Advanced Drug Delivery Reviews 182 (2022) 114118

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/adr

Reproducibility, relevance and reliability as barriers to efficient and credible biomedical technology translation

Ulrich Dirnagl [a,b,1], Georg N. Duda [c,d,1], David W. Grainger [e,f,*], Petra Reinke [c,g,1], Ronenn Roubenoff [h,1]

[a] Department of Experimental Neurology, Charité - Universitätsmedizin Berlin, Germany
[b] QUEST Center for Responsible Research, Berlin Institute of Health, Germany
[c] Berlin Institute of Health (BIH) Center for Regenerative Therapies (BCRT), Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Germany
[d] Julius Wolff Institute for Biomechanics and Musculoskeletal Regeneration, Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Germany
[e] Department of Pharmaceutics and Pharmaceutical Chemistry, Health Sciences, University of Utah, Salt Lake City, UT 84112 USA
[f] Department of Biomedical Engineering, University of Utah, Salt Lake City, UT 84112 USA
[g] Berlin Center for Advanced Therapies (BeCAT), Charité - Universitaetsmedizin Berlin, 13353 Berlin, Germany
[h] Novartis Institutes for Biomedical Research, Cambridge, Basel, Massachusetts, Switzerland

ARTICLE INFO

Article history:
Received 7 December 2021
Revised 14 January 2022
Accepted 15 January 2022
Available online 20 January 2022

Keywords:
Healthcare
Pharmaceutical
Data quality
Preclinical research
Medical device
Predictive
Clinical innovation

ABSTRACT

Biomedical research accuracy and relevance for improving healthcare are increasingly identified as costly problems. Basic research data quality, reporting and methodology, and reproducibility are common factors implicated in this challenge. Preclinical models of disease and therapy, largely conducted in rodents, have known deficiencies in replicating most human conditions. Their translation to human results is acknowledged to be poor for decades. Clinical data quality and quantity is also recognized as deficient; gold standard randomized clinical trials are expensive. Few solid conclusions from clinical studies are replicable and many remain unpublished. The translational pathway from fundamental biomedical research through to innovative solutions handed to clinical practitioners is therefore highly inefficient and costly in terms of wasted resources, early claims from fundamental discoveries never witnessed in humans, and few new, improved solutions available clinically for myriad diseases. Improving this biomedical research strategy and resourcing for reliability, translational relevance, reproducibility and clinical impact requires careful analysis and consistent enforcement at both funding and peer review levels.

Contents

* Corresponding author at: Department of Pharmaceutics and Pharmaceutical Chemistry, Health Sciences, University of Utah, Salt Lake City, UT 84112 USA.
E-mail address: david.grainger@utah.edu (D.W. Grainger).
[1] Names listed in alphabetical order; all authors equally contributed to the publication.

https://doi.org/10.1016/j.addr.2022.114118
0169-409X/

"Knowing is not enough; we must apply. Willing is not enough; we must do."
J.W v. Goethe

1. Introduction

Translation of biomedical research results into clinical benefits is the rallying cry of the modern medical research establishment [1]. Medical innovation is linked to effective translation of new discoveries about disease, and in how drugs and devices produce therapies. Effectively translating observations from fundamental experimental biomedical research protocols (e.g., *in vitro, ex vivo, in vivo, or* in silico, in animal preclinical studies) to address human diseases treatments and improvements in clinical routines has diverse challenges [2–9]. Myriad murine disease models are frequently used to herald new "cures" for diverse human diseases, that unfortunately for most, prove to be invalid [10,11]. Only a small fraction of animal study outcomes are deemed transferrable to relevant human responses, thus qualifying as "knowledge-gaining research" [12]. The remainder often claim "potential" relevance, yet are poorly convincing, unsupported, or too risky or ambiguous to attempt correlation or translation to human conditions, and without clear clinical impact. This widely recognized but worrisome chasm separating discovery from technology validation and clinical impact de-values the role and credibility of the biomedical scientist and erodes their contributions to addressing compelling healthcare challenges [13,14].

Comprehensive expert-level grasp of each medical challenge, its causality and progression, and how likely preclinical metholodologies and data are reliably translated to the clinically relevant context, are often required to confidently transition from preclinical testing to clinical validation. Accurate, validated scientific evidence generated in a timely manner in relevant biomedical research testbeds is required to address both unmet needs and also known divides between biomedical research and clinical challenges. "Translation" is defined as the "essential process of turning observations in the laboratory, clinic and community into new interventions that improve both the health of individuals and the public — from diagnostics and therapeutics to medical procedures and behavioral changes" [15]. Nonetheless, formidable barriers are frequently recognized that preclude ready achievement of this mission [16]. Traditionally, distinctions in translational cultures among investigators, regulatory hurdles, limited data access, reproducibility, usability, and poorly predictive research models have been identified [17]. Practically, a critical barrier surrounds the increasing complexity of biomedical information and limited research capabilities to integrate complex multi-factorial data across multiple research formats. Research strategies to effectively and comprehensively accommodate complex, dynamic models of health, disease and intervention do not yet exist in many cases. Biomedical research then often employs either a holistic whole organism approach, often equated with a black-box empirical study with control of few of myriad parameters, or a simplified reductionist approach where experimental design suffers from lack of relevance or realistic approximations to the actual human condition justifying the study.

2. Predicting translational success

A particularly acute facet of the translational research challenge is evident in the depressingly low rate of successful translation of preclinical models to human experiences [16,18–21]. Years ago, a systematic review identified that only approximately 33% of 67 highly cited animal research studies could translate accurately at the level of published randomized human clinical trials [22]. Translational predictiveness and research reliability shown in that study are poorer than the recently estimated replication rates, less reliable than a coin toss, for highly cited human studies [23,24]. Given these precarious features of translational science, extrapolating outcomes from animal research as models of human maladies into claims for approaches to treating human disease should be performed with caution [25]. These noted deficiencies certainly provide major insights and opportunities to address deficiencies in study designs and improve methodological qualities in much preclinical research that might improve their human relevance.

"The definition of insanity is doing the same thing over and over again, but expecting different results."

(attributed to 1981 *Narcotics Anonymous* pamphlet)

2.1. Changing decades of habit

To produce different, and better, results than those published, the research community must attempt different methodologies and approaches. Over 26 years ago, Altman asserted, "We need less research, better research, and research done for the right reasons" [26]. Yet, scientists like all other humans are creatures of routine who respond naturally to the incentives provided for performance: *it is constantly* challenging *to expect, develop and enforce different and hopefully improved, validated approaches to address long-standing challenges, particularly in medical research centered on humans.* One prominent hindrance is the dominant, pervasive incentive system for scientific recognition, promotion and success in academic research, relying on impact-agnostic numerical compilations and assessments of scholarly production [27–31]. Indeed, academic performance analytical tools now commonly employed by university administrative rankings and assessments use publications as "a currency they were never meant to be: a system of metrics to assess research, research programs and individual researchers" [32]. These research performance constraints perpetuate the long-standing insidious academic "publish-or-perish" culture, engaging 15,000,000 researchers publishing over 25,000,000 scientific papers in 1996–2011 alone [33], without much incentive to change either the metrics, merits, or the results. Furthermore, the audience for this mass of "discovery" literature is unappreciative and inattentive: the 10-year uncited rate for publications across all science disciplines, minus self-citation, is about 18% [34]. This excessive and under-appreciated global dissemination effort is openly acknowledged as a costly system that fails all involved in bringing the expected academic learnings, progress, innovation and research breakthroughs to benefit society. Nonetheless, when "researchers are rewarded primarily for publishing, then habits which promote publication are naturally selected. . .. they modify their methods to produce the largest possible number of publishable results rather than the most rigorous investigations" [29]. Enormously profitable scientific publishing business interests [35] and the dubious roles of researchers as both the producers and consumers in this publication business (and who pay in both roles) introduce orthogonal pressures on research systems as well. New, unanticipated consequences and unfiltered media hype, often through rapid social media dissemination of non-factual reports and non-peer reviewed evidence [36] produce new complexities for scientific accuracy. Breaking this performance pattern, removing the perverse incentives from for-profit external forces [27–31] and restoring biomedical research to originally envisioned more altruistic and impacting goals beyond publications will require concerted will and dedication from numerous stakeholders [30,37,38]. Publications are important dissimenation tools as critical reports of progress to their stakeholders, but these are not final products [32,37]. Publishers and journal editorial

boards could wield increasing influence in setting standards for acceptable research conduct and quality [39].

Researchers themselves, along with their peers and administrators who supervise and enforce the promotional and merit-review/reward systems, also bear collective responsibility to shift research performance criteria embedded for academic advancement. These are traditionally based on consistent, reliable research productivity, not necessarily reliable conduct. Junior faculty are compelled by necessity and security to seek incremental contributions to established research directions as a most direct route to publication. Original, more creative, and potentially high impact pursuits are deemed too risky as unproven directions for new researchers [40]. These two divergent, conflicting research career strategies pit traditional production against high-risk, high gain innovation to start a research career [41]. Incremental, conservative research yields few surprises beyond intuition, while risky studies represent a gamble lacking sufficient citation benefits to junior investigators to endorse such undertaking. Additionally, any serendipitous or unexpected findings should they be discovered, tend to be ignored as a spurious [42]. Beyond steering young researchers away from more creative work, innovation under such "tension" is compromised. Unexpected findings are known to be essential in supplanting established scientific theories (i.e., Kuhn's "scientific revolution") [43] Limiting the choices, risks and payoffs to researchers taking creative risks and disruptive thinking explains the low probability of reporting unexpected creative breakthroughs that power innovation [41]. As the current research environment incentivizes conservative research conduct opposing radical novel discoveries, ensuring early job security separately from arbitrary research productivity is known stimulate creativity (i.e., the "Bell Labs model" [44]).

Funding agencies that create and promote research programs influence heavily the directions, products and quality of the research enterprise through performance expectations, regular review and grant deliverables. Funding research based on failed research tools and strategies, and irrelevant disease models must cease. Funding for preclinical model characterization and validation should be an essential component preceding funded reliance on any model. Risk and creativity, as well as expectations for reproducibility and quality, can be titrated into funded research directions to balance more conservative modes of research through innovative programming and aggressive funding approaches that favor high-risk, high gain outcomes, and technical innovation. Finally, the peer-review community that critiques, approves and refuses both funding directions from research agencies, and the products of research in their dissemination must be consistently directed to best practices. This community is the guardian of research quality control, robust conduct, research vision, and performance expectations. Too often this expert group is seen as seemingly condoning the exaggerations of novelty and claims of impact while perpetuating certain systemic futility and wasted resources through on-going collective approval of compromised models, poor strategies and faulty techniques. Peer review itself has been called out as a culprit actively responsible for perpetuating the poor quality of published data [45]. Peer review expert groups must all be called upon to recognize their critical role and implement quality control and enact experimental changes in the global research system.

To best conduct research that is both reliable and reproducible, researchers must first understand the importance of reproducibility, replicability, and transparency, the adverse influence of various forms of bias and poor statistical support, and then be trained in best practices with the best tools available. Funding agencies have recently recognized this deficiency in training and mentoring systems, and either require or recommend specific competences in research ethics and conduct to be demonstrated by awardees. Educational institutions and others increasingly incorporate reproducibility training into courses and certifications, aimed both at aspiring junior researchers and senior investigators [46].

Nonetheless, poor researcher behavior is unlikely to improve substantially from simply encouraging or recommending responsible research practices. Long-implemented mandates from journals, instructing authors about acceptable publication standards and more accurate reporting expectations [47], have produced little notable impact: no improvement is observed in fractions of papers properly using or reporting statistics [15,16], or applying statistics to data in figures [17]. To date, only a dismal percentage of journals actually enforce transparent scientific reporting practices [18]. As a result, published biomedical research reporting remains largely inadequate [19,20,21]. Since researchers do not follow explicit recommended journal reporting practices, they are unlikely to adhere to responsible research practices expected but not required in grant proposals [48].

2.2. Addressing preclinical failure

Currently, the world's largest biomedical research funding body (National Institutes of Health, USA, NIH) spends approximately $12–14.5 billion yearly on animal experiments; roughly half of awarded NIH grants conduct animal–based research, a stable trend observed over a decade [49]. Use and enforcement of whole organism or animal research preceding human testing or to benefit the human condition has vast historical and ethical complexity, resulting in regulatory statutes, mandates and expectations for human translation [50]. Few experimental precedents show accepted capabilities to bypass this powerful historical, scientific and ethical animal preclinical "requirement" in moving preclinical biomedical research to possible clinical application.

Using previous publications uncritically to justify further *in vivo* work, regardless of their veracity, relevance, robustness or quality, is often the most rapid route to institutional animal study approval and to obtaining publishable data, even though these data may have no translational relevance. Research data production is enabled but relevance and impact are compromised. Low animal-to-human translational success rates continue to plague most aspects of biomedical research. Two opposed camps commonly attempt to explain translational failures. One advocates that the entire concept of animal-to-human research predictability is fundamentally flawed [51]. Support comes from more obvious aspects of specific molecular, genetic, and physiological differences that cannot be experimentally reconciled, as well as more philosophical arguments that animals and humans comprise chaotic systems for study that are inherently unpredictable, and that animal-to-human predictability is not yet scientifically validated [52]. Some argue that animal models are outdated, antiquated, and now-irrelevant, [53] supplanted by new, improved methods including digital twins and in silico physiology, and organ-on-chip test bed replacements.

The other camp claims that many medical research advances have necessarily relied on multiple different, predictable animal outcomes to benefit humans and this clinical advancement in drugs and devices therefore endorses the value of animal models [54]. Translational failures in this view result from faulty experimental designs or conduct [52] and the intrinsic lack of reproducibility of biomedical research [55]. Despite on-going, repeated critique of poorly designed but expensive animal studies, methodological flaws, poor analysis and reporting, and wasted resources and scientific distraction, [9,11,56–58] the problems persist without improvements. Under-powered animal studies, long known to be the bane of preclinical translational reliability [59–61], continue to be published and accepted as valid whole organism *in vivo* outcomes, despite poor methodological design [21,62], lack of validation and low or no human relevance. Appropriate

experimental designs accompanied by randomization, blinding and statistical methods are basic research requirements needed to achieve more reliable scientific results and ensure their validity for translational use. *A major goal for any model truly intent on human translation should be to reduce the number of costly false positives in preclinical studies that currently promote expensive but unjustified clinical trials.* Nonetheless, preclinical data may be seductive and apparently reinforcing to a stated clinical goal, yet not translate reliably or sufficiently to subsequent clinical research steps for diverse reasons [2,10,18,21,62–72]. An extensive and tabulated analysis of preclinical translation rates is provided in [52].

Given abundant questioning and critique published for many current animal models used in biomedical research, criteria for defining validity of improved animal models is essential, and refining best practies for their adoption and continual evolution should be a research community mandate [73,74]. Model validation and characterization are enormous investigative efforts and impacts in their own right, deserving of dedicated funding resources and ethical resolve before any further research pursuit. Some experimental uncertainties inherent in biomedical models lie beyond experimental control – physiology and pathology are indeed complex processes, often exhibiting non-linear or quasi-chaotic dynamics difficult to capture, control or reproduce in models [7,11]. Model organismal pathophysiology is often sufficiently distinct from humans to limit general comparisons beyond verified isolated mechanistic details. For these preclinical challenges, some research utility and validity can be justified for value other than often-claimed comprehensive human relevance. Focus on specific relevant mechanistic disease or healing signaling, pathway elucidation, pharmacology/toxicology profiles, or congruent genetic contributions are reasonable goals for non-human-equivalent investigative models. That both external validity (i.e., verified translation to other researchers conducting similar studies, in other study groups or preclinical species) and internal validity (i.e., reproducible, expert research protocol design, competent experimental performance, critical analysis of results, and factual, unbiased reporting) are necessary, and further that research results from animal models can only be reliably translated to human use if related preclinical studies exhibit both external and internal validity, have been argued [51,75].

> *"Quality is never an accident; it is always the result of high intention, sincere effort, intelligent direction and skillful execution; it represents the wise choice of many alternatives, the cumulative experience of many masters…"*
>
> William A. Foster

3. Reproducibility and robustness

3.1. Internal validity

Internal validity as represented by experimental reproducibility and robustness is an essential feature of reliable science and requisite for biomedical translation [76]. Substantial poor quality research of low internal validity and statistical power is shown to be a major contributor to translational attrition [22,33,59,61,62,70]. For example, a rigor and transparency analysis of 1.6 million papers reporting preclinical research results covering the period 1997–2019 showed that less than one third of studies describe experimental methods that might reduce outcome bias (i.e., randomization, powering, blinding, etc.) [77]. Most research appears to produce on average, a statistical power below 10%. As a result of poor experimental designs and resulting data quality, both false positive and negative results are common. Even for effects that are "real", their associated effect sizes may be grossly exaggerated [22,30,78].

Among notable consistent deficiencies in experimental design and data reporting, poor or wrong statistical analysis ranks highly. The biomedical field's continued, persistant reliance on null hypothesis significance testing (NHST), and consequent use and abuse of *P* values to demonstrate results significance [59–61,78,79] is claimed now to represent the "most widely perpetrated misdeed of statistical inference across all of science" [79]. Data cherry-picking strategies, improper selection bias, and p-hacking are increasingly employed to deceptively demonstrate study statistical validation [40]. Judicious, informed use of legitimate effect sizes and appropriately validated confidence intervals, methods to disclose false discovery rates, Bayesian techniques, and more stringent criteria to both determine and apply *P* values have all been advocated as best practices to avoid these increasingly reported and questionable problems undermining data quality, reliability and robustness [29,30,76]. Interestingly, given the recent vigorous discussions about data reproducibility and replicability, the American Statistical Association (ASA) broke with 177 years of precedent of not asserting positions on statistical methods to publish six principles about p-values, hoping to "shed light on an aspect of our field that is too often misunderstood and misused in the broader research community" [80]. The principles cover key issues in statistical reporting. A 2019 ASA publication now strongly discourages use of a statistical significance threshold in reporting research results due to overuse, abuse and broad misinterpretation of utility and applicability [81].

3.2. External validity

Research generalizability (i.e., external validity) is necessary for translational success and reliability. How well specific aspects of a chosen research model reflect critical factors of a relevant clinical setting is key to recapitulating disease pathophysiology, and hence for validating outcomes for possible therapeutic predictions. This includes – but is not limited to - sex, age, immune system status, microbiome, etc, as modifiers of an adequate model [82]. Expert knowledge of the specific human medical challenge, unique physiological and pathological traits and their possible variation are critical for identifying appropriate animal research analogues. Nevertheless, experimental heterogeneity was recently advocated to be more widely accepted as a means to improve model reproducibility, reliability and translatability. Rather than increasing reproducibility, experimental standardization so commonly emphasized may actually reduce variability within studies perhaps as intended, but also promote idiosyncratic, lab-specific results that are neither generally reproducible or translatable [83]. Hence, deliberate introduction of experimental heterogeneity (i.e., "heterogenization") into research may improve success rates in drug and medtech development and their subsequent translation.

Preclinical model capabilities to reliably predict human outcomes varies substantially across species, pathology, and intervention. Many human diseases produced in primary animal models (where they in fact never occur naturally) have been "cured" repeatedly using diverse interventional approaches, but notably without ever producing a successful human therapy or mitigation (e.g., Duchenne muscular dystrophy using the *mdx* mouse [84], the EAE-model of multiple sclerosis [85], several research models of sepsis and infection in animals [86,87], various animal solid organ transplantation models showing tolerance induction [88], hundreds of different murine tumor types and anti-cancer therapies [65]). Meanwhile, other medical challenges produce better model predictivity (e.g., inflammatory arthritis models predicting TNF inhibitor efficacy in rheumatoid arthritis [71,72,89], estrogen withdrawal on osteoporosis [90], CAR-T cells addressing liquid

tumors, and immunotherapies involving checkpoint inhibitors), despite the animal model eliciting few similarities to the actual human malady (e.g., most mammals fail to naturally manifest true osteoporosis [91] or, as in humans, spontaneously develop tumors [65]).

> *"The best material model of a cat is another, or preferably the same, cat."*
>
> (Norbert Wiener, with A. Rosenblueth, Philosophy of Science, 1945)

3.3. The end of the animal research Model?

Better predictive, reliable data needed to verify a hypothesis preclinically before moving to human medical assessments might ideally be sourced from orthogonally designed research - preferably using human genetic validation, or inferential human real world data (RWD) as biomedical evidence, or relating analogous pathway intervention outcomes observed in other human therapeutics, devices, etc. *Spending more time, effort and resources to establish new animal research models with more accurate preclinical relationships, relevance or equivalence to model human disease is not perceived to improve translational prospects*. This is a time-consuming, expensive, tedious distraction for the field, rarely performed adequately for past model validations, and likely to be rate-limiting to validating successful future translational models. Limitations long-recognized in most biomedical animal research models may never serve to duplicate any human disease entirely or accurately [18–22,65] – and importantly, need not do so to remain useful for research purposes [19,62]. Animal models have utility primarily for only select aspects of biomedical research validation and confirmation, often only limited mechanistic acute pharmacological, toxicological, or biomechanical features, and never an entire human scenario, either in disease propogation or healing response. Human translational forecasting therefore should not and cannot rely on such models. Yet, despite repeated criticisms, there does not appear to be the resolve in the research community to understand, directly address and alter many of the current challenges in translating animal results to human use: critical reviews, evaluations and collective recommendations of compiled evidence from animal research are methodologically inadequate [92]; few animal study *meta*-analyses are conducted compared to clinical trial *meta*-analyses to produce useful expert consensus on best-practices [20,62]. The only emerging consensus appears to be that most animal research is poorly performed and inadequate to inform human health applications, and that much preclinical research resources are wasted chasing elusive outcomes [18–25].

The pharmaceutical industry commonly finds it much more efficient and effective to access human assessments as quickly and directly as possible once limited supporting preclinical evidence of intended benefit is obtained from preclinical models to demonstrate that a targeted disease pathway is substantially corrected by the proposed candidate therapy. Companies are generally most willing to test their therapeutic strategies in the most relevant target species – humans – provided that essential pharmacological and toxicological safety parameters are satisfied in selected, relevant preclinical studies. Predictive pathologic pathway mechanistic information is more insightful in preclinical testing that actually establishing complete animal-human disease or healing equivalence. Ironically, toxicology studies needed for human trials are uniformly done in healthy animals, not in disease models. To address this goal, different critical issues – human safety, toxicity, dosing, effect size, and disease biomarkers – become the priority, and that do not entail more fundamental experiments using disease models.

Few studies are known to have proven clinical benefit despite preclinical animal models demonstrating none: few clinical studies would have the basis for ethical or regulatoy approvals to move ahead without strong animal proof of concept, safety and efficacy. A recent human-based *in vitro* model of retinal disease using retinal organoids from human stem cells better informed pharmacodynamic properties of lead drug candidates than the "optimal" rodent model addressing Leber's congenital amaurosis type 10 caused by CEP290 genetic mutations [93,94]. Discovery and subsequent development of lead drug molecule (Sepofarsen) resulted [95,96]. Preclinical animal models might well be further bypassed for translational benefits using current developments of digital twins [97], in silico physiological modeling [98], and organ-on-chip assays [99]. To both safely, ethically and rapidly facilitate the human tissue- and human experience-based data often classified as "translational research", further focus, attention and strategic resources are therefore needed beyond typical preclinical evidence that either directly equates animal response to human response, or identifies analogous isolated mechanisms [3,6,8–10], likely also considering also strengths of the probabilistic evidence for model predictivity in addition to possibly relevant mechanistic implications [52].

Regulators, clinical trials and academia: Worlds apart

Academic institutions typically dominate the fundamental and applied biomedical research forum. Further, more extensive, validating screening processes for therapeutic candidates are typically conducted within the pharma industry, with university researchers contributing to industry-sponsored or investigator-initiated clinical trials. The latter industrial process operates under the auspices of formalized and continuous regulatory oversight from health authorities. Academic research is generally regulated only by animal use committees, funding reviews, and institutional review boards. Increasing activities involving new advanced biologic therapies (i.e., cell and gene therapies, tissue engineering, medical biotechnologies, and various combinations) recently have crossed into both of these domains, eliminating the traditional gap between regulatory and academic operations. Many clinical innovations now originate and develop within academic labs, accompanied by newly emergent regulatory challenges [100], but also new translational dimensions to address in this operational continuum. Should this converging trend continue, then junior researchers should be necessarily trained uniquely to yield a mindset and skill set targeting translational medicine strategies and associated regulatory components [101,102]. Furthermore, translational de-risking and directional acceleration is enabled by bridging fundamental discovery-based research and their subsequent regulated clinical development, but only if sufficiently high-quality standards are enforced during transitional development phases [3,102]. The validity of preclinical mechanistic therapeutic data, its reliability, accuracy and reproducibility must, however, remain the essential base from which clinical research is built.

Beyond preclinical proof of concept, all new medicines, interventions and treatments result ultimately and necessarily from clinical trial volunteers. Human trials are conducted in clinical research using recruited, screened patient volunteers to address human-relevant medical issues. This evidence is required by national regulatory bodies for human use approvals. Randomized controlled trials (RCTs) - the gold-standard trial format – are the preferred, optimal approach to assess both safety and efficacy of new interventions. Trial design processes for RCT conduct minimize risks from confounding factors that might influence outcomes. As a result, RCTs are widely encouraged as the ideal methodology for causal inference and estimates of average treatment effects. Together with *meta*-analyses of pooled clinical trial data, the highest level of medical evidence is generated by high-quality RCTs designed with a low risk of systematic error (bias).

Analogous to other forms of biomedical research (*vide supra*), arguments considering their external and internal validity are published [103].

While RCTs may be considered the gold standard for generating clinical evidence, they are expensive, time-consuming, lengthy, tedious and difficult. Despite obvious ethical implications involving human quality of life, clinical research also has questionable quality, reliability and replicability challenges [24,33,104]. Because clinical acceptance, patient safety and treatment efficacy, and further research designs all depend directly on conclusions drawn from clinical research, and *meta*-analysis quality is intrinsically dependent on clinical trial evidence quality, clinical research reliability and veracity is essential. Nonetheless, the dangers of "spin" in deliberately distorting accuracy and validity in reporting clinical trial data remain highly concerning [105]. Interestingly, much of modern medical practice and established routine is not based on RCT vetting, as clinical medicine relies primarily on empirically "grandfathered" best practices and anecdotes collected and established from centuries of observational evidence. Hence, medical interest increases in using real-world evidence (RWE) to avoid the patient exclusion criteria used in RCTs, and to better reflect actual patient demographics, co-morbidities, protocol adherence, and concurrent treatment use in actual clinical environments, However, only a small fraction of RCTs are replicated to date in the real world [106] and these are typically using retrospective (observational) RWE patient data.

Regardless of clinical trial design and conduct, an essential contribution of clinical trial data comes from reliable trial conduct, including results reporting. Regulatory bodies are compelled to make detailed testing and evaluation information for regulated products available to the public to support trial enrollment, inform clinical care decisions, and accelerate future research. Nonetheless, despite nearly 275,000 clinical trials registered on ClinicalTrials.gov currently, fewer than 10% of these trials report trial results publicly back to the site annually. Furthermore, less than 50% of all clinical trial results have ever been published [107] and less than half of NIH-sponsored clinical trials are published in peer-reviewed journals within 30 months of trial completion [108,109]. Thousands of USA-conducted clinical trials are identified as noncompliant with regulatory agency results reporting requirements as of January 2021. The regulatory (and often funding agency) requirement and expectations that public recruitment and involvement in clinical research should be reported publicly in a timely manner has only recently begun to be enforced [110].

Ironically, clinical trials demonstrating positive outcomes are twice as likely to be published as those showing negative results [107,111]. Early published data for trials evaluating treatments of chronic medical conditions commonly demonstrate exaggerated treatment effects compared to later, subsequent similar trials [23,112]. In over a third of clinical studies analyzed, first or second clinical trial study outcomes reported an effect 2.67 times greater than outcomes from subsequent identical trials. Further, clinical trial data gleaned from small clinical trials published in major medical venues tend to be more exaggerated than results from equivalent studies published in other journals [113]. Trial results are most often published in English, and the likelihood of publication is frequently unrelated to sample size, funding mechanism, investigator rank or gender [114]. The notable lack of normal and consistent publication practices represents a significant publication bias, and in this case one that is highly unpredictable and non-uniform.

Unfortunately, scientific research validity is seriously threatened by such publication bias. Decision-making for clinical innovations based on studies showing exaggerated benefit, or biased by lack of studies published showing adverse effects, provides very low certainty in such recommendations [111]. Further, clincial data *meta*-analyses are more likely to consider more trials reporting positive findings than negative findings, producing estimated pooled effect sizes more exaggerated by publication bias. Resulting clinical recommendations resulting from evidence so influenced by publication bias falls victim to likely exaggerated benefits. Therefore, fair assessment of patient benefit and harm producing such recommendations is likely inaccurate. This risk is acknowledged in clinical trial assessments by reducing certainty in the quality of the evidence provided [115].

Success rates for biomedical technology translation are much lower than desired or expected given the magnitude of global resources dispensed to solve human-relevant medical problems [1,6–9,116]: less than 10% of pre-screened drugs entering phase I human clinical trials eventually become registered as new drugs, with most disappearing after costly failed trials [6,9,10,117]. Acknowledging the known limitations of preclinical models discussed above, early clinical trials must be designed and analyzed to provide key learnings by relying on carefully designed mechanistic studies employing safety, pharmacokinetic/pharmacodynamic (PK/PD), mechanism-of-action, and surrogate biomarker analyses. Given that proportions of early trial failures attributed to safety versus efficacy are likely roughly equal, and also not frequently posted to clinical registries, generating reliable evidence from human studies relies on exploiting responsibly validated biomarkers and their appropriate utility in tests and outcomes assessments – another important, often-overlooked opportunity to improve data quality and learnings [118]. Study outcomes based on competent trial design and data collection allow iterative

Table 1
Factors influencing quality biomedical research and translational success.

- Poor understanding of the underlying pathophysiology and medical need
- Researcher lack of awareness/denial of Data Quality and Research Reliability challenges
- Researcher lack of competence/training in Reliability/Reproducibility/Bias/Ethics/Statistics
- Inadequate enforcement of research ethics, statistical and experimental design training and qualifications for granting and data reporting
- Funding/Granting systems intrinsic disincentives for research rigor or change
- Academic promotional system and performance mandates and traditions enforce the status quo
- Early career research "Risk versus Quality versus Rapid Surety" in Research strategy and productivity
- Academic publishing mandates and for-profit publishing predation
- Research peer recognition/visibility in professional communities
- Social media and press dissemination; exaggeration of impacts to the public
- Poorly validated, characterized research models for best studying complex human conditions
- Poor validation of biomedical reagent quality
- Deficient research design and data reporting for animal studies to allow interpretation of results
- Poor clinical trial designs, conduct, results reporting and data access
- Intellectual property protection strategies as an essential distraction for translation
- Poor understanding of the translational collaborative research chain, partners and cultures

improvements to an intervention, therapy or device applications, and to better informing patient trial selection to better certify result using a "back-to-bench-forward-to-bed approach" (i.e., "refined translation") – a strategic tactical de-risking process [3,102]. Table 1 summarizes the major factors affecting how biomedical research is designed, conducted, compromised, reported and translated.

Table 2 describes approaches to improve research conduct through best practices and incentives from the diverse stakeholders.

> "*. . . despite the substantial resources invested into basic biomedical research, a vast majority of findings will never be tested in humans, let alone culminate in change in clinical practice.*"
> R. Ogier, W. Knecht, and M.E. Schwab [124].

Table 2
Practical Solutions to Improve Biomedical Research Rigor, Reliability and Reproducibility.

Create and sustain a professional biomedical research ecosystem

- Instill an understanding of how and biomedical research goes wrong, causation, mechanisms and its adverse impact on discovery, progress and public confidence.
- Instill a culture of professional responsibility and stewardship for the research process and system, from design to conduct to reporting and peer reviewing expectations, from trainees through to senior investigators.
- Establish standards of research reproducibility and replicability and teach investigators how to accomplish these in biomedical research practices.
- Engrain an understanding of different forms of bias and how they impact research quality.
- Establish targeted training, mentoring, and certification of established principal investigators (PIs) to reinforce application of best practices throughout the biomedical research process.
- Establish research funding agency expectations, especially NIH and leading global disease research organizations and foundations, requiring successful completion of researcher training courses at all levels, with refresher courses.

Underscore the Importance of the proper Experimental Design Strategy

- Improve academic training programs to emphasize and reinforce best practices in essential biomedical research skills, methods, technology, reagents and tools that produce reliability and accuracy and reduce likelihood of non-replicability.
- Emphasize necessity of scientific basic design principles, including randomization, blinding and statistical methods essential to achieve more reliable, better analyzed scientific results [52].
- Require regular training modules in animal ethics and proper handling in biomedical research, model selection, validation and characterization, and data collection and analysis.
- Require demonstrated competence in selecting and employing appropriate statistical methods for both appropriate experimental design and accurate data analysis.

Validate experimental reagents and reference materials

- Incentivize commercial reagent and instrument vendors to provision only validated biological research reagents (e.g., antibodies and cell lines) with utilization of such reagents by PIs as a documented best practice in research conduct.
- Require funding agencies to include explicit standards for study designs and experimental outcomes justified by documented use of validated research reagents, continued authentication throughout study duration, with funding dedicated for such additional costs.
- Define standard operating procedures for biological materials handling throughout the reagent and study lifecycle.
- Incentivize continued improvements in methods for biological reagent and whole organism validation using new genomics data and approaches.

Justify preclinical in vivo study design and conduct

- Scrutinize critical details of the actual human pathological complexity and condition (including heterogeneity of disease effects and outcomes) as necessary from the beginning to justify best animal model selection (if any), and essential physiological, genetic, and biochemical aspects of human disease and treatment.
- Characterize and report animal disease traits or symptoms and mortality, also using negative control treatments, and carefully note and report unexpected variations.
- Use mathematical models to guide experimental design, including how many animals must be included in any study to achieve reliable outcomes [119].
- Employ the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines and 20-point checklist for animal research reporting (from EQUATOR, the Enhancing the QUAlity and Transparency Of health Research network [120].
- Include information about effect size, sample size, rationale for subjects allocation to different study groups, and specific details about the strain of animals selected [121].
- Reverse-translate failed clinical approaches to guide improvement of preclinical animal model validity (https://pubmed.ncbi.nlm.nih.gov/25823810/)
- Avoid the "standardisation-translation paradox" in animal studies using syngeneic strains that compromise external validity. (https://doi.org/10.3390/ani10071170)
- Consider translational values of emerging alternative animal-free innovations and human-relevant biomedical science to validate better research strategies. https://doi.org/10.3390/ani10071170
- Incentivize and encourage both systematic and *meta*-analytical reviews of currently used animal models and the use of the resulting data within the research chain [122].

Improve Research Reporting Quality in Dissemination, Grant Writing and Publishing

- Require complete procedures that document reagent validation and lack of contamination, and complete preclinical animal experimental reporting details.
- Require all research reporting include expert analysis and discussion of experimental uncertainty in measurements and impact on conclusions.
- Require all research reporting include expert analysis and discussion of study limitations and impact on conclusions.
- Establish and implement standards of reproducibility and replicability to enforce as a recognized reporting priority

Better understand the translational pathway of biomedical basic to clinical research

- Consider the collaborative chain from basic research funding to patient application and understand the different "cultures of research" in this chain, and their modes of research engagement [52].
- Work with a "clinical champion" of the targeted human disease condition from the beginning to avoid "hammer seeking nail" issues in basic research relevance to the clinical challenge.
- Consider regulatory and intellectual property requirements along the entire performance pathway.

Encourage more effective, timely and less expensive clinical trial designs, reporting performance and access

- Require training of clinical researchers in modern clinical trial designs, patient recruitment/enrollment, expected reporting outcomes and format [123].
- Demand full and prompt reporting of clinical trial results regardless of outcome regardless of sponsor.
- Expect open data sharing of all human trials regardless of sponsor [107–110].
- Work with regulatory bodies to improve detailed testing protocols and evaluations for regulated products to better predict human performance and liabilities.

4. Conclusions

There are many challenges to successful translation in the biomedical research enterprise [3–8,101,102,125–127]. Research waste, lost opportunities for impact, and inefficiency are natural products of collective failure across the biomedical research enterprise to enact long-recognized changes necessary to improve translational processes and clinical impact. Scientists and clinical researchers operate in a global theatre in which biomedical research is governed by incentives that oppose such changes, enforcing the status quo to the detriment of best translational practices, patient welfare and quality of life, stewardship of resources and societal support, and scientific credibility. All of this ultimately focuses on the relationship between research rigor and researchers that conduct it, and methods that might actually improve it, instead of repeatedly using "lip-service" to describe the problem in many published critiques without changing the status quo.

A call for an improved "mind-set" and broader education of the next generation of biomedical and clinical researchers will in itself be insufficient to address data reproducibility, reporting, relevance and reliability challenges. Differences between academic and industrial missions, research strategies and conduct, and reward structures in biomedical research and translation must be appreciated and harmonized for translational congruence [124]. In some instances, a similar biomedical translational goal involving both academic and industry has resulted in surprisingly disparate intellectural property estates demonstrating orthogonal priorities and different translational strategies [94]. Stakeholder messaging and peer expectations for change provided to the translational research community must be consistent, persistent and focused. As global understanding of human disease mechanisms and biomarkers broadens and improves, periodic revisiting and critically evaluating 'standard' biomedical research models, and their expected deliverables, are critical. We must continually review whether existing models – in silico, *in vitro, ex vivo, in vivo* – and the underlying data and hypotheses that drive them remain valid as new data emerge. Continual re-evaluation and critique of research approaches, models and data reporting will continue to inform, but should be better enforced on the research community by diverse stakeholders, to evolve best practices. Towards this end, seeking actual clinical validity of novel interventions and therapies will be more convincing than continually improving pre-clinical models attempting to duplicate clinical reality. But engaging in such practices must also be reinforced by the proper research incentives to more efficiently steer the investigating and translating community and more reliably assess and report their medical utility and clinical benefit [124].

High quality biomedical evidence across the diverse different biomedical sources – molecular/cellular, preclinical whole organism, computational in silico, case series, clinical trials, *meta*-analyses, patient-clinician engagement, and societal evidence (e.g., clinical and patient advocacy groups) – must be reliably selected and supported, collected, vetted and fairly reported to achieve best evidence-based medical care. Numerous known research deficiencies preclude realization of a reliable, efficient biomedical research system. Critically, stakeholders must show the resolve and initiative to properly incentivize the research system to enable best practices systemically across the biomedical research spectrum. Only with dedicated and persistent focus on holistically improving the biomedical research process and resulting data quality and reliability emanating from "bench to bedside and back" will global biomedical translational efficiency and impact improve to benefit patient quality of life.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge valuable insights and expert guidance from numerous *Translate 2021!* conference (Berlin, Germany, 2021) partipants that influenced this work. Critical comments from U. Tölch, P. Ritter and J. Buechel, and graphics support from A. Ford are particularly appreciated.

References

[1] S.H. Woolf, The Meaning of Translational Research and Why It Matters, JAMA 299 (2) (2008), https://doi.org/10.1001/jama.2007.26.
[2] J. Kimmelman, J.S. Mogil, U. Dirnagl, D.R. Jones, Distinguishing between Exploratory and Confirmatory Preclinical Research Will Improve Translation, PLoS Biol. 12 (5) (2014) e1001863, https://doi.org/10.1371/journal.pbio.1001863.
[3] M. Abou-El-Enein, G.N. Duda, E.A. Gruskin, D.W. Grainger, Strategies for Derisking Translational Processes for Biomedical Technologies, Trends Biotechnol. 35 (2) (2017) 100–108, https://doi.org/10.1016/j.tibtech.2016.07.007.
[4] H. Hörig, E. Marincola, F.M. Marincola, Obstacles and opportunities in translational research, Nat. Med. 11 (7) (2005) 705–708, https://doi.org/10.1038/nm0705-705.
[5] D. Yu, Translational research: current status, challenges and future strategies, Am. J. Transl. Res. 3 (2011) 422–433.
[6] D.G. Contopoulos-Ioannidis, E.E. Ntzani, J.P.A. Ioannidis, Translation of highly promising basic science research into clinical applications, The American Journal of Medicine. 114 (6) (2003) 477–484, https://doi.org/10.1016/S0002-9343(03)00013-5.
[7] A.A. Seyhan, Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles, Translational Medicine Communications. 4 (2019) 18, https://doi.org/10.1186/s41231-019-0050-7.
[8] C.J. Vukotich, Challenges of T3 and T4 Translational Research, Journal of Research Practice. 12 (2016). http://jrp.icaap.org/index.php/jrp/article/view/552/454.
[9] T. Hartung, Look back in anger – what clinical studies tell us about preclinical work, ALTEX. 30 (3) (2013) 275–291.
[10] B. Fine, G. Vunjak-Novakovic, Shortcomings of Animal Models and the Rise of Engineered Human Cardiac Tissue, ACS Biomater. Sci. Eng. 3 (9) (2017) 1884–1897, https://doi.org/10.1021/acsbiomaterials.6b00662.
[11] J.P. Garner, B.N. Gaskill, E.M. Weber, J. Ahloy-Dallaire, K.R. Pritchett-Corning, Introducing Therioepistemology: the study of how knowledge is gained from animal research, Lab Animal. 46 (4) (2017) 103–113, https://doi.org/10.1038/laban.1224.
[12] M.L. Shuler, Organ-, body- and disease-on-a-chip systems, Lab Chip 17 (14) (2017) 2345–2346, https://doi.org/10.1039/C7LC90068F.
[13] D. Tallon, J. Chard, P. Dieppe, Relation between agendas of the research community and the research consumer, The Lancet. 355 (9220) (2000) 2037–2040, https://doi.org/10.1016/S0140-6736(00)02351-5.
[14] S.E. Gollust, J.W. Seymour, M.J. Pany, A. Goss, Z.F. Meisel, D. Grande, Mutual Distrust: Perspectives From Researchers and Policy Makers on the Research to Policy Gap in 2013 and Recommendations for the Future, INQUIRY: The Journal of Health Care Organization, Provision, and Financing. 54 (2017) 004695801770546. https://doi.org/10.1177/0046958017705465.
[15] About Translation | National Center for Advancing Translational Sciences, (n.d.). https://ncats.nih.gov/translation (accessed November 30, 2021).
[16] E.A. Zerhouni, US Biomedical Research, JAMA 294 (11) (2005) 1352, https://doi.org/10.1001/jama.294.11.1352.
[17] S. Homer-Vanniasinkam, J. Tsui, The Continuing Challenges of Translational Research: Clinician-Scientists' Perspective, Cardiology Research and Practice. 2012 (2012) 1–5, https://doi.org/10.1155/2012/246710.
[18] J.P.A. Ioannidis, Extrapolating from Animals to Humans, Sci. Transl. Med. 4 (151) (2012), https://doi.org/10.1126/scitranslmed.3004631.
[19] A. Knight, Systematic Reviews of Animal Experiments Demonstrate Poor Contributions Toward Human Healthcare, Rev. Recent Clin. Trials 3 (2008) 89–96, https://doi.org/10.2174/157488708784223844.
[20] P. Sandercock, I. Roberts, Systematic reviews of animal experiments, The Lancet. 360 (9333) (2002) 586, https://doi.org/10.1016/S0140-6736(02)09812-4.
[21] P. Perel, I. Roberts, E. Sena, P. Wheble, C. Briscoe, P. Sandercock, M. Macleod, L. E. Mignini, P. Jayaram, K.S. Khan, Comparison of treatment effects between

animal experiments and clinical trials: systematic review, BMJ 334 (7586) (2007) 197, https://doi.org/10.1136/bmj.39048.407928.BE.
[22] D.G. Hackam, D.A. Redelmeier, Translation of Research Evidence From Animals to Humans, JAMA 296 (14) (2006) 1727, https://doi.org/10.1001/jama.296.14.1731.
[23] J.P.A. Ioannidis, Contradicted and Initially Stronger Effects in Highly Cited Clinical Research, JAMA 294 (2) (2005) 218, https://doi.org/10.1001/jama.294.2.218.
[24] J.P.A. Ioannidis, Why Most Published Research Findings Are False, PLoS Med. 2 (8) (2005) e124, https://doi.org/10.1371/journal.pmed.0020124.
[25] C.H. Emmerich, L.M. Gamboa, M.C.J. Hofmann, M. Bonin-Andresen, O. Arbach, P. Schendel, B. Gerlach, K. Hempel, A. Bespalov, U. Dirnagl, M.J. Parnham, Improving target assessment in biomedical research: the GOT-IT recommendations, Nat. Rev. Drug Discovery 20 (1) (2021) 64–81, https://doi.org/10.1038/s41573-020-0087-3.
[26] D.G. Altman, The scandal of poor medical research, BMJ 308 (6924) (1994) 283–284, https://doi.org/10.1136/bmj.308.6924.283.
[27] N.S. Young, J.P.A. Ioannidis, O. Al-Ubaydli, Why Current Publication Practices May Distort Science, PLoS Med. 5 (10) (2008) e201, https://doi.org/10.1371/journal.pmed.0050201.
[28] Incentive malus | The Economist, (n.d.). https://www.economist.com/science-and-technology/2016/09/24/incentive-malus (accessed November 30, 2021).
[29] P.E. Smaldino, R. McElreath, The natural selection of bad science, R. Soc. Open Sci. 3 (9) (2016) 160384, https://doi.org/10.1098/rsos.160384.
[30] M.A. Edwards, S. Roy, Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition, Environ. Eng. Sci. 34 (1) (2017) 51–61, https://doi.org/10.1089/ees.2016.0223.
[31] D. Moher, F. Naudet, I.A. Cristea, F. Miedema, J.P.A. Ioannidis, S.N. Goodman, Assessing scientists for hiring, promotion, and tenure, PLOS Biology. 16 (2018) e2004089. https://doi.org/10.1371/journal.pbio.2004089.
[32] G. Pasterkamp, I. Hoefer, B. Prakken, Lost in the citation valley, Nat. Biotechnol. 34 (10) (2016) 1016–1018, https://doi.org/10.1038/nbt.3691.
[33] J.P.A. Ioannidis, How to Make More Published Research True, PLoS Med. 11 (10) (2014) e1001747, https://doi.org/10.1371/journal.pmed.1001747.
[34] R. Van Noorden, The science that's never been cited, Nature 552 (7684) (2017) 162–164, https://doi.org/10.1038/d41586-017-08404-0.
[35] S. Buranyi, Is the staggeringly profitable business of scientific publishing bad for science? | Science | The Guardian, (n.d.). https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science (accessed November 30, 2021).
[36] S. Dijkstra, G. Kok, J.G. Ledford, E. Sandalova, R. Stevelink, Possibilities and Pitfalls of Social Media for Translational Medicine, Frontiers in Medicine. 5 (2018), https://doi.org/10.3389/fmed.2018.00345.
[37] F.R.W. Kools, S. Mirali, S. Holst-Bernal, S.L. Nijhof, G. Cavalli, M.A. Grandner, Publications Are Not the Finish Line: Focusing on Societal Rather Than Publication Impact, Frontiers in Medicine. 5 (2018), https://doi.org/10.3389/fmed.2018.00314.
[38] P.M. Ridker, N. Rifai, Expanding Options for Scientific Publication, Circulation 127 (2) (2013) 155–156, https://doi.org/10.1161/CIRCULATIONAHA.112.155952.
[39] H. Aguinis, S.J. Vaschetto, Editorial Responsibility: Managing the Publishing Process to Do Good and Do Well, Management Organization Rev. 7 (3) (2011) 407–422, https://doi.org/10.1111/j.1740-8784.2011.00223.x.
[40] R.K. Merton, Priorities in Scientific Discovery: A chapter in the sociology of science, Am Soc Rev 22 (6) (1957) 635, https://doi.org/10.2307/2089193.
[41] T.S. Kuhn (1959) The Essential Tension: Tradition and Innovation in Scientific Research. University of Utah Research Conference on the Identification of Scientific Talent, ed. Taylor C (University of Utah Press, Salt Lake City), pp 162-174.
[42] J.G. Foster, A. Rzhetsky, J.A. Evans, Tradition and Innovation in Scientists' Research Strategies, Am Soc. Rev. (2015), https://doi.org/10.1177/0003122415601618.
[43] T.S. Kuhn, The Structure of Scientific Revolutions, University of Chicago Press, Chicago, 1962.
[44] J. Gernter, The Idea Factory: Bell Labs and the Great Age of American Innovation, Penguin Press, New York, 2012.
[45] J.P. Tennant, H. Crane, T. Crick, J. Davila, A. Enkhbayar, J. Havemann, B. Kramer, R. Martin, P. Masuzzo, A. Nobes, C. Rice, B. Rivera-López, T. Ross-Hellauer, S. Sattler, P.D. Thacker, M. Vanholsbeeck, Ten Hot Topics around Scholarly Publishing, Publications 7 (2) (2019) 34, https://doi.org/10.3390/publications7020034.
[46] National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and Replicability in Science. Washington, DC: The National Academies Press, page 110 https://doi.org/10.17226/25303.
[47] E. Wager, S. Kleinert, Responsible research publication: International standards for authors. A position statement developed at the 2nd World Conference on Research Integrity (Singapore, July 22–24, 2010), in Promoting Research Integrity in a Global Environment, 2011, eds. Mayer T, Steneck N (Imperial College Press/World Scientific Publishing, Singapore), pp 309–316.
[48] J. Diong, C.M. Kroeger, K.J. Reynolds, et al., Strengthening the incentives for responsible research practices in Australian health and medical research funding, Res. Integr. Peer Rev. 6 (2021) 11, https://doi.org/10.1186/s41073-021-00113-7.
[49] M., Bastasch, https://www.globalanimal.org/2013/10/07/guilty-government-practices/111386/#. (accessed 15 January 2022).
[50] N.H. Franco, Animal Experiments in Biomedical Research: A Historical Perspective, Animals (Basel). 3 (1) (2013) 238–273, https://doi.org/10.3390/ani3010238.
[51] P. Pound, M. Ritskes-Hoitinga, Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail, Journal of Translational Medicine. 16 (2018) 304, https://doi.org/10.1186/s12967-018-1678-1.
[52] C.H.C. Leenaars, C. Kouwenaar, F.R. Stafleu, A. Bleich, M. Ritskes-Hoitinga, R.B. M. De Vries, F.L.B. Meijboom, Animal to human translation: a systematic scoping review of reported concordance rates, J Transl Med. 17 (1) (2019), https://doi.org/10.1186/s12967-019-1976-2.
[53] M. Ritskes-Hoitinga, C. Leenaars, W. Beumer, T. Coenen-de Roo, F. Stafleu, F.L. B. Meijboom, Improving Translation by Identifying Evidence for More Human-Relevant Preclinical Strategies, Animals. 10 (7) (2020) 1170, https://doi.org/10.3390/ani10071170.
[54] C. Degeling, J. Johnson, Evaluating animal models: some taxonomic worries, J Med Philos. 38 (2) (2013) 91–106.
[55] L.P. Freedman, I.M. Cockburn, T.S. Simcoe, The Economics of Reproducibility in Preclinical Research, PLoS Biol. 13 (6) (2015) e1002165, https://doi.org/10.1371/journal.pbio.1002165.
[56] S.B. Green, Can animal data translate to innovations necessary for a new era of patient-centred and individualised healthcare? Bias in preclinical animal research, BMC Med Ethics. 16 (2015) 53, https://doi.org/10.1186/s12910-015-0043-7.
[57] A.J. Smith, R.E. Clutton, E. Lilley, K.E.A. Hansen, T. Brattelid, PREPARE: guidelines for planning animal research and testing, Lab Anim. 52 (2) (2018) 135–141, https://doi.org/10.1177/0023677217724823.
[58] C. Kilkenny, W.J. Browne, I.C. Cuthill, M. Emerson, D.G. Altman, Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research, PLoS Biol. 8 (6) (2010) e1000412, https://doi.org/10.1371/journal.pbio.1000412.
[59] J.P.A. Ioannidis, What Have We (Not) Learnt from Millions of Scientific Papers with P Values?, The American Statistician 73 (sup1) (2019) 20–25, https://doi.org/10.1080/00031305.2018.1447512.
[60] J.P.A. Ioannidis, The Proposal to Lower P Value Thresholds to.005, JAMA 319 (14) (2018) 1429, https://doi.org/10.1001/jama.2018.1536.
[61] G.A. Akerlof, P. Michaillat, Persistence of false paradigms in low-power sciences, Proc Natl Acad Sci USA 115 (52) (2018) 13228–13233.
[62] M.B. Bracken, Why animal studies are often poor predictors of human reactions to exposure, J. Royal Soc. Med. 102 (3) (2009) 120–122, https://doi.org/10.1258/jrsm.2008.08k033.
[63] D. Hammaker, G.S. Firestein, "Go upstream, young man": lessons learned from the p38 saga, Ann. Rheum. Dis. 69 (Suppl 1) (2010) i77–i82, https://doi.org/10.1136/ard.2009.119479.
[64] P. Zilla, D. Bezuidenhout, P. Human, Prosthetic vascular grafts: Wrong models, wrong questions and no healing, Biomaterials 28 (34) (2007) 5009–5027, https://doi.org/10.1016/j.biomaterials.2007.07.017.
[65] A. Kamb, What's wrong with our cancer models?, Nat Rev. Drug Discov. 4 (2) (2005) 161–165, https://doi.org/10.1038/nrd1635.
[66] C. Abboud, A. Duveau, R. Bouali-Benazzouz, K. Massé, J. Mattar, L. Brochoire, P. Fossat, E. Boué-Grabot, W. Hleihel, M. Landry, Animal models of pain: Diversity and benefits, J. Neurosci. Meth. 348 (2021) 108997, https://doi.org/10.1016/j.jneumeth.2020.108997.
[67] J.S. Mogil, Animal models of pain: progress and challenges, Nat. Rev. Neurosci. 10 (4) (2009) 283–294, https://doi.org/10.1038/nrn2606.
[68] A. Schmidt-Pogoda, N. Bonberg, M.H.M. Koecke, J.-K. Strecker, J. Wellmann, N.-M. Bruckmann, C. Beuker, W.-R. Schäbitz, S.G. Meuth, H. Wiendl, H. Minnerup, J. Minnerup, Why Most Acute Stroke Studies Are Positive in Animals but Not in Patients: A Systematic Comparison of Preclinical, Early Phase, and Phase 3 Clinical Trials of Neuroprotective Agents, Ann. Neurol. 87 (1) (2020) 40–51, https://doi.org/10.1002/ana.25643.
[69] U. Dirnagl, Thomas Willis Lecture, Stroke 47 (8) (2016) 2148–2153, https://doi.org/10.1161/STROKEAHA.116.013244.
[70] J.P.A. Ioannidis, B.Y.S. Kim, A. Trounson, How to design preclinical studies in nanomedicine and cell therapy to maximize the prospects of clinical translation, Nature Biomed. Eng. 2 (11) (2018) 797–809, https://doi.org/10.1038/s41551-018-0314-y.
[71] C.B. Little, D.J. Hunter, Post-traumatic osteoarthritis: from mouse models to clinical trials, Nature Rev. Rheumatol. 9 (8) (2013) 485–497, https://doi.org/10.1038/nrrheum.2013.72.
[72] M.J. Makarczyk, Q.i. Gao, Y. He, Z. Li, M.S. Gold, M.C. Hochberg, B.A. Bunnell, R. S. Tuan, S.B. Goodman, H. Lin, Current Models for Development of Disease-Modifying Osteoarthritis Drugs, Tissue Engineering Part C, Methods 27 (2) (2021) 124–138, https://doi.org/10.1089/ten.tec.2020.0309.
[73] K.A. Taylor, M. Emerson, Refinement of a mouse cardiovascular model: Development, application and dissemination, F1000 Res. 7 (2018) 593. https://doi.org/10.12688/f1000research.14456.1.
[74] J.A. Auer, A. Goodship, S. Arnoczky, S. Pearce, J. Price, L. Claes, B. von Rechenberg, M. Hofmann-Amtenbrinck, E. Schneider, R. Müller-Terpitz, F. Thiele, K.-P. Rippe, D.W. Grainger, Refining animal models in fracture research: seeking consensus in optimising both animal welfare and scientific validity for appropriate biomedical use, BMC Musculoskeletal Disorders. 8 (2007) 72, https://doi.org/10.1186/1471-2474-8-72.

[75] G.S. Ferreira, D.H. Veening-Griffioen, W.P.C. Boon, E.H.M. Moors, P.J.K. van Meer, Levelling the Translational Gap for Animal to Human Efficacy Data, Animals. 10 (2020) 1199, https://doi.org/10.3390/ani10071199.

[76] Reproducibility and Replicability in Science, National Academies Press, Washington, D.C., 2019. https://doi.org/10.17226/25303.

[77] J. Menke, M. Roelandse, B. Ozyurt, M. Martone, A. Bandrowski, The Rigor and Transparency Index Quality Metric for Assessing Biological and Medical Science Methods, IScience. 23 (11) (2020) 101698, https://doi.org/10.1016/j.isci.2020.101698.

[78] D. Colquhoun, An investigation of the false discovery rate and the misinterpretation of p -values, Royal Soc. Open Sci. 1 (3) (2014) 140216, https://doi.org/10.1098/rsos.140216.

[79] D. Chavalarias, J.D. Wallach, A.H.T. Li, J.P.A. Ioannidis, Evolution of Reporting P Values in the Biomedical Literature, 1990–2015, JAMA 315 (11) (2016) 1141, https://doi.org/10.1001/jama.2016.1952.

[80] R. L. Wasserstein, N.A Lazar, (2016) The ASA Statement on p-Values: Context, Process, and Purpose, Am Stat., 70:2, 129 133, DOI: 10.1080/00031305.2016.1154108.

[81] R.L. Wasserstein, A.L. Schirm, N.A. Lazar (2019) Moving to a World Beyond "p < 0.05", Am. Stat., 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

[82] C.H. Bucher, C. Schlundt, D. Wulsten, F.A. Sass, S. Wendler, A. Ellinghaus, T. Thiele, R. Seemann, B.M. Willie, H.-D. Volk, G.N. Duda, K. Schmidt-Bleek, Experience in the Adaptive Immunity Impacts Bone Homeostasis, Remodeling, and Healing, Front. Immunol. 10 (2019), https://doi.org/10.3389/fimmu.2019.00797.

[83] B. Voelkl, L. Vogt, E.S. Sena, H. Würbel, Reproducibility of preclinical animal research improves with heterogeneity of study samples, PLOS Biol 16 (2018) e2003693, https://doi.org/10.1371/journal.pbio.2003693.

[84] E. Rybalka, C. Timpani, D. Debruin, R. Bagaric, D. Campelj, A. Hayes, The Failed Clinical Story of Myostatin Inhibitors against Duchenne Muscular Dystrophy: Exploring the Biology behind the Battle, Cells. 9 (2020) 2657, https://doi.org/10.3390/cells9122657.

[85] C.S. Constantinescu, N. Farooqi, K. O'Brien, B. Gran, Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS), Br. J. Pharmacol. 164 (2011) 1079–1106, https://doi.org/10.1111/j.1476-5381.2011.01302.x.

[86] J.R. Swearengen, Choosing the right animal model for infectious disease research, Animal Models and Experimental Medicine. 1 (2) (2018) 100–108, https://doi.org/10.1002/ame2.12020.

[87] A. Dyson, M. Singer, Animal models of sepsis: Why does preclinical efficacy fail to translate to the clinical setting?, Crit Care Med. 37 (Supplement) (2009) S30–S37, https://doi.org/10.1097/CCM.0b013e3181922bd3.

[88] C.I. Kingsley, S.N. Nadig, K.J. Wood, Transplantation tolerance: lessons from experimental rodent models, Transpl. Int. 20 (10) (2007) 828–841, https://doi.org/10.1111/j.1432-2277.2007.00533.x.

[89] P. Li, E.M. Schwarz, The TNF-alpha transgenic mouse model of inflammatory arthritis, Springer Semin. Immunopathol. 25 (2003) 19–33, https://doi.org/10.1007/s00281-003-0125-3.

[90] R. Diaz Brinton, Minireview: Translational Animal Models of Human Menopause: Challenges and Emerging Opportunities, Endocrinology. 153 (2012) 3571–3578. https://doi.org/10.1210/en.2012-1340.

[91] S. Reinwald, D. Burr, Review of Nonprimate, Large Animal Models for Osteoporosis Research, J. Bone Min. Res. 23 (9) (2008) 1353–1368, https://doi.org/10.1359/jbmr.080516.

[92] L.E. Mignini, K.S. Khan, Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research, BMC Med. Res. Method. 6 (2006) 10, https://doi.org/10.1186/1471-2288-6-10.

[93] A. Garanto, S.E.C. van Beersum, T.A. Peters, R. Roepman, F.P.M. Cremers, R.W.J. Collin, T. Li, Unexpected CEP290 mRNA splicing in a humanized knock-in mouse model for Leber congenital amaurosis, PLoS ONE 8 (11) (2013) e79369, https://doi.org/10.1371/journal.pone.0079369.

[94] A. Garanto, L. Duijkers, R.W.J. Collin, Species-dependent splice recognition of a cryptic exon resulting from a recurrent intronic CEP290 mutation that causes congenital blindness, Int. J. Mol. Sci. 16 (2015) 5285–5298, https://doi.org/10.3390/ijms16035285.

[95] A.V. Cideciyan, S.G. Jacobson, A.V. Drack, A.C. Ho, J. Charng, A.V. Garafalo, A.J. Roman, A. Sumaroka, I.C. Han, M.D. Hochstedler, W.L. Pfeifer, E.H. Sohn, M. Taiel, M.R. Schwartz, P. Biasutto, W.d. Wit, M.E. Cheetham, P. Adamson, D.M. Rodman, G. Platenburg, M.D. Tome, I. Balikova, F. Nerinckx, J.D. Zaeytijd, C. Van Cauwenbergh, B.P. Leroy, S.R. Russell, Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect, Nat. Med. 25 (2) (2019) 225–228, https://doi.org/10.1038/s41591-018-0295-0.

[96] K. Dulla, M. Aguila, A. Lane, K. Jovanovic, D.A. Parfitt, I. Schulkens, H.L. Chan, I. Schmidt, W. Beumer, L. Vorthoren, R.W.J. Collin, A.V. Garanto, L. Duijkers, A. Brugulat-Panes, M. Semo, A.A. Vugler, P. Biasutto, P. Adamson, M.E. Cheetham, Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models, Mol. Ther. Nucleic Acids. 12 (2018) 730–740, https://doi.org/10.1016/j.omtn.2018.07.010.).

[97] J. Masison, J. Beezley, Y. Mei, H.A.L. Ribeiro, A.C. Knapp, L. Sordo Vieira, B. Adhikari, Y. Scindia, M. Grauer, B. Helba, W. Schroeder, B.R. Laubenbacher Mehrad, A modular computational framework for medical digital twins, Proc. Natl. Acad. Sci., USA 118 (20) (2021), e2024287118, https://doi.org/10.1073/pnas.2024287118.

[98] J.C. Madden, S.J. Enoch, A. Paini, M.T.D. Cronin, A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications. Altern Lab Anim. 48 (4) (2020) 146–172, https://doi.org/10.1177/0261192920965977.

[99] Q. Wu, J. Liu, X. Wang, L. Feng, J. Wu, X. Zhu, W. Wen, X. Gong, Organ-on-a-chip: recent breakthroughs and future prospects, BioMed Eng OnLine 19 (1) (2020), https://doi.org/10.1186/s12938-020-0752-0.

[100] M. Abou-el-Enein, A. Angelis, F.R. Appelbaum, N.C. Andrews, S.E. Bates, A.S. Bierman, M.K. Brenner, M. Cavazzana, M.A. Caligiuri, H. Clevers, E. Cooke, G.Q. Daley, V.J. Dzau, L.M. Ellis, H.V. Fineberg, L.S.B. Goldstein, S. Gottschalk, M.A. Hamburg, D.E. Ingber, D.B. Kohn, A.R. Krainer, M.V. Maus, P. Marks, C.L. Mummery, R.I. Pettigrew, J.L. Rutter, S.A. Teichmann, A. Terzic, F.D. Urnov, D. A. Williams, J.D. Wolchok, M. Lawler, C.J. Turtle, G. Bauer, J.P.A. Ioannidis, Evidence generation and reproducibility in cell and gene therapy research: A call to action, Mol. Ther. Methods Clin. Dev. 22 (2021) 11–14, https://doi.org/10.1016/j.omtm.2021.06.012.

[101] H.-D. Volk, M.M. Stevens, D.J. Mooney, D.W. Grainger, G.N. Duda, Key elements for nourishing the translational research environment, Sci. Transl. Med. 7 (282) (2015) 282cm2, https://doi.org/10.1126/scitranslmed.aaa2049.

[102] G.N. Duda, D.W. Grainger, M.L. Frisk, L. Bruckner-Tuderman, A. Carr, U. Dirnagl, K.M. Einhäupl, S. Gottschalk, E. Gruskin, C. Huber, C.H. June, D.J. Mooney, E.T. Rietschel, G. Schütte, W. Seeger, M.M. Stevens, R. Urban, A. Veldman, G. Wess, H.-D. Volk, Changing the Mindset in Life Sciences Toward Translation: A Consensus, Sci. Transl. Med. 6 (264) (2014) 264cm12, https://doi.org/10.1126/scitranslmed.aaa0599.

[103] A. Deaton, N. Cartwright, Understanding and misunderstanding randomized controlled trials, Soc. Sci. Med. 210 (2018) 2–21, https://doi.org/10.1016/j.socscimed.2017.12.005.

[104] J.P.A. Ioannidis, Why Most Clinical Research Is Not Useful, PLoS Med. 13 (6) (2016) e1002049, https://doi.org/10.1371/journal.pmed.1002049.

[105] I. Boutron, P. Ravaud, Spin in biomedical literature, Proc. Natl. Acad. Sci. USA 115 (11) (2018) 2613–2619, https://doi.org/10.1073/pnas.1710755115.

[106] V.L. Bartlett, S.S. Dhruva, N.D. Shah, P. Ryan, J.S. Ross, Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence, JAMA Network Open. 2 (10) (2019) e1912869, https://doi.org/10.1001/jamanetworkopen.2019.12869.

[107] A. Fortunato, D.W. Grainger, M. Abou-El-Enein, Enhancing patient-level clinical data access to promote evidence-based practice and incentivize therapeutic innovation, Adv. Drug Deliv. Rev. 136–137 (2018) 97–104, https://doi.org/10.1016/j.addr.2018.01.017.

[108] P. Doshi, S.N. Goodman, J.P.A. Ioannidis, Raw data from clinical trials: within reach?, Trends Pharmacol Sci. 34 (12) (2013) 645–647, https://doi.org/10.1016/j.tips.2013.10.006.

[109] D.A. Zarin, Participant-Level Data and the New Frontier in Trial Transparency, N. Engl. J. Med. 369 (5) (2013) 468–469, https://doi.org/10.1056/NEJMe1307268.

[110] R. Ramachandran, C.J. Morten, J.S. Ross, Strengthening the FDA's Enforcement of ClinicalTrials.gov Reporting Requirements, JAMA 326 (21) (2021) 2131, https://doi.org/10.1001/jama.2021.19773.

[111] M.H. Murad, H. Chu, L. Lin, Z. Wang, The effect of publication bias magnitude and direction on the certainty in evidence, BMJ Evidence-Based Medicine. 23 (3) (2018) 84–86, https://doi.org/10.1136/bmjebm-2018-110891.

[112] F. Alahdab, W. Farah, J. Almasri, P. Barrionuevo, F. Zaiem, R. Benkhadra, N. Asi, M. Alsawas, Y. Pang, A.T. Ahmed, T. Rajjo, A. Kanwar, K. Benkhadra, Z. Razouki, M.H. Murad, Z. Wang, Treatment Effect in Earlier Trials of Patients With Chronic Medical Conditions: A Meta-Epidemiologic Study, Mayo Clin. Proc. 93 (3) (2018) 278–283, https://doi.org/10.1016/j.mayocp.2017.10.020.

[113] K.C. Siontis, E. Evangelou, J.P. Ioannidis, Magnitude of effects in clinical trials published in high-impact general medical journals, Int. J. Epidemiol. 40 (2011) 1280–1291, https://doi.org/10.1093/ije/dyr095.

[114] S. Hopewell, K. Loudon, M.J. Clarke, A.D. Oxman, K. Dickersin, Publication bias in clinical trials due to statistical significance or direction of trial results, Cochrane Database Systemat. Rev. 2010 (2009), https://doi.org/10.1002/14651858.MR000006.pub3.

[115] G.H. Guyatt, A.D. Oxman, V. Montori, G. Vist, R. Kunz, J. Brozek, P. Alonso-Coello, B. Djulbegovic, D. Atkins, Y. Falck-Ytter, J.W. Williams, J. Meerpohl, S.L. Norris, E.A. Akl, H.J. Schünemann, GRADE guidelines: 5. Rating the quality of evidence—publication bias, J. Clin. Epidemiol. 64 (2011) 1277–1282, https://doi.org/10.1016/j.jclinepi.2011.01.011.

[116] R.V. Ghaemi, L.C. Siang, V.G. Yadav, Improving the Rate of Translation of Tissue Engineering Products, Adv. Healthcare Mater. 8 (19) (2019) 1900538, https://doi.org/10.1002/adhm.201900538.

[117] M. Steedman, K. Taylor, M. Stockbridge, M. Joao Cruz, S. Shah, W. Miranda, Ten years on measuring the return from pharmaceutical innovation 2019. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-ten-years-on-measuring-return-on-pharma-innovation-report-2019.pdf (accessed November 30, 2021).

[118] V.B. Kraus, Biomarkers as drug development tools: discovery, validation, qualification and use, Nature Rev. Rheumatol. 14 (6) (2018) 354–362, https://doi.org/10.1038/s41584-018-0005-9.

[119] S. Perrin, Preclinical research: Make mouse studies work, Nature 507 (7493) (2014) 423–425, https://doi.org/10.1038/507423a.

[120] https://www.equator-network.org/reporting-guidelines, accessed January, 2022.

[121] Centre for Replacement Refinement & Reduction of Animals in Research, https://nc3rs.org.uk/sites/default/files/documents/Guidelines/NC3Rs%20ARRIVE%20Guidelines%202013.pdf, accessed January 2022.
[122] Z. Bahadoran, P. Mirmiran, K. Kashfi, A. Ghasemi, Importance of Systematic Reviews and Meta-analyses of Animal Studies: Challenges for Animal-to-Human Translation, J Am Assoc Lab Anim Sci. 59 (5) (2020) 469–477, https://doi.org/10.30802/AALAS-JAALAS-19-000139.
[123] J. Williams, L. Weekes, S. Harvey, A. Kumar, M. Leigh, F. Thiele, D. Sabanathan, Clinical trials best practice checklist: Guidance for Australian clinical research sites from CT:IQ, Contemp Clin Trials Commun 20 (2020) 100651, https://doi.org/10.1016/j.conctc.2020.100651.
[124] R. Ogier, W. Knecht, M.E. Schwab, Academic leadership: (with)holding the keys to translational medicine?, Nat Med. 25 (12) (2019) 1812–1813, https://doi.org/10.1038/s41591-019-0670-5.
[125] T. Greenhalgh, J. Raftery, S. Hanney, M. Glover, Research impact: a narrative review, BMC Medicine. 14 (2016) 78, https://doi.org/10.1186/s12916-016-0620-8.
[126] Z.S. Morris, S. Wooding, J. Grant, The answer is 17 years, what is the question: understanding time lags in translational research, J. Royal Soc. Med. 104 (12) (2011) 510–520, https://doi.org/10.1258/jrsm.2011.110180.
[127] S.R. Hanney, S. Castle-Clarke, J. Grant, S. Guthrie, C. Henshall, J. Mestre-Ferrandiz, M. Pistollato, A. Pollitt, J. Sussex, S. Wooding, How long does biomedical research take?, Studying the time taken between biomedical and health research and its translation into products, policy, and practice, Health Research Policy and Systems 13 (2015) 1, https://doi.org/10.1186/1478-4505-13-1.