Investigator brochures for phase I/II trials lack information on the robustness of preclinical safety studies

Sören Sievers1 | Susanne Wieschowski1 | Daniel Strech1,2,3

1Hannover Medical School, Institute for Ethics, History, and Philosophy of Medicine, Hannover, 30625, Germany
2QUEST – Center, Berlin Institute of Health (BIH), Berlin, 10178, Germany
3Charité Universitätsmedizin Berlin, Berlin, 10117, Germany

Correspondence
Sören Sievers, Hannover Medical School, Institute for Ethics, History, and Philosophy of Medicine, Carl-Neuberg Str. 1, 30625, Hannover, Germany.
Email: sievers.soeren@mh-hannover.de

Funding Information
German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), Grant/Award Numbers: STR 1070/4-1, Project Number 2605250811 German Federal Ministry of Education and Research, (BMBF 01GP1611A, https://www.bmbf.de/).

[Aim: Meaningful and ethical phase I/II trials can only be conducted with supportive prospective risk-benefit assessment. This relies largely on preclinical animal studies addressing the safety and efficacy of treatments. These studies are reported in an Investigator’s Brochure (IB) to inform ethics review boards and regulatory authorities. Our study investigated the extent, reporting quality and accessibility of preclinical safety studies (PCSSs) compiled in IBs.

Methods: We analysed a sample of 46 IBs for phase I/II trials approved at a leading German university medical centre from 2010 to 2016. We extracted all PCSSs presented in the 46 IBs and assessed them for reporting on methodological measures to reduce validity threats.

Results: The 46 IBs included 777 PCSSs. Blinded outcome assessment, randomization and sample size calculation were reported for fewer than 1% of studies. Only 5% of the PCSSs provided a reference to published data. Compliance with Good Laboratory Practice (GLP) guidance was reported for 52% of PCSSs, but the GLP document itself does not include any relevant methodological requirements for the reduction of validity threats.

Conclusion: Scarce reporting in IBs and the very limited publicly available data on PCSSs make it almost impossible for investigators to critically evaluate the robustness of preclinical evidence of drug safety. Combined with recent findings on the presentation of preclinical efficacy studies in IBs, we conclude that the current reporting patterns in IBs strongly limit the independent review of evidential support for early human trials. Regulatory authorities and IRBs should require better reporting in IBs.

KEYWORDS
bioethics, drug development, drug safety, research ethics, translational research

1 | INTRODUCTION

Early-phase clinical research provides important knowledge gain to inform late-phase clinical trials. This knowledge gain not only requires monetary and personal resources, but also poses burdens and sometimes serious risks to participating subjects. Hence, it is not only an economic but also an immediate ethical obligation for Institutional Review Boards and regulatory bodies to assess in advance whether the planned clinical study shows a favourable risk-benefit ratio. Because research experiments inevitably entail uncertainty, this ratio can only be estimated.

Preclinical safety studies (PCSSs) primarily inform risk assessments. A closer look at these preclinical safety data is worthwhile because adverse drug reactions that were not anticipated by PCSSs make safety
issues the second most common cause for failure in clinical drug development (28% of failures), subsequently decelerating drug development and increasing prices.\textsuperscript{1,2} Additionally, preclinically undetected risks can harm healthy trial participants, with 1% experiencing severe adverse events\textsuperscript{3} or hospitalization and 6.7% experiencing serious adverse drug reactions.\textsuperscript{4–6} Reviews and umbrella studies of meta-data for translational success in studies of adverse events and toxicology show very limited predictive value of negative results in animals but a higher likelihood of observing toxicity in humans if it has already been detected in animal models.\textsuperscript{7–10} However, concordance rates showed high variability, with better prediction in certain species, toxicity fields or organs.\textsuperscript{11}

Over the past decade, manifold concerns have been raised about the reliability of the preclinical body of evidence.\textsuperscript{12,13} Meta-research studies found sparse reporting of essential measures to reduce bias on the individual study level, such as a priori sample size calculation, blinding of outcome assessment, randomization and data handling.\textsuperscript{14–17} These studies also show that only a fraction of conducted animal studies is published, which overestimates effect sizes and distorts perception of the chances of success.\textsuperscript{18–20}

High failure rates during clinical drug development are frequently linked to the abovementioned observations.\textsuperscript{21–23} However, knowledge about the amount, quality and accessibility of evidence used to support clinical trial applications is scarce, although highly relevant to examine the process of risk-benefit assessment. Anecdotal evidence suggests that pharma companies do not rely on what is believed to be poor-quality academic research and apply more rigorous methods in preclinical studies to safeguard their commercial interests.\textsuperscript{21,24,25}

Investigator brochures (IBs) aim to inform principal investigators, regulatory agencies, IRBs, and data safety and monitoring boards of the evidence that has been obtained on an investigational product (IP) for the purpose of risk-benefit assessment. To provide guidance for industry, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) adopted the Guideline for Good Clinical Practice E6 including the requirements for display of evidence: “The information should be presented in a concise, simple, objective, balanced, and non-promotional form that enables a clinician, or potential investigator, to understand it and make his/her own unbiased risk-benefit assessment of the appropriateness of the proposed trial.” An IB’s section on nonclinical studies is required to contain a summary of all relevant studies, including toxicological studies, which “should address the methodology used [...].”\textsuperscript{26}

A recent study conducted by this working group found poor reporting of measures to diminish threats to validity in preclinical efficacy studies in both IBs and referenced publications.\textsuperscript{27} This study also indicated a selection bias in IBs towards preclinical efficacy studies that demonstrate the desired outcomes. The Good Laboratory Practice (GLP) guidance document does not require reducing validity threats through randomization, sample size calculation or blinding. It requires reporting of some items in the study plan, which can be requested by an IRB when in doubt.\textsuperscript{28} Although the ICH provides further guidance specifically for the conduct of PCSSs in their work products S1 to S12, it also does not explicitly require randomization, blinding or sample size calculation.

**What is already known about this subject**

- Investigators are required to perform a risk-benefit assessment for planned early phase clinical trials based on preclinical evidence compiled in an Investigator’s Brochure (IB).
- A recent study showed that the reporting of preclinical efficacy studies in IBs does not allow readers to critically appraise this type of supporting evidence for the potential benefit of an intervention.

**What this study adds**

- This study showed that also for preclinical safety studies the reporting of IBs does also not allow investigators, Institutional Review Boards or regulatory authorities to critically appraise the validity of these safety studies.
- Blinded outcome assessment, randomization and sample size calculation were reported for fewer than 1% of studies.

The primary objective of this study is to investigate the extent, reporting quality and accessibility of PCSSs that are presented via IBs to undergo ethical review for early-phase clinical trials.

**2 | METHODS**

**2.1 | Sampling of IBs**

Access to material for clinical trial applications is difficult to obtain since IBs and study protocols are treated with a high level of confidentiality by academic medical centres and their IRBs.\textsuperscript{29,30} In a recent paper by our working group, we described how we obtained a sample of IBs representing the complete range of all $n = 97$ phase I/II trials approved by the local IRB of a leading German university hospital between the years 2010 and 2016.\textsuperscript{27} For feasibility reasons further described in the discussion, we narrowed the 97 IBs to a stratified subsample of 46 IBs consisting of (a) all IBs of phase I trials and all IBs showing characteristics of both phase I and II ($n = 20$), (b) a random subsample of five IBs from the largest companies based on global sales\textsuperscript{31} and (c) another random subsample of 21 IBs (Figure 1).

**2.2 | Data protection**

Confidentiality statements were signed by all members of the research team. The IB documents were stored on secure institutional servers. The results are presented in an aggregate manner and do not allow the identification of IPs, sponsors, investigators or other commercially sensitive information. To designate the University...
Medical Center in which the IBs have been submitted could lead to
de-anonymization and is therefore not possible.

2.3 Selection and rating of PCSSs

For selection of PCSSs from IBs, we employed the following inclusion
criteria: (a) conduct in nonhuman animals and (b) relevant to inter-
preting the safety of the IP (eg, identifying, characterizing and quanti-
fying undesirable effects or safety hazards). We excluded preclinical
studies if they were (c) pharmacokinetic studies only, (d) efficacy
studies or (e) in vitro/ex vivo studies.

We analysed the included PCSSs based on a matrix of nine items
that represent practices to reduce the risk of bias in preclinical studies.
Practices were grouped under three types of validity threats regarding
(a) internal validity, (b) construct validity and (c) external validity. The
development of the matrix referred to our recent study27 and
acknowledged a review of methodological criteria for safety studies.32
Use of randomization or choosing a meaningful sample size represent
practices that can be rated at the individual PCSS level. The reporting
on replication studies or dose-response relationships applies to a
cluster of evidence from more than one PCSS and was therefore rated
at the IB level.

We further evaluated whether an IB reported compliance with
GLP for individual PCSSs.

Furthermore, assessing the safety and toxicity of an IP is a multi-
faceted enterprise compared to the rather specific evaluation of
efficacy.33,34 To determine how the IBs covered the components of
toxicity testing we rated, at the level of IBs, whether they reported
studies of “single dose”, “repeated dose”, “carcinogenicity”, “reproduc-
tive toxicity” and “genotoxicity (mutagenicity)”. The rating of PCSS reporting in IBs was conducted by S.S. for an
initial sample of five IBs. Findings were discussed with the working
group, and the items to rate were clarified and noted in a codebook.

After the analysis of 46 IBs by S.S., a random subsample of five was
assigned to a coauthor, who independently rated the reporting to check
for disagreement. Discordance in rating of items ranged from 0% to
17% per item, with an overall inter-rater reliability of 92.5%. Unclear
cases were further discussed and were resolved by consensus.

If IBs cited safety-relevant preclinical studies that (a) matched the
inclusion criteria and (b) were published in peer-reviewed journals or
were otherwise publicly available, we located them and rated them
through the same matrix.

Descriptive statistics were applied. The complete set of rated
items is available online as Supporting Information.

3 RESULTS

3.1 Characteristics of IBs and PCSSs

We analysed 46 IBs in total, which included n = 12 IBs for phase I,
n = 8 IBs for trials that had characteristics of phases I and II, and n = 26
IBs for phase II clinical trials. While the majority of IBs (91%) mentioned
at least some clinical evidence for the IP, our sample included four stud-
ies (9%) that were “first-in-human”. Information about funding sources
showed private funding for all trials. These comprised 25 IBs (54%)
submitted by 14 of the 25 largest pharma companies based on global
sales.31 The IBs represented studies from nine out of 12 therapeutic
areas characterized by the European Medicines Agency (EMA). Figure 2
presents data for the study phases, initiators and therapeutic areas.

The 46 IBs included a total of n = 777 PCSSs with a median of 17
PCSSs per IB (range = 1-40 PCSSs per IB). Figure 3 presents the num-
ber of reported PCSSs for each IB.

Compliance with GLP was explicitly reported for 52% (n = 407) of
all PCSSs, while 24% (n = 183) explicitly stated noncompliance with
**FIGURE 2** IB sample characteristics

![IB sample characteristics chart](image)

**FIGURE 3** Number of reported PCSSs per IB

![Number of reported PCSSs chart](image)
GLP. For another 24% (n = 186) of all PCSSs, we did not find any statement on whether the study adhered to GLP.

### 3.2 Reporting on practices used to address validity threats

Table 1 shows the degree to which essential measures to reduce validity threats were reported for the 777 PCSSs. Of all PCSSs, 49% (n = 384 studies) reported the sample size, which had a median of 32 animals per experiment, including multiple groups. A calculation or explanation of the sample size was never given. Neither did any IB report blinded outcome assessment for PCSSs. Only 8% (n = 62) of the PCSSs declared the exclusion of data and <1% (n = 3) mentioned randomization.

Among the most frequently reported items were the employed animal species in 98% (n = 759), application route in 89% (n = 692) and dosing scheme in 69% (n = 532) of PCSSs.

Baseline characterization of animals was reported in 43% (n = 334), where the majority mentioned only the sex of the animal model (33% of IBs, n = 256). The outcome choice was mentioned implicitly or explicitly in 85% of PCSSs (n = 658).

A dose-limiting toxicity that could be linked to the IP was found in 19% of PCSSs (n = 150), while the remaining 81% did not show such a toxicity or safety issue that clearly limits the applicable dose. Dose-limiting toxicity was defined here as mortality or genotoxic or fetotoxic effects in one of the treatment groups or a statement of the authors that one or more effects of the tested treatment precluded such a toxicity or safety issue that clearly limits the applicable dose.

Reporting about methods to increase validity of clinical inference, which rather apply to an ensemble of studies, had to be rated on the level of the 46 IBs each, as described in the Methods section. Only 7% of the IBs (n = 3) included a report of at least one study where the age of the animal model was matched to that of patients (Table 2). Replication of an experiment concerning the safety of an IP with the same model and application route was mentioned in 9% of IBs (n = 4). Another application route to test safety was used in 61% of the IBs (n = 28), and the species of employed animal models was two or more in 96% of IBs (n = 44).

Tables 1 and 2 present the reporting of validity items also for the subgroup of first-in-human studies, where randomization was reported for 6% of studies. In the Supporting Information, further subgroups are presented, such as "IBs of phase I and phase I/II" and "IBs submitted by the largest pharma companies". The data show similar reporting patterns for the most essential methodological items randomization, blinding and sample size calculation, which were reported for less than 1% of studies.

### 3.3 References to data outside IBs

From n = 777 PCSSs reported, 52% included a reference. The majority (47%) of references pointed to in-house data that were not publicly available. In four IBs (9% of all IBs), we found references for 37 PCSSs (5% of all PCSSs) pointing to peer-reviewed publications. If such references were available, the IBs’ reports about the respective PCSSs were notably scarcer than average for essential information to judge the PCSS’s methodological quality, such as the sample size (0% for PCSSs with peer-reviewed publication vs 49% for PCSSs without such a publication), baseline characterization (8% vs 43%), control group (3% vs 51%) and application route (27% vs 81%). In turn, the respective peer-reviewed literature reported more often on baseline characterization and controls for PCSSs than the IBs themselves. Essential quality information, however, was also not reported in the peer-reviewed literature, such as sample size calculation (0% reported in peer-reviewed literature), randomization (24%) and blinding (3%). For details, see "Reporting within peer-reviewed literature for PCSSs cited in IBs" and "Reporting including external literature" in Table 1.

### 4 DISCUSSION

In this study, we assessed the reporting of 777 PCSSs included in 46 IBs for phase I/II trials. We found that the vast majority of IBs do not report on measures to reduce validity threats in the PCSSs. The absence of reporting for important measures such as sample size calculation (0%), randomization (<1% of studies) and blinding (0% of studies) makes it almost impossible to critically appraise the robustness of the PCSSs included in IBs. The lack of essential information on methodology such as randomization and blinding also stands in contrast to the Good Clinical Practice (ICH E6) guidelines requirement that the summary information on PCSSs “should address the methodology used”.

While some of the criteria used to determine the risk of validity threats only rely on expert judgement, the neglect of randomization and blinding is empirically proven to introduce systematic error and influence effect sizes. In particular, allocation concealment and randomization protect against selection bias, and blinding of outcome assessors and investigators reduces detection and performance bias. Insufficient sample sizes do not themselves produce systematic but random error. With small sample sizes, multiple replications of the same study yield toxic results with a greater variety. However, this imprecision can turn into a systematic error when publication bias joins through selective reporting of studies. Recent studies have demonstrated that the results of completed animal studies are often not disseminated. In the case of animal efficacy studies, this lack of information can result in overstatements of efficacy. For preclinical safety studies, such investigations on the extent of selective publication and its practical impact are not yet available. Since there is no obligation for prospective and transparent registration of animal studies, the regulatory authorities and IRBs have no opportunity to check whether the PCSSs reported in IBs are subject to selection bias or not. The first animal study registries were launched recently and this topic is receiving increased attention.

Because 95% of PCSSs presented in IBs do not provide a reference to a published report, there is no way for reviewers to access
**TABLE 1** Reporting on internal and construct validity items on the preclinical safety study (PCSS) level

| Validity items: PCSS level | Description | All IBs (n = 46) | IBs for first-in-human studies (n = 4) | Subset of PCSS citing peer-reviewed literature | PCSSs in IBs combined with information from peer-reviewed literature |
|---------------------------|-------------|----------------|-------------------------------|-----------------------------------------------|--------------------------------------------------|
|                           |             | n %            | N %                          | N %                                          | n %                                              |
| Sample size<sup>a</sup>   | Is sample size reported? | 384 49.4% | 14 29.8% | 0 0% | 34 91.9% |
|                           | If yes, is sample size calculation reported? | 0 0% | 0 0% | 0 0% | 0 0% |
| Blinded assessment<sup>b</sup> | Is concealment of allocation reported? | 0 0% | 0 0% | 0 0% | 1 2.7% |
|                           | Is blinding for outcome assessment reported? | 0 0% | 0 0% | 0 0% | 1 2.7% |
| Exclusion of data from analysis<sup>a</sup> | Is exclusion of data reported? | 62 8% | 2 4.3% | 0 0% | 3 8.1% |
| Randomization<sup>b</sup>  | Is randomization reported? | 3 0.4% | 3 6.4% | 0 0% | 9 24.3% |
| Controls<sup>a</sup>       | Is a control group reported? | 404 52% | 15 31.9% | 1 2.7% | 32 86.5% |
| Model choice<sup>b</sup>   | Is the species of animals reported? | 759 97.7% | 37 78.7% | 33 89.2% | 37 100% |
|                           | Is the application route reported? | 692 89.1% | 25 53.2% | 10 27.0% | 36 97.3% |
|                           | Is the dosing scheme reported? | 532 68.5% | 19 40.4% | 12 32.4% | 36 97.3% |
| Baseline characterization<sup>b</sup> | Is baseline characterization reported? | 334 43% | 13 27.7% | 3 8.1% | 35 94.6% |
| Outcome choice<sup>b</sup> | Is the outcome choice reported? | 658 84.7% | 25 53.2% | 32 86.5% | 37 100% |
| GLP compliance             | Is compliance of the study with good laboratory practice reported? | 407 52.4% | 12 25.5% | 0 0% | 0 0% |

Abbreviations: GLP, good laboratory practice; IB, investigators brochure; NOAEL, no observed adverse effect level; PCSS, preclinical safety study.
<sup>a</sup>Internal validity item.
<sup>b</sup>Construct validity item.
The statement of GLP compliance does not inform the investigator about the study methodology. This is because the OECD GLP Principles do not include reporting obligations for the IB or specific methodological requirements to rely on but rather regulate responsibilities, the facility, the use of a study plan and the contents of a final report.

The low number of published PCSSs, 5%, is incompatible with several ethical and scientific guidelines which highlight the importance of published preclinical evidence for early-phase trial launch.\textsuperscript{28, 29} We acknowledge that there are obstacles to complying with this requirement, such as the protection of sensitive information, low incentive to publish and the potential disinterest of journals and peers for a battery of toxicity studies. It is nevertheless possible to disclose data in a repository or to publish preprints. However, analysing the reasons for nonpublication is beyond the scope of this analysis, since we focus on the question of whether the application material is sufficient to perform an independent and meaningful risk-benefit analysis.

An important problem in the discussion of translational failure is the question of whether low predictivity of human outcomes is due to flawed methodology of preclinical research or the general failure of animal models themselves because these models have never been scientifically evaluated.\textsuperscript{40} However, the deficient reporting of internal and construct validity items found in this study does not allow us to attribute failure to one cause or the other. The predictivity of animal models can only be evaluated if animal studies use and report rigorous methodology.

Combined with the results of our recent study,\textsuperscript{27} we can conclude that the current reporting in IBs on both PCSSs and preclinical efficacy studies strongly limits the external review of the evidential support for early human trials.

Our study has the following limitations. Due to the tight restrictions described above, we were not able to access a random sample of IBs\textsuperscript{29, 30} but the full sample of all 97 IBs for phase I/II trials approved by one German IRB. For the in-depth reporting analysis of all PCSS mentioned in one IB we needed to reduce the number of analysed IBs for feasibility reasons. We draw a stratified-randomized subsample of 46 IBs that altogether reported on 777 PCSS. For the following reasons we believe that our findings on the reporting of PCSS in IBs are generalizable and not specific for IBs submitted to one German IRB. Most importantly, the same IBs that we analysed were also submitted to the German regulatory agency Bundesinstitut fuer Arzneimittel und Medizinprodukte and thus adhere to standards of the EMA. The IRB that approved the analysed IBs does not have a specific focus on the steps that have been taken to increase the clinical generalizability. The IRB that approved the analysed IBs does not have a specific focus on the steps that have been taken to increase the clinical generalizability of PCSSs. The relevant information for such tables would be the employed model and outcomes as well as further items to critically appraise the robustness of the respective PCSSs, such as randomization, blinding, data handling and registry numbers.

### Table 2 Reporting on internal, construct and external validity items at the IB level

| Validity items | Description | All IBs Total = 46 IBs | IBs for first-in-human studies (n = 4) Total = 4 IBs |
|----------------|-------------|------------------------|-----------------------------------------------|
| Dose response\textsuperscript{a} | Is dose response reported? | 29 63% | 1 25% |
| | Is NOAEL reported? | 34 73.9% | 2 50% |
| Age matched to patients\textsuperscript{b} | Is the age of animals reported? | 14 30.4% | 1 25% |
| | Is the age of animals matched to that of patients? | 3 6.5% | 0 0% |
| Replication of experiment\textsuperscript{c} | Is there a replication of experiment? | 4 8.7% | 0 0% |
| Replication in different route\textsuperscript{c} | Is there replication in a different application route? | 28 60.9% | 2 50% |
| Replication in different species\textsuperscript{c} | Is there replication in a different species? | 44 95.6% | 4 100% |

Abbreviations: IB, investigators brochure; NOAEL, no observed adverse effect level.

\textsuperscript{a}Internal validity item.
\textsuperscript{b}Construct validity item.
\textsuperscript{c}External validity item.
ACKNOWLEDGEMENTS
We would like to thank the IRB for granting access to IBs. This project was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). Grant Number: STR 1070/4-1, Project Number 2605250811 and by the German Federal Ministry of Education and Research (BMBF 01GP1611A, https://www.bmbf.de/).

CONTRIBUTORS
Sören Sievers, Susanne Wieschowski, Daniel Strech: conceptualization, methodology and writing – review & editing. Susanne Wieschowski and Sören Sievers: Formal analysis and Investigation. Daniel Strech: funding acquisition, project administration, supervision and validation. Sören Sievers: writing – original draft.

COMPETING INTERESTS
There are no competing interests to declare.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in the Supporting Information of this article.

ORCID
Sören Sievers https://orcid.org/0000-0002-6462-1223
Susanne Wieschowski https://orcid.org/0000-0003-0128-2938
Daniel Strech https://orcid.org/0000-0002-9153-079X

REFERENCES
1. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. 2013.
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? 2004.
3. Emanuel EJ, Bedarida G, Macci K, Gabler NB, Rid A, Wendler D. Quantifying the risks of non-oncology phase I research in healthy volunteers: meta-analysis of phase I studies. BMJ. 2015;350:h3271. https://doi.org/10.1136/bmj.h3271
4. Hartung T. Toxicology for the twenty-first century. Nature. 2009;460 (7252):208–212. https://doi.org/10.1038/460208a
5. Hartung T. Look back in anger – what clinical studies tell us about preclinical work. ALTEX. 2013;30(3):275–291. https://doi.org/10.14573/altext.2013.3.275
6. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. 1998.
7. Bailey J, Thew M, Balls M. An analysis of the use of dogs in predicting human toxicology and drug safety. Alternative Lab Anim. 2013;41(5): 335-330.
8. Bailey J, Thew M, Balls M. An analysis of the use of animal models in predicting human toxicology and drug safety. Alternative Lab Anim. 2014;42(3):181-199.
9. Bailey J, Thew M, Balls M. Predicting human drug toxicity and safety via animal tests: can any one species predict drug toxicity in any other, and do monkeys help? Alternative Lab Anim. 2015;43(6): 393-403.
10. Clark M. Prediction of clinical risks by analysis of preclinical and clinical adverse events. J Biomed Inform. 2015;54:167-173.
11. Leenaars CHC, Kouwenaar C, Stafleu FR, et al. Animal to human translation: a systematic scoping review of reported concordance rates. J Transl Med. 2019;17:223.
12. Pusztai L, Hatzis C, Andre F, Andre F. Reproducibility of research and preclinical validation: problems and solutions. Nat Rev Clin Oncol. 2013;10(12):720-724.
13. van der Worp HB, Macleod MR. Preclinical studies of human disease: time to take methodological quality seriously. J Mol Cell Cardiol. 2011; 51(4):449-450.
14. Kilkenny C, Parsons N, Kadyrzewski E, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS ONE. 2009;4(11):e7824. https://doi.org/10.1371/journal.pone.0007824
15. van der Worp HB, Howells DW, Sena ES. Can Animal Models of Disease Reliably Inform Human Studies?. PLoS Medicine. 2010;7(3): e1000245. https://doi.org/10.1371/journal.pmed.1000245
16. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. JAMA. 2006;296(14):1731-1732.
17. Hess KR. Statistical Design Considerations in Animal Studies Published Recently in Cancer Research. Cancer Research. 2011;71(2): 625–625. https://doi.org/10.1158/0008-5472.can-10-3296
18. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol. 2010;8(3):e1000344.
19. ter Riet G, Korevaar DA, Leenaars M, et al. Publication bias in laboratory animal research: a survey on magnitude, drivers, consequences and potential solutions. PLoS ONE. 2012;7(9):e43404.
20. Wieschowski S, Biernt S, Deutsch S, et al. Publication rates in animal research. Extent and characteristics of published and non-published animal studies followed up at two German university medical centres. PLoS ONE. 2019;14(11):e0223758.
21. Osherovich L. Hedging against academic risk. Science-Business eXchange. 2011;4(15):416. https://doi.org/10.1038/scibx.2011.416
22. van der Worp HB, de Haan P, Morreema E, Kalkman CJ. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. J Neurol. 2005;252(9):1108-1114.
23. Howells DW, Sena ES, Macleod MR. Bringing rigour to translational medicine. Nat Rev Neurol. 2014;10(1):37-43.
24. Freedman S, Mullane K. The academic-industrial complex: navigating the translational and cultural divide. Drug Discov Today. 2017;22(7): 976-993.
25. Levin LA, Behar-Cohen F. The academic-industrial complexity: failure to launch. Trends Pharmacol Sci. 2017;38(12):1052-1060.
26. European Medicines Agency. Guideline for good clinical practice E6 (R2). London: 2016;53.
27. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: do they enable risk-benefit assessment? PLoS Biol. 2018;16:e2004879.
28. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG) [Medicinal Products Act, Germany] §42, Accessed November 7, 2020. https://www.gesetze-im-internet.de/amg_1976/.../42.html
29. Chan AW, Upshur R, Singh JA, Gherbi D, Chapsis F, Altman DG. Research protocols: waiving confidentiality for the greater good. BMJ. 2006;332(7549):1086-1089.
30. Cohen D. Oxford TB vaccine study calls into question selective use of animal data. BMJ. 2018;360(8136):j5845. https://doi.org/10.1136/bmj.j5845
31. Global Data. Top 25 Pharma Companies By Global Revenue: 2020. Available at: http://www.pmlive.com/top_pharma_list/global_revenues [Accessed 7 November 2020].
32. Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environ Health Perspect. 2013;121(9):985-992.
33. Ahuja V, Sharma S. Drug safety testing paradigm, current progress and future challenges: an overview. J Appl Toxicol. 2014;34(6): 576-594.
34. Stark C, Steger-Hartmann T. Nonclinical safety and toxicology. Handb Exp Pharmacol. 2016;232:261-283.

35. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirmagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke. 2008;39(10):2824-2829.

36. Bert B, Heinl C, Chmielewska J, et al. Refining animal research: the animal study registry. PLoS Biol. 2019;17(10):e3000463.

37. Baker M. Animal registries aim to reduce bias. Nature. 2019;573(7773):297-298.

38. (STAIR) STAIR. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. Stroke. 1999;30:2752-2758.

39. Daley GQ, Hyun I, Apperley JF, et al. Setting global standards for stem cell research and clinical translation: the 2016 ISSCR guidelines. Stem Cell Rep. 2016;6(6):787-797.

40. Shanks N, Greek R, Greek J. Are animal models predictive for humans?. Philos Ethics Humanit Med. 2009;4(1):2. https://doi.org/10.1186/1747-5341-4-2

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sievers S, Wieschowski S, Strech D. Investigator brochures for phase I/II trials lack information on the robustness of preclinical safety studies. Br J Clin Pharmacol. 2021;87:2723–2731. https://doi.org/10.1111/bcp.14615