

Limitations and Flaws of Preclinical Pulmonary Hypertension Studies

Tsukasa Shimauchi, MD, PhD

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

François Potus, PhD

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

Sébastien Bonnet, PhD, FAHA

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

Steeve Provencher, MD, MSc, FRCPC

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

Roxane Paulin, PhD

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

Olivier Boucherat, PhD

*Pulmonary Hypertension Research Group
Institut Universitaire de Cardiologie et de
Pneumologie de Québec Research Center
Department of Medicine
Université Laval
Québec City, Canada*

Despite advances in our understanding of the disease, significant therapeutic gaps remain for pulmonary arterial hypertension (PAH). Indeed, no cure exists yet for this devastating disease, and very few innovative therapies beyond the traditional pathways of endothelial dysfunction have reached late clinical trial phases in PAH. While there are inherent limitations to the currently available animal models of PAH, the delayed translation of innovative therapies to the clinic may also relate to flawed preclinical research methodologies. The present article discusses the limitations and flaws in the design of preclinical PH trials and discusses opportunities to create preclinical studies with improved predictive value in identifying key mechanisms involved in PAH development and progression and guiding early phase drug development in PAH patients. The implementation of rigorous study design will need support not only from researchers, peer reviewers, and editors, but also from academic institutions, funding agencies, and animal ethics authorities.

INTRODUCTION

The number of scientific publications related to pulmonary arterial hypertension (PAH) increased exponentially over the last decades, leading to significant advances in our understanding of its pathophysiology¹ and management,² allowing the delay of clinical worsening,³ and likely improving survival.⁴ However, long-term prognoses of PAH patients can be further improved.⁵⁻⁷ Significant translational and therapeutic gaps between preclinical research and improved patient outcomes thus persist, as very few innovative therapies have reached late clinical trial phases in PAH.⁸⁻¹¹ This translational gap is not unique to PAH. Since the beginning of the millennium, the number of new drugs approved yearly by health authorities has declined¹² despite marked increases in total research and development expenditures.¹³ Clinical drug development is notoriously arduous, with fewer than 5% of *high impact* basic science discoveries¹⁴ and fewer than 10% of development paths in Phase 1¹² being eventually approved by health authorities. Several reasons may explain

this phenomenon, including higher regulatory efficacy hurdles and increased complexity and cost of clinical trials.¹⁵ There are also inherent limitations to the currently available *in vitro* and animal models, which imperfectly mimic the full spectrum of the human disease.¹⁶ Moreover, it has been proposed that the failing might also be related to the study design, implementation, and analysis, ultimately weakening our confidence in preclinical studies to identify promising therapeutic targets.¹⁷ Bias in study design, analytical methods, and reporting practices may indeed compromise scientific validity¹⁸ and data reproducibility,¹⁹ and ultimately jeopardize translation to human studies. Given the limited financial resources, the persistent medical need for improved therapy in PAH, and the restricted study population available for clinical trials, there is a need for reducing the number of false positive signals in preclinical studies and for optimizing the development of innovative therapeutic targets through performance of clinical trials based on more robust experimental data.

Key Words—pulmonary arterial hypertension, pulmonary hypertension, animal models of human disease, methodological rigor, reproducibility, confirmatory study, preclinical study, statistical analysis plans

Correspondence: steeve.provencher@criucpq.ulaval.ca

Disclosure: No funding was received to write this manuscript. TS and FP have no conflicts of interest to declare. SB received consultant fees from Actelion Pharmaceuticals and research grants from Janssen. SP has received research grants from Actelion Pharmaceuticals, AstraZeneca, and Resverlogix, and has received speaker fees from Actelion Pharmaceuticals. RP is a scientist of the Heart and Stroke Foundation. OB is a scientist of the Fonds de Recherche en Santé du Québec and has received research grants from the Fondation Québécoise en Santé Respiratoire (FQSR) and Boehringer-Ingelheim.

RESEARCH BIAS

In research, bias occurs when “a systematic error is introduced into sampling or testing by selecting or encouraging one outcome or answer over others.”²⁰ Bias can cause estimates of association to be either larger or smaller than the true association; in some cases, bias can even cause a perceived association that is directly opposite of the true association. Importantly, bias is relatively independent of both study power and statistical significance in contrast to imprecision, which relates to a random error. Thus, studies may produce precise but biased results because of flaws in study design and execution. Conversely, a study may be free of significant bias but yield an incorrect effect estimate due to low statistical power. While some degree of

bias is nearly always present in a study, researchers should make every effort to identify, quantify, and/or eliminate bias through proper study design and data analysis, and to acknowledge its occurrence when unavoidable.

IDENTIFYING AND LIMITING BIAS IN IN-VITRO PAH PRECLINICAL STUDIES

The access to human samples of high quality from PAH patients and appropriate controls represents an invaluable resource for improving our understanding of PAH and validating emerging hypotheses. However, human PAH tissues and cells have been most commonly obtained at the time of lung transplantation from patients with long-standing disease that may no longer be repre-

sentative of mechanisms accounting for PAH development or progression, introducing a significant selection bias (Figure 1). More importantly, most experimental studies using human samples are performed with small sample sizes due to the scarcity of specimens. Given that PAH is a heterogeneous disease in terms of background genetic defects, concomitant diseases predisposing to PAH, as well as genetic variations influencing response to therapy,^{11,21} limited sample size may easily amplify the effects of selection bias and lead to erroneous conclusions (Figure 2A). Therefore, collaborative studies allowing the exploration of promising targets and the real interindividual heterogeneity in a larger number of samples are thus essential (Figure 2B).²² The creation

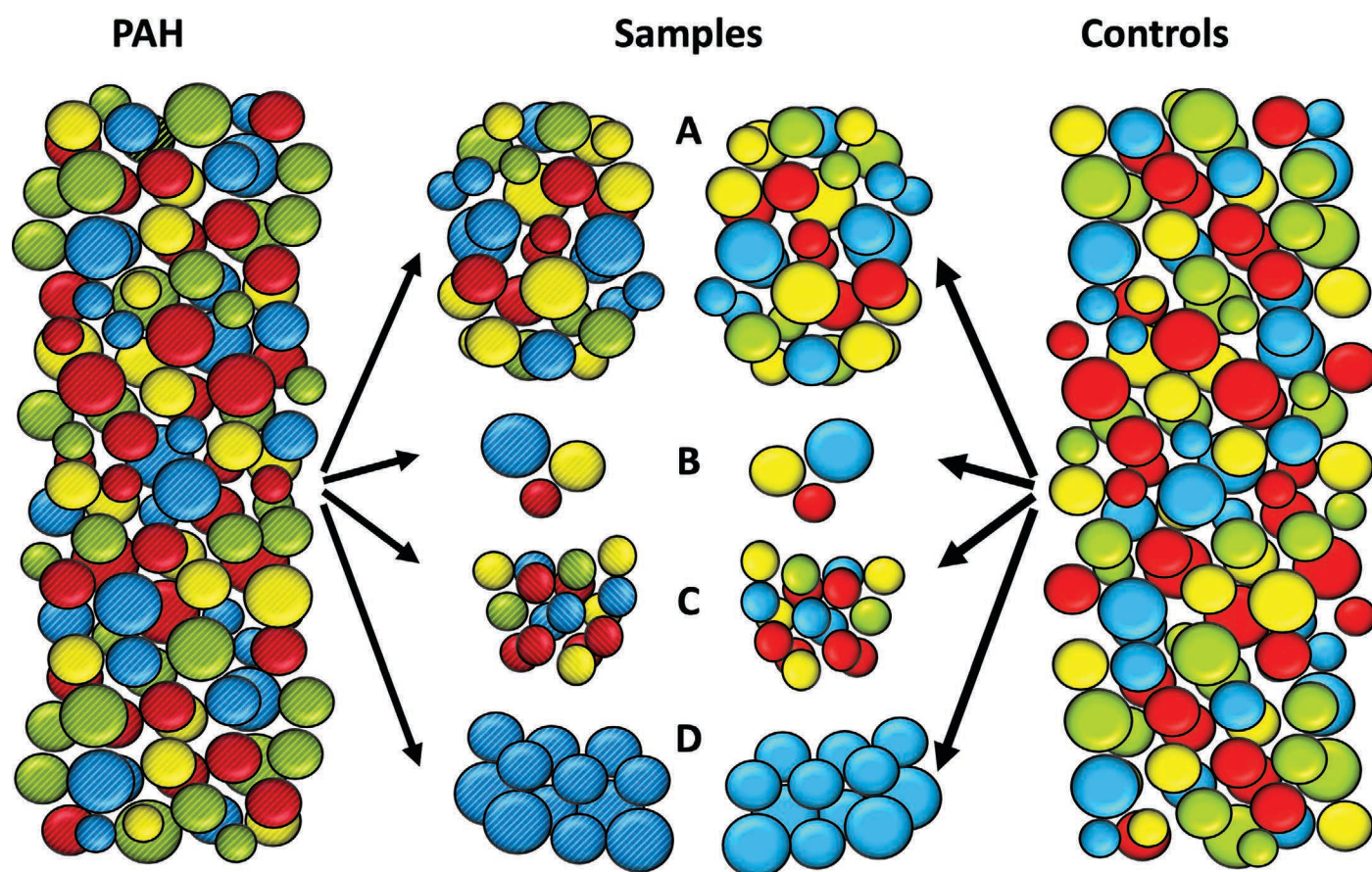


Figure 1: Selection bias in research. In research, selection bias frequently occurs when a systematic error is introduced by sampling methodologies. (A) Selecting a sample that is truly representative of both diseased subjects and matched controls is thus mandatory before observed differences are believed to be representative of the disease process, and the observed results should be considered to only apply to subjects with characteristics comparable to the study population. This is most commonly performed using a large representative and random sampling. In pulmonary arterial hypertension (PAH), human tissues and cells have been most commonly obtained at the time of lung transplantation from patients with long-standing disease that may no longer be representative of the mechanisms accounting for PAH development or progression, introducing a significant selection bias. (B) Using small sample size or (C–D) biased samples (eg, samples from only men or patients with end-stage disease) is unlikely to be representative of mechanisms accounting for PAH development or progression, or to take into account the significant heterogeneity of the disease in humans.

of a structured PAH network (eg, the Pulmonary Hypertension Breakthrough Initiative or the International Consortium for Genetic Studies in PAH) and biobank facilities dedicated to harvesting and preserving explanted lung tissues, facilitating access to human tissue, and ensuring homogeneity in tissue processing, is thus warranted since human tissues are currently underexploited in PAH experimental research.

The choice of the cells and tissues to which PAH samples are compared is also crucially important. Indeed, there are often systematic differences between the groups being compared, known as confounding, so much so that differences in signaling pathways or outcomes may result from these differences rather than actual pathobiological abnormalities. Minimizing these inherent differences is thus essential. Control samples should ideally be matched for age/sex and for their underlying disease. In many cases, PAH researchers have relied on resected lung tissue for cancer. However, special attention is required to obtain tissues sufficiently distal from the tumor that may significantly influence the phenotype and genotype of neigh-

boring cells.²³ In all cases, equivalent tissue specimens should be collected from the same organ areas. This is particularly important within the lungs as distal versus more proximal pulmonary arteries may significantly differ phenotypically. In addition, the same handling and processing has to be used. Taken together, careful selection of control tissues that most likely represent healthy lungs/tissues is crucial.

IDENTIFYING AND LIMITING BIAS IN IN-VIVO PAH PRECLINICAL STUDIES

Despite the importance of scientific results obtained from animal models, most of these studies have been hampered by the fact that these models do not entirely encompass the typical features of human PAH. This may explain why animal models are frequently considered poor predictors of whether an experimental drug can become an effective treatment. Sometimes, though, the real reason is that confirmatory preclinical studies were not rigorously designed. Accordingly, the statistical and methodological rigor should be adapted, and in many ways, confirmatory studies should resemble

clinical trials. Indeed, only the most rigorously conducted trials can completely exclude bias as an alternate explanation for the promising results observed following an intervention. Thoughtful subject eligibility criteria, sample size estimation, randomization and treatment allocation concealment, blinding, standardized outcome assessment, proper data handling, and transparent reporting methods have profoundly improved the validity of clinical trial results over the years.²⁴ Such improvements are also essential in confirmatory preclinical research.

Matching Models to Human Manifestations of PAH

Recruiting a study population representative of future patients to be treated while minimizing confounding effects is the first step of an appropriately designed prospective study. Despite the limitations of current animal models previously discussed,^{16,25} a detailed characterization and reporting of animal traits at baseline and appropriate controls using animal characteristics that are representative of the human disease should be promoted for a better standardization of the experimental design, enhanced reproducibility,

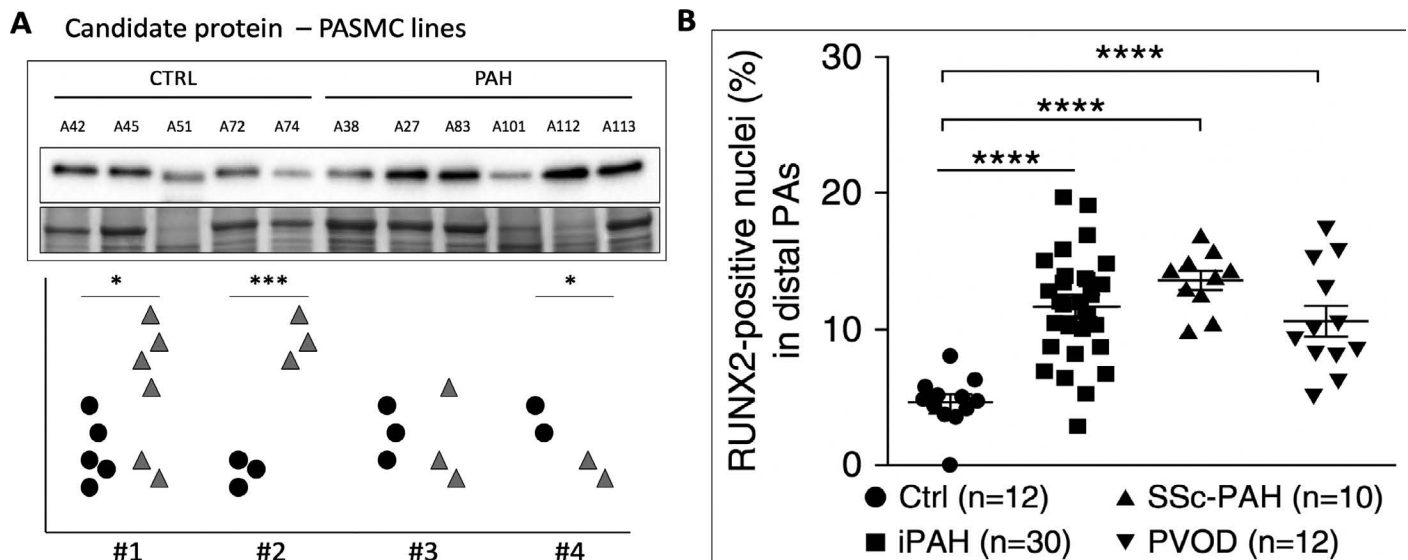


Figure 2: Influence of selection bias and sample size on the presumed validity of pulmonary arterial hypertension (PAH) preclinical results. PAH is a heterogeneous disease in terms of background genetic defects, concomitant diseases predisposing to PAH, and genetic variations influencing response to therapy. **(A)** While even large sample size may be associated with significant selection bias, this may easily be amplified with limited study samples. For example, experiment #1 suggests that the expression of a candidate protein is increased when evaluated on 5 independent cell lines. These results appear to be overestimated when the analysis was restricted to 3 selected cell lines (experiment #2), whereas no difference (experiment #3) or even a downregulation of the candidate protein (experiment #4) were observed when other cell lines were selected. **(B)** This is exemplified in a recent collaborative work to assess the role of RUNX2 in a large number of PAH samples.²² In that study, significant overlap in the proportion of RUNX2-positive cells was observed between PAH and control tissues obtained at the time of lung transplantation. Although such analysis may not be representative of the early disease processes, a smaller sample size would have increased the risk of overestimating or underestimating differences between groups.

and greater predictive ability. Currently, variations in disease induction and the potential for persistent and unrecognized confounders, including considerable inconsistencies in animals' ages and weights, how pulmonary hypertension (PH) is induced, and when the intervention is initiated and terminated,²⁶ represent important sources of bias in PH preclinical research. Care must thus be taken to prespecify eligibility criteria before animals are enrolled. In confirmatory preclinical studies, it is also reasonable to randomize animals to novel therapies when irreversible PH is expectedly fully established and following prior confirmation (eg, by echocardiography). In addition, the rationale for choosing models should be stated,^{27,28} and performing studies using more than one model and across different animal strains is encouraged. Ultimately, large animal models may share some common features of the human disease and are often the last step before translating novel drug candidates to clinical trials.²⁹ While the need to include women is now a well-established requirement in clinical trials,³⁰ analogous standards have not been equally enforced in preclinical stages of research. While the vast majority of preclinical PH studies still use male rodents only,^{26,31} inferring experimental findings to both sexes when a single sex is studied could disadvantage women by biasing our understanding of disease processes toward male-predominant patterns. This is especially problematic in PAH, where there is a significant female predominance in humans.³² The landscape of clinical trials in PAH also dramatically changed over the last decade, and future compounds will almost necessarily be tested on top of currently available therapies in clinical trials leading to drug approval.^{2,33} Although new targets can be alternatives to the currently approved therapies in humans, the demonstration of additive or synergic effects of novel therapeutic targets nowadays appears desirable for confirmatory preclinical studies.

Randomization and Allocation Concealment

The starting point for an unbiased interventional study is the use of a mechanism that ensures that the same sorts of

participants receive each intervention. Even an apparently homogeneous group of animals may have inherent differences when the intervention is introduced. Thus, processes need to be considered to allow proper balance between groups and, as for humans, random animal allocation generally minimizes bias and balances characteristics that may influence response to treatment if properly done in a large enough sample. Techniques used to implement the allocation sequence (ie, allocation concealment) are also essential to avoid selection bias being introduced by selecting animals based on the upcoming intervention assignment. There is indeed empirical evidence from preclinical research^{34,35} that either inadequate generation or concealment of allocation sequence yield to exaggerated estimates of intervention effects. Therefore, researchers should ideally report measures of successful randomization and allocation concealment.

OTHER POTENTIAL BIAS IN PAH PRECLINICAL STUDIES

Blinding of Outcome Assessment

Blinding refers to the process by which the study personnel are kept unaware of intervention allocations. Lack of blinding in clinical trials is associated with exaggerated estimates of intervention effects,³⁶ especially when the outcome of interest is subjective.³⁷ Importantly, many apparently objective outcome measures in preclinical PAH studies remain subject to interpretation. Unconscious bias can thus creep into evaluation of unblinded experiments even when performed by scientists of high integrity. Although blinding the investigator administering the treatment/intervention may not be possible in all instances, blinded assessment of imaging, hemodynamics, and histological outcomes is almost universally possible through independent team members performing outcome ascertainment.

Study Readouts and Interstudy Standardization

Even with rigorous attention to study design, studies may not have translational validity if the endpoint specified is not valid or is not measured using robust techniques. Importantly, outcome measures should match the clinical realm

using relevant measures (eg, comprehensive hemodynamics in in-vivo studies). Secondary readouts are generally used to provide supportive information (eg, to ensure hemodynamic endpoints are correlated with histological, anatomic, and biochemical findings postmortem) or exploratory, hypothesis-generating information. Obviously, the exploratory nature of some experiments makes sample size calculation impossible or meaningless. Conversely, the importance of prespecified sample size calculation, referred to as a power calculation, in confirmatory experiments cannot be overemphasized, although it is rarely performed in preclinical PH studies.^{26,31} Using human samples or exposing lab animals to research is only justifiable if there is a realistic chance that the study will yield useful information. Importantly, inappropriate samples will result in an inconclusive study, whereas an unnecessary large sample size will accrue excessive cost. Many researchers are thus tempted to perform interim analyses to subsequently increase the sample size as necessary. However, interim analyses enhance the risk of false positive results due to multiple analyses. Therefore, the primary endpoint of preclinical confirmatory PH studies must be decided before the study begins, as well as the effect size of the intervention for which the study is powered, and should be provided in the *methods* section of confirmatory experiments. Similarly, empirical work has confirmed marked heterogeneity in the methodologies used to assess study outcomes in preclinical PH studies,²⁶ including pulmonary hemodynamics, markers of right ventricular function, and pulmonary remodeling.²⁶ The majority of in-vivo studies also fail to appropriately monitor for toxicity. Obviously, these elements cannot be unilaterally dictated but require a consensus process to take place, with experts in the field agreeing on best practice, as has been previously developed for preclinical in-vivo evaluation of pharmacological active drugs in other fields.³⁸⁻⁴⁰

Multiplicity, Interim Analyses, and P Value Adjustments

Because multiple readouts are necessary to fully evaluate pathophysiological

pathways and the effects of interventions, multiple endpoints are frequently measured. However, conducting multiple tests of significance progressively increases the probability that a null hypothesis is rejected when the null hypothesis is actually true (ie, false positive result). Consideration must be given to controlling the risk of false positive conclusions, and adjustment for multiplicity will typically be necessary, especially for confirmatory studies.⁴¹⁻⁴³ Similarly, interim analyses frequently used to incorporate what is learned during the course of a study increase the risk of falsely rejecting the null hypothesis. Researchers should thus avoid unplanned interim analyses, and preliminary results should be presented without formal statistical analyses unless nominal *P* values have been adjusted accordingly. Collaboration with a statistician at the design stage and throughout analyses is thus crucial, and the selected procedure must be prespecified in the statistical analysis plan before undertaking any analyses of the data.

Handling of Missing Data

Attrition and exclusions frequently occur in preclinical PH studies when animals die or are withdrawn from the experiments or assessment does not provide relevant data. The risk of bias from incomplete outcome data depends on several factors, including the amount and distribution of missing data across intervention arms and the reasons for missing outcome data. Researchers should consider using a flow diagram showing the number of animals in intervention and control groups at each experimental step from randomization to outcome assessment. A timeline of experimentation is also desirable to inform whether all animals within each experimental group were analyzed together. For confirmatory studies, an intention-to-treat analysis may be considered as potentially the least biased way to estimate intervention effects in randomized trials.⁴⁴ However, true intention-to-treat analyses generally require imputation, which can also lead to serious biases unless conservative methods are used. Thus, where imputation is used, both the *per protocol* and the *intention-to-treat* analyses should

be presented, and the methods and assumptions for imputing data should be defined a priori and appropriately described.

Interpretation of the Results

The high pressure to find low *P* values, combined with a common misunderstanding of how to correctly interpret *P* values, frequently distorts the interpretation of *significant* results.⁴⁵ A low *P* value is considered strong evidence against the null hypothesis. However, a *P* value of .05 is frequently incorrectly interpreted as meaning that there is 95% chance that the observed difference is true, rather than indicating a 5% probability that the difference is observed even if the null hypothesis is true. Previous studies estimated that a *P* value of .05 corresponds to a false positive rate of *at least* 23% (and typically close to 50%).⁴⁶ Thus, a single statistically significant hypothesis test often provides insufficient evidence to confidently discard the null hypothesis, and study replication, especially by independent investigators, enhances the confidence that study results are true findings. Therefore, investing more time in replicating results (those of others as well as our own) and synthesizing data through systematic reviews and meta-analyses should be incentivized.^{47,48}

Reporting and Publication Bias

Reporting biases arise when the dissemination of research findings is influenced by the nature and direction of results. While publication bias (occurring when entire studies are not published, are published in obscure journals, are rarely cited, or are inappropriately indexed in databases) is the most obvious form of reporting bias, *within-study publication bias* may be one of the most substantial biases affecting results from individual studies,⁴⁹ analyses with statistically significant findings, or extensive magnitude of effects being more likely to be reported compared to *uninteresting* or *unwelcome* findings. Reporting bias almost inevitably leads to major overstatements of efficacy,⁵⁰ including in preclinical PH research.²⁶ Intriguingly, selective submission by the authors rather than selective acceptance by the reviewers may

predominantly contribute to publication and reporting bias.^{51,52} Conversely, some journals indirectly contribute to this phenomenon by relegating *less interesting* findings to the supplement section. Publication and selective reporting biases also prevent others from learning about negative study results (which, by the way, should be allowed by editors to be published even in big impact-factor journals), with implications for animal ethics and research funding. To minimize publication and reporting bias, study preregistration was developed for clinical trials, limiting researcher ability to modify planned experimental design and analysis afterwards. As a result, the International Committee of Medical Journal Editors now considers only those clinical trials for publication that have been registered before the start of patient recruitment.⁵³ In preclinical studies, preregistration in a public repository at the study inception is a debated issue. Indeed, while a finding is more convincing when it was predicted, breakthrough findings have been made through exploration with limited a priori hypotheses.

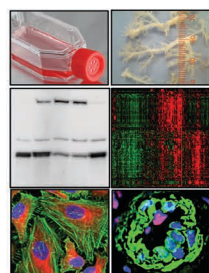
ADAPTING STATISTICAL AND METHODOLOGICAL RIGOR TO THE PROGRAMMATIC PURPOSE OF RESEARCH

While investigators seek to provide a better understanding of the pathophysiological processes and identify key cellular and molecular signaling pathways/targets involved in disease development in exploratory research, detailed and reproducible information on efficacy, dosing, and toxicity of potential drug candidates are required in confirmatory investigation to decide whether the drug could be tested in clinical trials. The statistical and methodological rigor should thus be adapted according to the nature of the study (Figure 3). Nonetheless, even at an exploratory stage, significant attention should be paid to identify, avoid, and acknowledge potential bias.

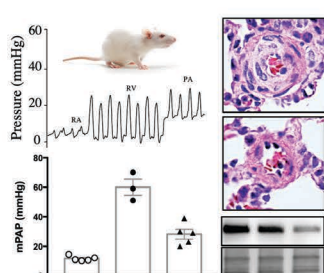
A CALL FOR CHANGES IN PRECLINICAL PH STUDIES

Scientific irreproducibility is a growing concern among academics and in the general population.⁵⁵ Biases and poorly

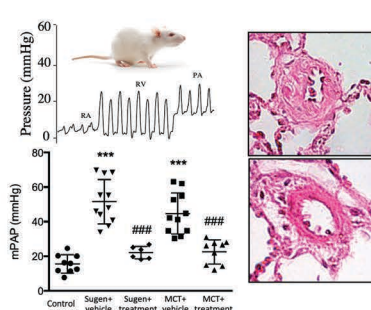
Exploratory



Proof-of-concept



Confirmatory study



Clinical trial

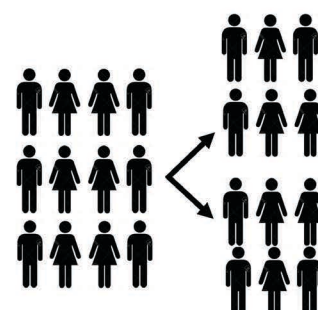


Figure 3: Adaptation of statistical and methodological rigor according to the nature of the study. For confirmatory studies, the most rigorous application of randomized controlled trial-like approaches should be applied, including prespecified study design and a statistical analysis plan describing eligibility criteria, sample size calculation, relevant study readouts, interim analysis (if any), and corrections for multiple analyses, as well as randomization and concealment procedures, validation of the results, and potentially study preregistration. However, this type of approach may be expensive and cumbersome, and applying randomized controlled trial-like standards to studies focusing on discovery and elucidating fundamental biological processes may not add the same level of value. Thus, methodological rigor must be appropriately calibrated to the goals of the study. In all cases, however, the choice of an appropriate animal model that is representative of the human disease, the development of standardized operating procedures, blinded outcome assessment, and transparent reporting are critical.⁵⁴

designed preclinical studies likely contribute to experimental irreproducibility, wasted resources, and erroneous conclusions.⁵⁶ In response to these issues, the National Institutes of Health (NIH) proposed a set of guidelines and funding policies as minimum reporting requirements to promote rigor, reproducibility, and transparency of preclinical research that have been endorsed by prominent academic societies and scientific journals with editorial commitment to complying.⁵⁷⁻⁵⁹ These include guidelines and checklists to improve methodology and reporting.^{57,58} Multicenter preclinical studies,^{47,60} systematic reviews, and meta-analyses are also preconized.^{48,61} Practical solutions to improve preclinical research quality and research translation have also been specifically proposed in PH preclinical studies.^{16,54} Implementing such requirements will involve a paradigm shift for scientists, their institutions, journals, and funding agencies.

CONCLUSION

In preclinical research, methodological sources of potential bias and imprecision are prevalent and frequently overlooked

by researchers, potentially contributing to the significant discordance between preclinical and clinical results. Although not unique to PAH, concerted efforts to address this problem are needed for more effective translation of preclinical research findings into sustainable improvements in patient outcomes, including rigorous study designs, methodological standardization, appropriate data interpretation, and statistical analysis plans, as well as transparent reporting of preclinical studies.

References

- Schermler RT, Ghofrani HA, Wilkins MR, Grimminger F. Mechanisms of disease: pulmonary arterial hypertension. *Nat Rev Cardiol.* 2011;8(8):443-455.
- Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: the Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur Respir J.* 2015;46(6):903-975.
- Lajoie AC, Lauziere G, Lega JC, et al. Combination therapy versus monotherapy for pulmonary arterial hypertension: a meta-analysis. *Lancet Respir Med.* 2016;4(4):291-305.
- Galie N, Manes A, Negro L, et al. A meta-analysis of randomized controlled trials in pulmonary arterial hypertension. *Eur Heart J.* 2009;30(4):394-403.
- Humbert M, Sitbon O, Chouat A, et al. Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. *Circulation.* 2010;122(2):156-163.
- Humbert M, Sitbon O, Yaici A, et al. Survival in incident and prevalent cohorts of patients with pulmonary arterial hypertension. *Eur Respir J.* 2010;36(3):549-555.
- Benza RL, Miller DP, Barst RJ, et al. An evaluation of long-term survival from time of diagnosis in pulmonary arterial hypertension from the REVEAL Registry. *Chest.* 2012;142(2):448-456.
- Hoepfer MM, Barst RJ, Bourge RC, et al. Imatinib mesylate as add-on therapy for pulmonary arterial hypertension: results of the randomized IMPRES study. *Circulation.* 2013;127(10):1128-1138.
- Ghofrani HA, Morrell NW, Hoepfer MM, et al. Imatinib in pulmonary arterial hypertension patients with inadequate response to established therapy. *Am J Respir Crit Care Med.* 2010;182(9):1171-1177.
- Granton J, Langleben D, Kutryk MB, et al. Endothelial NO-synthase gene-enhanced

- progenitor cell therapy for pulmonary arterial hypertension: the PHACeT trial. *Circ Res*. 2015;117(7):645-654.
11. Michelakis ED, Gurtu V, Webster L, et al. Inhibition of pyruvate dehydrogenase kinase improves pulmonary arterial hypertension in genetically susceptible patients. *Sci Transl Med*. 2017;9(413).
12. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32(1):40-51.
13. Pharmaceutical Research and Manufacturers of America. *Annual Report 2011*. Washington, DC: Pharmaceutical Research and Manufacturers of America; 2011.
14. Contopoulos-Ioannidis DG, Ntzani E, Ioannidis JP. Translation of highly promising basic science research into clinical applications. *Am J Med*. 2003;114(6):477-484.
15. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. 2010;87(3):272-277.
16. Bonnet S, Provencher S, Guignabert C, et al. Translating research into improved patient care in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2017;195(5):583-595.
17. Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov*. 2011;10(6):428-438.
18. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. *PLOS Med*. 2013;10(7):e1001489.
19. Ioannidis JP. Why most published research findings are false. *PLOS Med*. 2005;2(8):e124.
20. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg*. 2010;126(2):619-625.
21. Benza RL, Gomberg-Maitland M, Demarco T, et al. Endothelin-1 pathway polymorphisms and outcomes in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2015;192(11):1345-1354.
22. Ruffenach G, Chabot S, Tanguay VF, et al. Role for runt-related transcription factor 2 in proliferative and calcified vascular lesions in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2016;194(10):1273-1285.
23. Kadara H, Shen L, Fujimoto J, et al. Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev Res*. 2013;6(1):8-17.
24. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732.
25. Stenmark KR, Meyrick B, Galie N, Mooi WJ, McMurtry IF. Animal models of pulmonary arterial hypertension: the hope for etiological discovery and pharmacological cure. *Am J Physiol Lung Cell Mol Physiol*. 2009;297(6):L1013-1032.
26. Sztuka K, Jasinska-Stroschein M. Animal models of pulmonary arterial hypertension: a systematic review and meta-analysis of data from 6126 animals. *Pharmacol Res*. 2017;125(Pt B):201-214.
27. Mitani Y, Mutlu A, Russell JC, et al. Dexfenfluramine protects against pulmonary hypertension in rats. *J Appl Physiol*. 2002;93(5):1770-1778.
28. Pullamsetti SS, Berghausen EM, Dabral S, et al. Role of Src tyrosine kinases in experimental pulmonary hypertension. *Arterioscler Thromb Vasc Biol*. 2012;32(6):1354-1365.
29. Van der Velden J, Snibson KJ. Airway disease: the use of large animal models for drug discovery. *Pulm Pharmacol Ther*. 2011;24(5):525-532.
30. Pinn VW. Sex and gender factors in medical studies: implications for health and clinical practice. *JAMA*. 2003;289(4):397-400.
31. Ramirez FD, Motazedian P, Jung RG, et al. Sex bias is increasingly prevalent in preclinical cardiovascular research: implications for translational medicine and health equity for women: a systematic assessment of leading cardiovascular journals over a 10-year period. *Circulation*. 2017;135(6):625-626.
32. Humbert M, Sitbon O, Chaouat A, et al. Pulmonary arterial hypertension in France: results from a national registry. *Am J Respir Crit Care Med*. 2006;173(9):1023-1030.
33. Galie N, Barbera JA, Frost AE, et al. Initial use of Ambrisentan plus Tadalafil in pulmonary arterial hypertension. *N Engl J Med*. 2015;373(9):834-844.
34. Crossley NA, Sena E, Goehler J, et al. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke*. 2008;39(3):929-934.
35. Rooke ED, Vesterinen HM, Sena ES, Egan KJ, Macleod MR. Dopamine agonists in animal models of Parkinson's disease: a systematic review and meta-analysis. *Parkinsonism Relat Disord*. 2011;17(5):313-320.
36. Pildal J, Hrobjartsson A, Jorgensen KJ, et al. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *Int J Epidemiol*. 2007;36(4):847-857.
37. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336(7644):601-605.
38. Liu S, Zhen G, Meloni BP, Campbell K, Winn HR. Rodent stroke model guidelines for preclinical stroke trials (1st edition). *J Exp Stroke Transl Med*. 2009;2(2):2-27.
39. Ludolph AC, Bendotti C, Blaugrund E, et al. Guidelines for preclinical animal research in ALS/MND: a consensus meeting. *Amyotroph Lateral Scler*. 2010;11(1-2):38-45.
40. Moreno B, Espejo C, Mestre L, et al. Guidelines on the appropriate use of animal models for developing therapies in multiple sclerosis. *Rev Neurol*. 2012;54(2):114-124.
41. Holm S. A simple sequentially rejective multiple test procedure. *Scand J of Stat*. 1979;6(2):65-70.
42. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*. 1988;75(4):800-802.
43. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*. 1988;75(2):383-396.
44. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol*. 1992;21(5):837-841.
45. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31(4):337-350.
46. Sellke T, Bayarri M, Berger J. Calibration of p values for testing precise null hypotheses. *Am Stat*. 2001;55(1):62-71.
47. Van der Feen DE, Kurakula K, Tremblay E, et al. Multicenter preclinical validation of BET inhibition for the treatment of pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2019;200(7):910-920.
48. Sandercock P, Roberts I. Systematic reviews of animal experiments. *Lancet*. 2002;360(9333):586.
49. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7479):753.
50. 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.
51. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ*. 2005;331(7507):19.
52. Weber EJ, Callahan ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA*. 1998;280(3):257-259.
53. De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2004;351(12):1250-1251.
54. Provencher S, Archer SL, Ramirez FD, et al. Standards and methodological rigor in pulmonary arterial hypertension preclinical and translational research. *Circ Res*. 2018;122(7):1021-1032.
55. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452-454.
56. Gonzalez Martin-Moro J. The science reproducibility crisis and the necessity to publish negative results. *Arch Soc Esp Oftalmol*. 2017;92(12):e75-e77.
57. National Institutes of Health. Rigor and Reproducibility: Principles and Guidelines for Reporting Preclinical Research. <https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines->

- reporting-preclinical-research. Accessed April 2, 2020.
58. Bolli R. New initiatives to improve the rigor and reproducibility of articles published in circulation research. *Circ Res*. 2017;121(5):472-479.
 59. Vahidy F, Schabitz WR, Fisher M, Aronowski J. Reporting standards for preclinical studies of stroke therapy. *Stroke*. 2016;47(10):2435-2438.
 60. Llovera G, Hofmann K, Roth S, et al. Results of a preclinical randomized controlled multi-center trial (pRCT): anti-CD49d treatment for acute brain ischemia. *Sci Transl Med*. 2015;7(299):299ra121.
 61. Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke*. 2004;35(5):1203-1208.