).

In what follows, we sought to capture and characterize all published animal efficacy studies for a cohort of 47 novel drugs where the first reported trials were published between 2000 and 2003, inclusive. We find a large volume of published animal efficacy studies; however, much of the evidence supporting specific trials is published belatedly, if at all. We close with recommendations for improving the reporting of preclinical efficacy studies.

Methods

Overview

Our primary goal was to identify and characterize all published *in vivo* animal efficacy studies for a random sample of 50 drugs. To capture animal studies that were published as long as 12 years after the first published trial, we restricted our sample to agents where the first ever trials were published between January 2000 and December 2003. We created a convenience sample of 50 novel drugs by searching MEDLINE and Embase for early phase trials of drugs, and then used secondary searches of PubMed to screen for drugs that had no reported trials published prior to 2000 (see Supporting Information Table S1). From our overall sample of 340 novel drugs, we randomly selected 50 using a random number generator. We chose this sample size because it seemed a reasonable balance between tractability and comprehensiveness. For each drug, we used abstracts and titles to identify the disease indication for the first published trial (see Supporting Information Table S2). For studies testing healthy volunteers, the full publication was curated to determine the presumed disease area. If none could be found, the indication stated in the next published trial was recorded. We performed a PubMed search of trials for each drug to identify the second disease indication tested in a published trial.

Capture of animal efficacy studies

Our search strategy, described in Supporting Information Table S3, was a modified version of the filters described by Hooijmans *et al.* (2010) and de Vries *et al.* (2011). It used relevant Medical Subject Headings (MeSH) terms and subheadings to identify investigations in an exhaustive list of animal species. This search was then combined with individual drug names and all known aliases identified using the PubChem Compound search function of PubMed (Bolton *et al.*, 2008). Searches were performed between 28 June 2012 and 11 August 2012 using the OvidSP search platform. Data-

bases queried included Ovid MEDLINE In-Process & Other Non-Indexed Citations, Ovid MEDLINE (dates of coverage from 1948 to 2012), Embase Classic + Embase database (dates of coverage from 1974 to 2012) and BIOSIS Previews (dates of coverage from 1969 to 2012). Captured entries were exported into EndNote and then imported into Excel for screening following removal of duplicate entries. Trial references were also screened for preclinical studies not identified by the above method.

Screening and validation

Titles and abstracts were screened, and articles retained if they met the following inclusion criteria: they (i) measured disease response; (ii) contained primary data; and (iii) included at least one *in vivo* non-human animal experiment using the agent of interest. To ensure intercoder agreement in screening, entries from a pilot database were coded independently by two individuals (CF and BC). Double coding continued until the rate of agreement between the two coders exceeded 95%, at which time the remaining preclinical entries were split and screened individually by the same two individuals. To test the specificity of our search, we used five review articles for one of the drugs in our cohort – the stroke agent NXY-059, selected because of the availability of several preclinical systematic reviews (Macleod *et al.*, 2008; Bath *et al.*, 2009) – to identify all referenced animal efficacy studies of this agent. Our search strategy captured all 13 published studies cited in review articles.

Extraction and analysis of efficacy studies

We extracted basic items from titles and abstracts of eligible studies, including publication date, publication type (i.e. abstract vs. full-text article), indication tested and species used. We also extracted basic information from the initial clinical trial for each of the agents, including publication date, trial phase, funding source and indication tested.

Drug trials and animal efficacy studies often sampled different disease indications. The relationship between timing of animal efficacy publication and timing of trial publication was probed by creating a series of indication-matched animal study/clinical trial dyads for all drugs in our cohort, and determining the length of time between each. We created two dyad types: (i) the initially published clinical trial indication and first corresponding animal efficacy study; and (ii) the second published clinical trial indication and corresponding animal efficacy study. Rules on how indications in preclinical studies were matched to trials are provided in Supporting Information Note S1. All animal/clinical trial dyads are presented in Supporting Information Table S4. CF and JK performed all pairings independently and discrepancies were resolved through discussion until consensus was reached. In all cases, we also combed the bibliography of trials for animal efficacy studies in matching indications. The searching of citations led to the identification of three animal efficacy studies not previously captured in our literature search. For those agents where no indication-matched efficacy studies could be found, minimal publication lag was determined by presuming a preclinical publication date of May 2013.

Last, we tested the *a priori* formulated hypothesis that the volume of published animal studies was correlated with how

far an agent had progressed in clinical development. When the phase was not explicitly stated, judgments were made (by CF) based on sample size, participant type and outcome measures. Industry sponsorship was determined by curating both clinical and preclinical studies; when a funding source was not explicitly stated, the study was marked as industry sponsored if either the first or corresponding author reported an industry-related affiliation. Spearman's rank-order correlation was used to test our hypothesis; *P*-values of <0.05 were considered to be statistically significant.

Results

Description of drugs used for study

We began with a random convenience sample of 50 agents entering the published record between 2000 and 2003, inclusive. Three agents were later excluded; two trials had been published earlier than 2000. Another was excluded because the agent was used as a challenge for a physiological study, not as an intervention. The properties of the remaining 47 interventions are presented in Table 1. Our cohort demonstrated a tendency to progress further in clinical development than the general average for new drugs (e.g. 23% of drugs in our cohort were ultimately licensed, whereas only 11% of drugs entering clinical development are licensed) (Kola and Landis, 2004). More than half of the initial clinical trials of these agents (57%, $n = 27$) reported funding by industry sponsors.

Volume of accessible efficacy studies

In total, 2462 efficacy studies were identified for the 47 agents in our sample (Figure 1). Figure 2 presents the volume of efficacy studies for each agent. Most studies appeared as full publications (94%, $n = 2314$), while 6% ($n = 148$) were published as abstracts. More than half of the studies tested interventions in mice (55%, $n = 1347$), 27% ($n = 676$) used rats and a small proportion utilized other rodents such as hamsters (1%, $n = 32$), guinea pigs (<1%, $n = 19$) and groundhogs (<1%, $n = 3$). Six per cent of studies ($n = 137$) were conducted in

rabbit, pig and dog models. One-hundred and twenty-three studies tested interventions in other animal species such as zebrafish, sheep and ferrets. Although non-human primates were used in only 2% ($n = 55$) of studies, they accounted for nearly a third of studies involving neurological drugs.

The mean number of efficacy studies that were accessible in biomedical databases for each drug was 52 (SD = 96). However, we were unable to identify any published studies for three drugs (6%). Cancer drugs had the largest average volume of studies (76 per agent, SD = 127). The volume of published studies was related to how far the agent advanced in clinical development (Spearman's rank-order correlation coefficient = 0.66, $P < 0.0001$). Interventions that stalled in phase 1 testing had a mean of seven efficacy studies (SD = 4) whereas drugs that progressed to Food and Drug Administration (FDA) approval had a mean of 161 studies per agent (SD = 145).

Relationship of animal efficacy studies to trials

Presumably, all trials were preceded by efficacy studies in animals. We next sought to compare the timing of publication for efficacy studies against published trials. The vast majority of efficacy studies were published after initial trial reports. For instance, a mean of five efficacy studies (SD = 6) were published *before* the first trial, whereas a mean of 36 studies (SD = 80) were published at least 5 years *after* the first trial. The timing of publication for the first animal efficacy study relative to the first published trial is presented in Figure 3. The first published animal efficacy study preceded the first published trial by a mean of 3 years (SD = 5).

To probe this relationship further, we examined timing of publication for efficacy studies relative to trials matched for disease indication. Many efficacy studies are not aimed at supporting trials, but are rather aimed at exploring novel indications or contained within basic science studies (indeed, that 87% of efficacy studies were published after the first trial suggests that novel drugs become useful research reagents after a drug is launched in clinical development). We therefore determined the accessibility and timing of efficacy studies that matched the first and second disease indications

Table 1

Clinical characteristics of the random sample of drugs included in our analysis ($n = 47$)

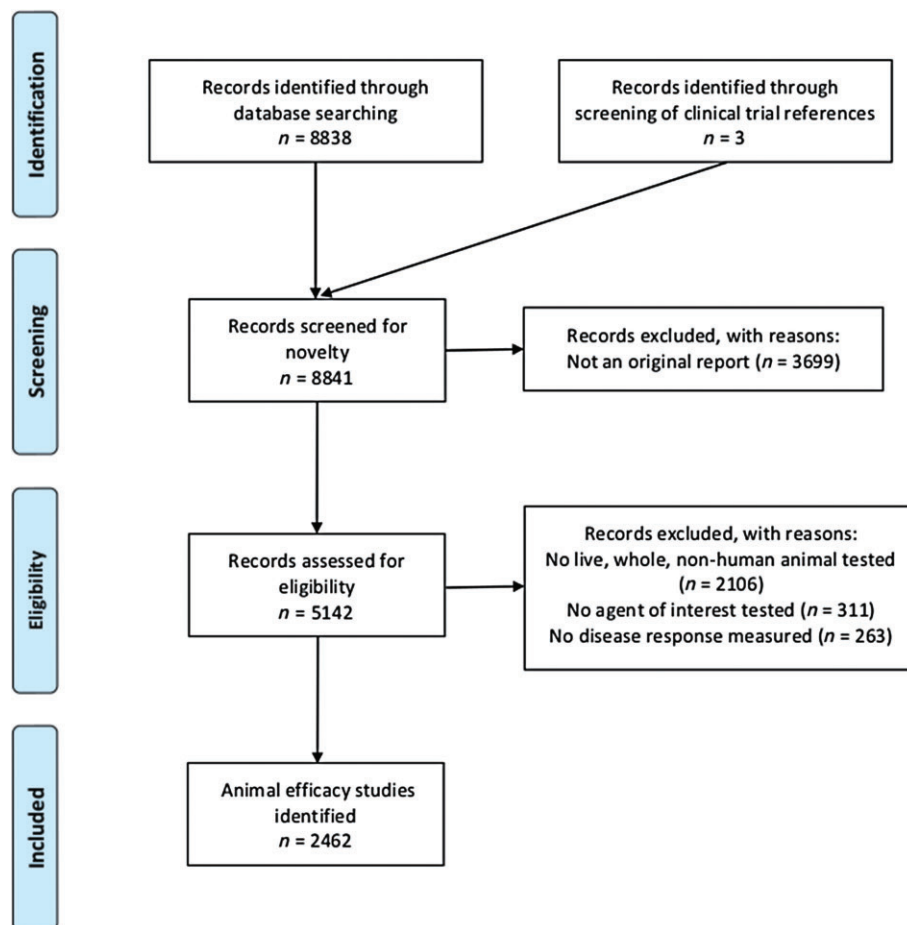| Indication	*n*	Furthest stage of development (%)		
		Phase 2	Phase 3	FDA approval (%)
Cancer	19	9 (47%)	9 (47%)	5 (26%)
Cardiovascular	9	3 (33%)	5 (56%)	2 (22%)
Infection	9	2 (22%)	7 (78%)	3 (33%)
Neurological	4	0	2 (50%)	1 (25%)
Sepsis	3	3 (100%)	0	0
Pain	2	0	2 (100%)	0
Diabetes	1	1 (100%)	0	0
Total	47	18 (38%)	25 (53%)	11 (23%)

Figure 1
PRISMA (Preferred Reporting Items of Systematic Reviews and Meta-Analyses) flow diagram describing database searches and eligibility screening for preclinical efficacy studies. Sample size at the identification stage reflects the output of de-duplicated searches of MEDLINE, Embase and BIOSIS databases.
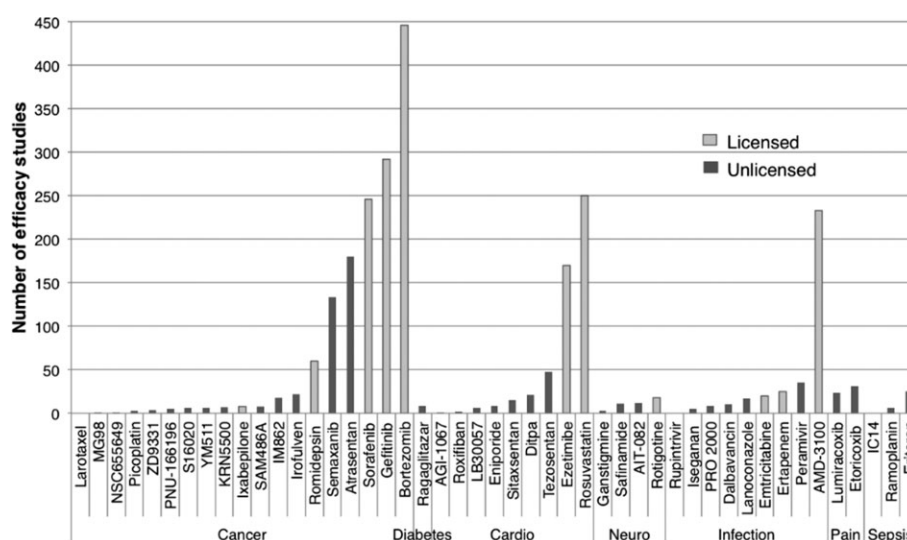

Figure 2
Number of preclinical efficacy studies per agent, by clinical indication and licensure status ($n = 47$).
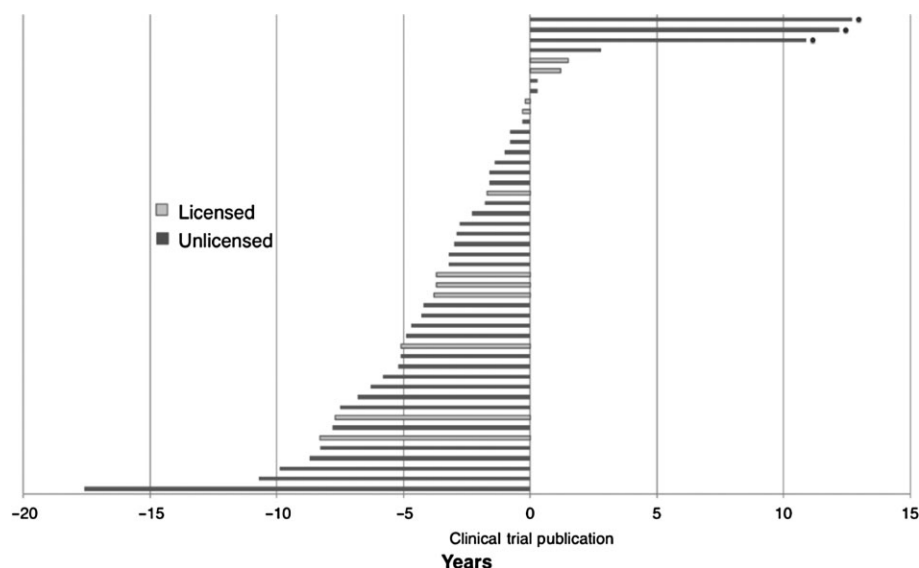
Figure 3

Lag between publication of first animal efficacy study (of any disease indication) and initial clinical trial for each agent in our cohort. Lag represents the amount of time that elapsed between the first publication demonstrating efficacy in animals, regardless of disease indication, and the first reported clinical trial (presented in Supporting Information Table S2). Bars extending to the left of the X-axis origin represent instances where efficacy studies were published *before* the first published trial; bars extending to the right represent instances where efficacy studies were published *after*. In three cases, we were unable to find *any* animal efficacy studies. When this occurred, we defined lag as time between when trial was published and when we performed literature searches. These studies are marked with a '•'.
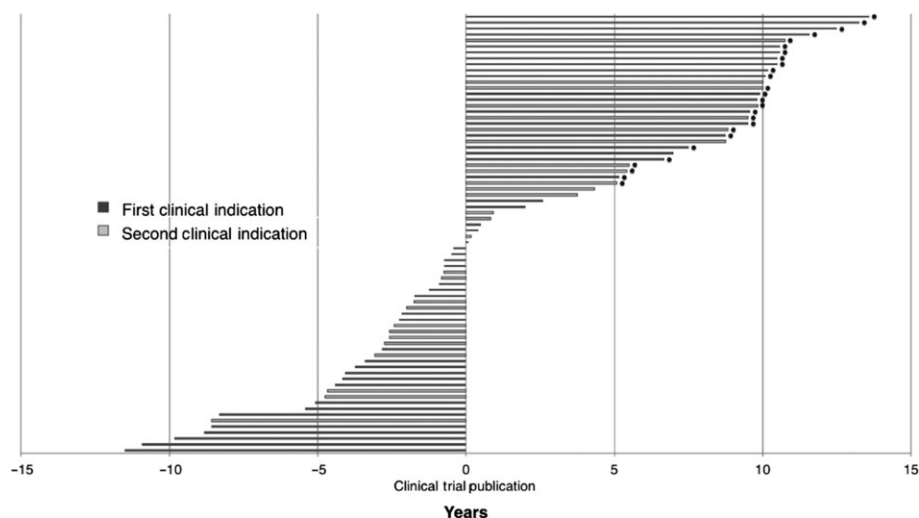

Figure 4

Lag between indication-matched efficacy study publication and clinical study publication, by first and second clinical indication. Lag represents the amount of time that elapsed between the publication of the first indication-matched preclinical efficacy study and the reported clinical trial for both first and second clinical indications (presented in Supporting Information Table S4). Bars extending to the left of the X-axis origin represent instances where efficacy studies were published *before* the first published trial; bars extending to the right represent instances where efficacy studies were published *after*. In 26 cases, we were unable to find *any* matched indication animal efficacy studies. When this occurred, we defined lag as time between when trial was published and when we performed literature searches. These studies are marked with a '•'.

tested in published trials. As indicated in Figure 4, efficacy studies tended to be published after publication of the first-matched trial, if at all. For 36% (*n* = 17) of the drugs in our sample, we were unable to access any matching efficacy studies for the first clinical indication tested; of the 27 drugs that were tested in a second clinical indication, we were unable to access matching efficacy studies for 67% (*n* = 9) of them. In total, indication-matched efficacy studies were

inaccessible for 35% (*n* = 26) of the interventions in our sample. Of the 47 trials in our analysis that named an industrial sponsor in the publication, only 40% (*n* = 19) had indication-matched efficacy studies supported by the same sponsor. Using the generous assumption that studies would be published by mid-2013 (after we completed our searching), efficacy studies were published an average of 2 years after publication of an indication-matched trial (SD = 7). Publication of matched efficacy studies for FDA-approved agents tended to be more forthcoming (6 months vs. 2.6 years for unlicensed agents).

Discussion

Efficacy studies provide the evidentiary and moral basis for exposing volunteers to unproven substances in trials. Efficacy studies also promote a favourable risk/benefit balance in trials by supporting the interpretation of trial findings. For example, re-analysis of preclinical evidence has been used to address uncertainties about the application of the cancer drug cetuximab in patients whose tumours contain uncommon mutant alleles of KRAS, a member of the Ras family of GTPases that may affect response to cetuximab (Woo *et al.*, 2013).

In this study, we sought to determine the accessibility of efficacy studies supporting the clinical development for 50 novel drugs. Our findings belie the common perception that efficacy studies stop after a drug is advanced into clinical development. Many commentators urge the systematic review of animal efficacy studies (Piper *et al.*, 1996; Sandercock and Roberts, 2002; Ludolph *et al.*, 2010), and our findings suggest that systematic review may be feasible in principle. Nevertheless, efficacy studies are scarce for drugs failing in early development: of the 22 drugs in our sample that only reached phase 1 or phase 2 testing, we were unable to access any published efficacy studies for two (9%), and could access fewer than 10 for 77% (*n* = 17) of agents. We note that many guidelines for preclinical testing encourage at least one attempt to independently replicate preclinical findings in an adequately powered sample, and at least one attempt to quasi-replicate efficacy using a different model (O'Collins *et al.*, 2006; Henderson *et al.*, 2013). Together, this would suggest that at least three indication-matched efficacy studies should be performed before clinical development, although the precise number and extent of testing will depend on trial particulars. In our sample, 53% (*n* = 25) reported three or more preclinical studies (of any indication) prior to the first published trial.

We also discovered that efficacy studies are frequently published *after* clinical testing. For more than a third of the agents in our sample (36%, *n* = 17), there were fewer than two published efficacy studies before the first clinical trials were published, and for 17% (*n* = 8) of drugs, no efficacy studies were published prior to trial publication. When matched for clinical indication used in trials, efficacy studies were even more inaccessible. For 18% of trials (*n* = 13), efficacy studies in the same indication were published after trials were published; for another 35% (*n* = 26), no indication-matched animal efficacy studies were accessible. That industrial sponsorship was not identical for many animal efficacy trial dyads

leads us to speculate that the efficacy evidence actually driving clinical development goes unpublished.

This pattern of non- and delayed publication might have two explanations. First, research teams might not be performing efficacy studies until after trials are initiated and/or published. Although this would seem surprising and inconsistent with ethics policies, FDA regulations do not emphasize review of animal efficacy data when approving conduct of phase 1 trials. For instance, FDA guidelines state that 'summary reports . . . without individual study results . . . usually suffices [for an IND application] . . . lack of . . . potential effectiveness information should not generally be a reason for a Phase 1 IND to be placed on clinical hold' (FDA, 1995, p. 11). We are aware of several instances where drugs seem to have been advanced into trials before animal efficacy studies were performed (Horn *et al.*, 2001; O'Collins *et al.*, 2006; Wheble *et al.*, 2008).

Another explanation is that drug developers precede trials with animal studies, but withhold them or publish only after trials are complete. This hypothesis is consistent with the reports of publication bias and withholding in preclinical research (Dyson and Singer, 2009; Sena *et al.*, 2010; van der Worp *et al.*, 2010). This interpretation also raises concerns, as delay of publication circumvents mechanisms such as peer review and replication, that promote systematic and valid risk/benefit assessment for trials (Doggrell, 2008).

Under the current system, there is little incentive for private drug developers to publish preclinical studies. For one, publication entails costs and does not directly serve market objectives, especially when the licensing of a drug is uncertain. Preclinical efficacy studies also contain information that may be of relevance to competitors and companies are understandably reluctant to compromise their strategic edge. Because preclinical studies do not directly inform clinical practice, the case for obligatory publication of preclinical efficacy studies is weaker than for trials. We nevertheless believe that current non-disclosure of preclinical efficacy findings is ethically problematic. First, clinical investigators are ethically obligated to ensure a favourable risk–benefit balance. Non- and delayed publication frustrates quality control measures, such as peer review, through which evidence supporting clinical testing is vetted. And as noted above, guidance documents leave uncertain the degree to which regulatory authorities access and review efficacy studies. Second, enormous public and private resources are committed to researching interventions that are never vindicated in trials; publication helps ensure the knowledge acquired from these investments can be integrated into a broader body of scientific knowledge. Last, we believe there are important ways that preclinical studies do guide clinical decision making. Many bedside decisions are informed by theories of pharmacology and pathophysiology, and preclinical efficacy studies – including those of unsuccessful drugs – often enrich these theories. Private drug developers are likely to bristle at our analysis, but we suggest that, collectively, they too have much at stake in reducing waste in clinical translation. We believe policy makers and the research community should commit greater effort to devising solutions that unlock the social utility of preclinical efficacy publication without unduly antagonizing commercial imperatives.

Our study has several limitations. First, it was aimed at measuring the public accessibility of animal evidence studies, not the total volume, and we cannot exclude the possibility that many more efficacy studies matched to trial indications are performed, and reviewed, prior to trials. Second, matching trial disease indications to those for animal studies involves judgment, and others might have created animal/clinical dyads differently. Moreover, animal models may not exist for some diseases investigated in trials. We invite readers to judge our matching – and to judge whether models existed for indications we were unable to match – by reviewing Supporting Information Table S4. Third, for several drugs, less than 6 years elapsed between the second indication trial and our search for preclinical studies; it is possible that, with more time, preclinical studies supporting these agents might yet be published. Last, although we designed a state-of-the-art search for preclinical studies, we did not contact drug companies to confirm non-publication of efficacy studies.

In conclusion, we find that the volume of published efficacy studies associated with new drugs is substantial, especially for licensed drugs. However, few studies are likely to be directly relevant for assessing risk/benefit or interpreting outcomes for early trials. Animal efficacy studies supporting specific trials are often published long after the trial itself is published, if at all. Whether because relevant efficacy studies are not performed, or because they are withheld from publication, this represents a threat to human protections, animal ethics and scientific integrity. Animal care committees, ethics review boards and biomedical journals should take measures to correct these practices.

Acknowledgements

Author contributions

J. K. and D. A. F. conceived and designed the study. C. A. F. and B. C. performed the data collection. C. A. F. and J. K. analysed the data. C. A. F. and J. K. interpreted the data and wrote the paper. D. A. F. provided statistical guidance and edited the manuscript.

Conflict of interest

The authors report no conflict of interest.

References

Bath PM, Gray LJ, Bath AJ, Buchan A, Miyata T, Green AR (2009). Effects of NXY-059 in experimental stroke: an individual animal meta-analysis. Br J Pharmacol 157: 1157–1171.

Bolton E, Wang Y, Thiessen P, Bryant S (2008). PubChem: integrated platform of small molecules and biological activities. Annu Rep Comput Chem 4: 217–241.

Doggrell SA (2008). The failure of torcetrapib: is there a case for independent preclinical and clinical testing? Expert Opin Pharmacother 9: 875–878.

Dyson A, Singer M (2009). Animal models of sepsis: why does preclinical efficacy fail to translate to the clinical setting? Crit Care Med 37: S30–S37.

FDA (1995). Guidance for industry: content and format of investigational new drug applications (INDs) for phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products.

Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG (2013). Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. PLoS Med 10: e1001489.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M (2010). Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Lab Anim 44: 170–175.

Horn J, de Haan RJ, Vermeulen M, Luiten PG, Limburg M (2001). Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. Stroke 32: 2433–2438.

Kimmelman J, Anderson JA (2012). Should preclinical studies be registered? Nat Biotechnol 30: 488–489.

Kola I, Landis J (2004). Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3: 711–715.

Ludolph AC, Bendotti C, Blaugrund E, Chio A, Greensmith L, Loeffler JP *et al*. (2010). Guidelines for preclinical animal research in ALS/MND: a consensus meeting. Amyotroph Lateral Scler 11: 38–45.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39: 2824–2829.

O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW (2006). 1,026 experimental treatments in acute stroke. Ann Neurol 59: 467–477.

Piper RD, Cook DJ, Bone RC, Sibbald WJ (1996). Introducing Critical Appraisal to studies of animal models investigating novel therapies in sepsis. Crit Care Med 24: 2059–2070.

Prinz F, Schlange T, Asadullah K (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10: 712.

Sandercock P, Roberts I (2002). Systematic reviews of animal experiments. Lancet 360: 586.

Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR (2010). Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol 8: e1000344.

Steward O, Popovich PG, Dietrich WD, Kleitman N (2012). Replication and reproducibility in spinal cord injury research. Exp Neurol 233: 597–605.

de Vries RB, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M (2011). A search filter for increasing the retrieval of animal studies in Embase. Lab Anim 45: 268–270.

Wheble PC, Sena ES, Macleod MR (2008). A systematic review and meta-analysis of the efficacy of piracetam and piracetam-like compounds in experimental stroke. Cerebrovasc Dis 25: 5–11.

Woo J, Palmisiano N, Tester W, Leighton JC Jr (2013). Controversies in antiepidermal growth factor receptor therapy in metastatic colorectal cancer. Cancer 119: 1941–1950.

van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V et al. (2010). Can animal models of disease reliably inform human studies? PLoS Med 7: e1000245.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

http://dx.doi.org/10.1111/bph.12771

Table S1 Search strategy for identifying drugs entering clinical development.
Table S2 Initial clinical trials.
Table S3 Search strategy for identifying animal efficacy studies.
Table S4 Clinical/preclinical dyads.
Note S1 Rules used to create clinical/preclinical dyads.