This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Diversity And Inclusion for Rodents: How Animal Ethics Committees Can Help Improve Translation

By Monika Piotrowska

Abstract: Translation failure occurs when a treatment shown to be safe and effective in one type of population does not produce the same result in another. We are currently in a crisis involving the translatability of preclinical studies to human populations. Animal trials are no better than a coin toss at predicting the safety and efficacy of drugs in human trials and the high failure rate of drugs entering human trials suggests most of the suffering of laboratory animals is futile, creating no commensurate benefit for human patients. Here, I argue that animal ethics committees have a role to play in getting us out of this crisis. Inadequate representation is a known contributor to translation failures and is a matter of both scientific and ethical concern. Ethical review committees have the authority to address it by reprioritizing the values already enshrined in their guiding principles.

1. Introduction

The high failure rate of drugs entering human trials suggests laboratory animals are not being effectively used in preclinical research, and that their suffering does not usually contribute to clinical benefit.[see 1, 2] Even so, animal ethics committees designed to ensure the responsible use of animals in research are stuck in a pattern of approval, giving the thumbs up to proposals regardless of the anticipated value of the research for human patients. One reason for this acquiescence is that in the U.S., committees such as Institutional Animal Care and Use Committee (IACUC) and Institutional Review Board (IRB) are expected to "stay in their lane" and concentrate on their specific areas of responsibility, overseeing the ethics of research involving animal and human subjects, respectively. In other words, ethical review committees are expected to review ethical issues while leaving scientific review to the scientists.[see 3, 4] As a result, issues like the probability of translation and clinical relevance are left to the scientists, not the ethicists, because it is the former who are qualified to assess such matters. But animal

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

trials are no better than a coin toss at predicting the safety and efficacy of drugs in human trials [5], and translation failures continue to occur. This suggests a failure of assessment with moral implications. Indeed, Joseph Garner has called translation failure, "the greatest laboratory animal welfare issue of our day."[6, p.349][1]

To break the pattern of ethical review committees rubber stamping risky or harmful research regardless of potential translation failures, authors working on the preclinical side of these issues have suggested adding new principles to existing ethical guidelines. These principles would explicitly allow animal ethics committees to assess the scientific merit of proposals.[8-10] But such suggestions are problematic since many committee members lack the necessary expertise to conduct scientific merit reviews. To resolve this general problem, authors have suggested restructuring the composition of animal ethics committees to ensure they are composed of members with the requisite scientific knowledge [11] or overhauling the entire review process [12]. While these suggestions are promising and might be an effective long-term solution, implementation will require significant effort, political will, and time.

Authors working on the clinical side of these issues [see 4, 13] have proposed a quicker and easier solution. Instead of restructuring review boards or overhauling the review process to evaluate the scientific merit of research proposals, these authors have suggested that IRBs have an inherent ethical responsibility to address the issue of translation failure. The issue is, as it were, already within the wheelhouse of ethical review boards because inadequate representation in clinical trials is a major contributor to translation failures and improving diversity and

[1] One response to failures in translation has been to develop alternative nonanimal models. But as Margaret Landi and colleagues argue, reducing our dependence on animals should not take priority over improving the human relevance of animal models. They write, "the complexity of mammalian biology will remain elusive to even the most sophisticated modeling systems, requiring that we still conduct scientifically defensible and judicious animal research."[7, p. 61]

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

inclusion is an ethical issue. I agree and aim to argue here that animal ethics committees can adopt a similar approach. Just as IRBs have an inherent responsibility to consider diversity and inclusion, IACUCs have a similar responsibility, one that supports the authority of IACUCs to address translation failures as an ethical issue. In short, I argue that insufficient representation in preclinical trials—as a significant contributor to translation failures—supports the obligation of animal ethics committees to address this issue.[2]

The paper unfolds as follows: In the next section, I discuss the issue of underrepresentation in research, its causes, and persistence. Subsequently, I argue that underrepresentation is a matter of both scientific and ethical concern, and that ethical review committees have the authority to address it. This requires prioritizing the values already enshrined in the guiding principles of ethical review boards. Finally, I contend that, while improving representation in animal research presents unique challenges, it is relatively easier to address than underrepresentation in human trials.

2. Causes and Consequences of Underrepresentation

Discrimination based on race, color, religion, national origin, or sex has been illegal in federally funded U.S. clinical trials since the Civil Rights Act of 1964, although the initial enforcement of the act was partial and inconsistent. The result was the overrepresentation of vulnerable populations in clinical research. This fact gained national attention in 1972 when it was revealed that the 40-year Tuskegee Syphilis Study, funded by the government and conducted by the U.S. Public Health Service, involved untreated low-income African American men who

[2] Insufficient representation is a significant contributor to translation failures, but there are others. For a comprehensive list see [14].

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

continued to be research subjects long after effective treatment for syphilis had become available.[15] In response, the U.S. Congress passed the National Research Act of 1974, which established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The commission was tasked with laying a foundation for determining when research on human subjects could be ethical. After reviewing the evidence, the commission made the following recommendation:

> [T]he selection of research subjects needs to be scrutinized to determine whether some classes (e.g., welfare patients, particular racial and ethnic minority population groups, or persons confined to institutions) are being systematically selected simply because of their easy availability, their compromised position, or their manipulability, rather than for reasons directly related to the problem being studied.[16]

The idea that burdens of research should not unduly fall to members of vulnerable populations is embodied in the principle of justice, which was one of three principles proposed by the Belmont Report.[16] The other two principles were respect for persons, which mandates voluntary and informed consent from capable research participants and the protection of those incapable of providing such consent, and beneficence, which requires minimizing possible harms and maximizing possible benefits of research. Together, these three principles—respect for persons, beneficence, and justice—are designed to ensure that biomedical and behavioral research involving human subjects is conducted ethically.

Although the Belmont Report was well-intentioned, the principle of justice it articulated has inadvertently contributed to the problem of inadequate representation in clinical research.[4]

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Indeed, responding to historical abuses and unfair distribution of past burdens by being overly cautious about recruiting members of racial and ethnic minorities has led to the underrepresentation of certain groups in clinical research. This itself has had unintended harmful consequences for members of those groups. A notable example of excessive caution leading to harmful results is warfarin, a medication used to prevent blood clots. Although warfarin was approved for human use in the early 1950s, dosing algorithms were based on evidence from clinical trials that enrolled subjects predominantly of European ancestry, and the findings didn't generalize to the more diverse U.S. population. As a result, warfarin became known for its adverse effects over the next 60 years of its distribution, including increased risk of bleeding, hospitalization, and death, due to the 20-fold interpatient variability in therapeutic dose requirements.[17] About half of the variability can be explained by genetic differences that map onto genetic ancestries, with populations of greater genetic African ancestry requiring a higher daily dose of warfarin and populations of greater genetic Asian ancestry requiring a lower dose than the homogenous European population enrolled in the clinical trials. In fact, because of the clinical trial population's homogeneity, the varied response to warfarin did not register and its effects on a more varied population could not be known.

The example highlights the limited generalizability of study findings when certain populations are inadequately represented in clinical trials. Given that populations excluded from trials can have distinct disease presentations or health circumstances that can affect their response to treatment [see 18, Chapter 2], clinical trial participants should be diverse enough to capture those variabilities. Excessive cautiousness in participant selection can undermine this demand.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Inadequate representation is also a problem in preclinical trials. While many exclusions in clinical trials arise from attempts to prevent historical abuses and injustices from recurring,[3] inadequate representation in preclinical trials is almost always intentional and designed to bring about scientific rigor. During the early stages of discovery, scientists eliminate potential confounding variables by using similar test subjects and keeping them in similar environments, which is standard practice in animal research. This allows researchers working with laboratory animals to eliminate variation to a much greater extent than what can be done in human trials. Even so, continuing to eliminate variation in the confirmatory preclinical stage can decrease the generalizability of findings, similar to inadequate representation in clinical trials.

Consider a couple of examples. First, almost all animal experiments that inform clinical trials employ rodents and their homogeneity extends beyond class and species membership. Indeed, most are also genetically identical. The most common strain of mice used in U.S. laboratories is Black-6 (C57BL/6) [21], which is a highly inbred strain due to the far greater number of generations of brother and sister (or parent and offspring) matings than other strains of mice, making it essentially genetically homozygous at all loci. To make things even more uniform, until recently, scientists have been excluding female mice whose fluctuating hormone levels may add unnecessary noise.[4] The epistemic advantage of using genetically uniform males is that any difference in experimental outcomes cannot be attributed to genetic differences or sex differences but (ideally) to the therapies being tested. Second, to ensure nothing else can account

[3] Many but not all. There is also a history of exclusions for the sake of scientific rigor in clinical trials. For example, women were often excluded from clinical trials because their hormonal cycles were deemed to introduce unwanted variability [see 19], and participants with co-morbidities continue to be excluded because their preexisting health conditions might interfere with study outcomes [see 20].

[4] This trend has recently come to an end in the U.S., at least for federally funded research. In 2016, the U.S. National Institutes of Health mandated that grant recipients must use both sexes in animal studies.[22] Similar policies were previously passed by the Canadian Institutes of Health Research and the European Commission.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

for differences in outcome, scientists also standardize the environmental conditions of their experiments. Although eliminating all potential confounds has been shown to be an unachievable ideal [23], eliminating most differences increases the odds of being able to measure the effect produced by the intervention of interest. Therefore, rodents are housed in identical cages, stored in rooms with identical temperatures, and fed the same diets, among other things.

The problem, however, is that attempting to eliminate all variation in preclinical test subjects rests on the mistaken assumption that there is a single, pure treatment effect that can be measured once all differences have been removed. There is no such effect (other than the one generated by the interplay between a particular set of genes and environmental conditions).[see 24, 25] When scientists conduct preclinical trials in a controlled setting with uniform test subjects, they are measuring a specific norm of reaction. They are discovering that under certain conditions, a particular genotype is more likely to produce a particular effect when given a specific treatment. But gene-environment interactions are complex and treatment effects that remain consistent across significant biological variations are rare. The upshot is that homogenous preclinical trials generate limited, localized "truths" that only provide insights into expected outcomes under narrowly defined conditions.[26] The point for our purposes is that even though eliminating variation may be useful for eliminating extraneous noise and exploring new ideas, excessive zeal in eliminating variation during the confirmatory preclinical stage may hinder translation and increase translation failure. This is because variation is a fundamental aspect of biology, and if we expect preclinical trials to generalize to human patients, they must be more representative of biological reality.

Garner suggests thinking of rodents as human patients to help us see the ways in which uniformity in preclinical trials hinders translation. He writes:

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

> Is our animal population a realistic representation of human variability? Humans are variable. Indeed it is our variability in risk for illness and response to treatment that is the focus of modern medicine…We would never perform a human drug trail in 42-year-old white males with identical educational levels, identical socioeconomic statuses, identical jobs, identical houses with identical (locked) thermostats, identical wives, identical diets, identical exercise regimes, in the same small town in Wisconsin, who all incidentally had the same grandfather. So can we realistically expect mice in exactly this kind of Stepford experiment to tell us anything about humans in general, or variability in risk or response in particular?[6, p. 442]

Garner's imaginative exercise exposes how seemingly reasonable efforts to eliminate confounding variables are likely hurting our ability to generalize findings to human patients. Even so, it's not as if creating experimental conditions that mimic human lives would resolve the issue of translation failures. Indeed, the very use of rodents in research could not possibly create the kind of perfect mimicry of human biology or disease needed to resolve the issue. But by creating experiments that are more closely representative of varied biological reality we can check if differences in the subject and/or environment will affect treatment outcomes.[see 24, 26] Such information can, in turn, inform our judgments about the likelihood of translation.

3. The Ethics of Representation

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

If incorporating variation into preclinical research to improve translation has widely recognized benefits, why do proposals that rely on homogeneous populations still receive ethical approval? For one thing, the composition of study participants and their environments are part of experimental design. Ethics committees are not meant to comment on scientific design, only on ethical considerations. Even the Belmont Report, from our earlier example, states that the investigators are the ones responsible for assessing whether their proposed research is properly designed. In contrast, the role of ethical review committees is to determine whether the risks associated with the proposed research design are outweighed by the potential benefits for research subjects.

Similar ethical considerations are meant to inform the deliberations of IACUCs. Although William Russell and Rex Burch's 3Rs [27]—Reduce, Refine, and Replace—are supposed to ensure the humane design of animal research studies, the 1985 amendments of the U.S. Animal Welfare Act encourage IACUCs to stay in their lane and leave experimental design to researchers.[28] Indeed, to the extent that IACUCs scrutinize experimental design, it's generally only with the aim of determining whether less painful or distressing alternatives are available to achieve the same experimental results. In practice, this means that IACUCs are much more likely to request a protocol change if the issue is within the realm of veterinary care than if it's within the realm of experimental design and analysis, even though scrutinizing the latter can have greater benefits for the animals than the former.[see 7, 29] The result is that preclinical experimental designs are widely approved by animal ethics committees, even if the research is harmful or risky to the animals and unlikely to benefit human patients, and the inability (or reluctance) of committees to assess experimental design may be a contributing factor in this.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Indeed, this inability has given rise to a pattern of approval where ethical review committees approve research proposals without justifying the burdens placed on the subjects by experimental design. Margaret Waltz and colleagues argue that ethical review committees lack guidance on how to break out of this pattern, which compounds the problem.

> [W]hile the mission of both IRBs and IACUCs require, in different ways, assurance of the quality and value of the science, neither group is given explicit guidance regarding how to accomplish this key ethical goal as part of their evaluation of research protocols.[30, p. 40]

Based on recent empirical findings by these authors, neither IRBs nor IACUCs are conducting overall ethical assessments of the value of human and animal research.[30] The fact that they are told to stay in their lane may be a significant part of this problem. It is therefore understandable that the preclinical side's response has been to call for new principles to be added to existing guidelines, which would explicitly grant IACUCs the authority to evaluate the scientific merit of proposals.

But granting review committees authority to review the scientific merit of proposals may not resolve the issue, since there are other explanations for why ethical review committees continue to approve research that is unlikely to benefit human patients. One of these alternative reasons has less to do with the separation of science and ethics and more to do with how the committees interpret their guiding principles. Indeed, as I argue below, the committees may be limiting themselves from exercising their power by overemphasizing certain principles while neglecting others.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Let's start with clinical ethical review. As previously mentioned, IRBs follow the guiding principles of the Belmont Report, which include respect for persons, beneficence, and justice. These principles are critical as even research conducted on willing participants entails inherent risks. To address these risks, the Belmont Report poses the critical question, "Who ought to receive the benefits of research and bear its burdens?"[16] Under the principle of justice, there is significant emphasis given to protecting those deemed most vulnerable. Given the history of exploitation in research involving human subjects, IRBs have an obligation to prevent vulnerable populations from bearing the brunt of research burdens. However, the equitable distribution of burdens should be complemented by the equitable distribution of benefits, as outlined in the Belmont Report. Excluding vulnerable populations from research, even for their own protection, may lead to the unfair distribution of benefits.

There are several reasons for this. First, excluding groups from research denies them immediate access to benefits, particularly in cases where the research could potentially benefit participants. Second, underrepresented groups being excluded from clinical trials may result in them being denied access to the benefits of research as the research may not apply to them.[see 18, 31] Thus, IRBs are faced with competing demands when it comes to ensuring the equitable distribution of burdens and benefits. On the one hand, as Phoebe Friesen and colleagues note, “IRBs are increasingly aware and invested in the importance of inclusion, especially as it relates to ensuring good, generalizable, and clinically applicable, and therefore ethical, science.”[4, p. 6] After all, inclusion contributes to equitable benefits.[5] On the other hand, IRBs are responsible for

[5] There are a few exceptions where inclusion does not necessary contribute to equitable benefits. According to David Strauss and colleagues, those exceptions include, “small exploratory, proof-of-concept, early phase studies, or research that seeks to learn about specific communities.”[13, p. 1211]

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

excluding vulnerable populations from possible exploitation, thereby ensuring the equitable distribution of research burdens.

The demographics of U.S. trial sites suggest that the well-justified reasons for inclusion are currently underemphasized by IRBs, while the protection of vulnerable research participants is overemphasized.[32] As Friesen and colleagues have argued, IRBs can strike a better balance between inclusion and protection by ensuring the equitable distribution of burdens and benefits of research as a matter of justice.[4][6] If the equitable distribution of benefits is directly tied to the problem of underrepresentation, it is within the purview of IRBs to address underrepresentation in experimental design by demanding better inclusion and diversity in clinical research without involving scientific review. It's worth noting that IACUCs can also address the problem of underrepresentation without engaging in scientific review by taking a similar approach.

The guiding principles of IACUCs—the 3Rs mentioned above—also place competing demands on members of ethical review committees. To gain a better understanding of these competing demands and current trends in emphasis, it is useful to examine the context of Russell and Burch's book, where the 3Rs were first introduced. Although the book's legacy is centered around minimizing inhumane treatment of animals during research, the authors were equally concerned with maximizing efficiency through the responsible use of animals. In their opening pages, they state that "the intimate relationship between humanity and efficiency in experimentation will recur constantly as a major theme in the present book."[27, p. 3-4 as cited in 33, p. 123] This relationship exists because a research project that adheres to the three principles is supposed to result in both high-quality research and humane treatment of animals.

[6] It is also a matter of beneficence, according to The Belmont Report: “Beneficence thus requires that we protect against risk of harm to subjects and also that we be concerned about the loss of the substantial benefits that might be gained from research.”[16]

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

But quality science and the ethical treatment of animals under the rubric of the 3Rs do not always converge in practice. Consider, for example, a scientist trying to fulfill the reduction principle, which says that being humane involves reducing the numbers of animals used in research. While keeping the number of experimental animals low seems appropriate, using animals too few in number to obtain scientifically rigorous findings can compromise the quality of the research and waste resources, which is unkind in the short term because it wastes individual animal lives and in the long-run by wasting larger numbers of animals. The wasteful use of animals then goes against the refinement principle, which says that animals should not be harmed unnecessarily.

That said, good science and the ethical treatment of animals do sometimes converge. Indeed, the humane treatment of animals can often improve the quality of research. For instance, minimizing pain and distress is typically undertaken to benefit animals, but it turns out that agitated animals also tend to disrupt experimental outcomes. Consider, for example, that rodents housed in temperatures that are too cold for them—which happens to be the room temperature maintained in most animal facilities—must use metabolic activity for heat generation, which in turn reduces the energy available for other functions, such as immune responses.[34][7] The realization that certain environmental conditions can distort experimental results inspired Russell to recommend that researchers not only minimize distress but also avoid physiological disturbances that might "upset the experimental results."[35, p. 277] The upshot is that treating animals decently can sometimes produce better scientific outcomes.

Similarly, minimizing disturbances to research animals often entails maintaining a consistent environment to reduce variability. This, in turn, enhances statistical power and allows

[7] Thanks to Gunnar Babcock for this example.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

researchers to detect effects with smaller sample sizes, thereby satisfying the reduction principle. Minimizing disturbances also aids in fulfilling the refinement principle by reducing unnecessary stress. Since undue stress can interfere with experimental results, minimizing disturbances leads to improved research quality. Thus, the elimination of variability through humane animal treatment adheres to the reduction and refinement principles and showcases how science and ethics can align to improve research outcomes under the 3Rs. In a later publication, Russell comes back to this example:

> Hitherto, it had always been supposed that to make animals uniform it was only necessary to keep them in the same environment. *Chance discovered that some environments are more favorable to uniformity than others*. The most uniform populations of all were those kept in *an environment optimal for their wellbeing*. In this respect, the goal of reduction is precisely the same as the *goal of refinement*.[36, p. 283, italics in original]

The idea expressed in the above passage, that happy animals are more uniform than unhappy ones, echoes the sentiment expressed in the opening line of Leo Tolstoy's novel Anna Karenina: "Happy families are all alike; every unhappy family is unhappy in its own way."[37, p. 1] While this may also be applicable to animal research, it's important to note that introducing variability into animal experiments, a stance I support in this paper, need not involve causing distress to animals. There are numerous ways to introduce variation without compromising animal welfare. Even so, the excerpt from Russell effectively illustrates how ethics and science can intersect. By

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

keeping research animals content, a researcher can reduce the inhumane treatment of animals while optimizing efficiency.

These points, however, are mostly germane only to exploratory research and not to preclinical research, where generalizability is crucial. Let me explain. At the preclinical stage, keeping everything uniform is helpful to satisfy the reduction principle, but that uniformity comes at the price of efficiency and refinement because preclinical experiments that don't generalize to human populations needlessly harm animals. If the outcome of an experiment does not translate to human patients in need of treatment, then the animals that were used in the experiment have by and large been wasted. Despite this, homogenized preclinical studies are routinely approved by IACUCs. This trend can be explained by the fact that the values expressed in Russell and Burch's 3Rs, like the values in the Belmont Report, can be weighed and prioritized in different ways. Homogenized preclinical studies can pass ethical review if the importance of reduction is overemphasized and the importance of refinement and efficiency are underemphasized.[see 38] The good news is that ending the pattern of approving potentially valueless experiments doesn't require engaging in scientific review. As with IRBs, ending the pattern is a matter of reprioritizing the guiding ethical principles already in use. Rather than prioritizing reduction, which has incentivized researchers to keep everything uniform, more emphasis should be placed on refinement and efficiency, which can be achieved by adding variability.

4. Implementation

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

The high failure rate of drugs entering human trials means most of the suffering of laboratory animals is futile, creating no commensurate benefit for human patients. Of course, most of the burden of improving translation falls on the scientific community, and to their credit, they have taken significant steps to improve translation—e.g., by creating guidelines for the planning of animal experiments and improving transparency.[39, 40] Be that as it may, broad scale changes to improve translation likely require restructuring the incentives of those who approve, regulate, fund, and publish scientific research, and ethical considerations provide one such incentive. Indeed, excessive homogeneity in preclinical and clinical trials is an ethical issue because, for one thing, it wastes animal lives with no offsetting benefit for human beings. In addition, homogenous clinical studies also fail to promote the equitable distribution of benefits, which is an ethical consideration since distribution of publicly funded benefits is an issue of justice. In either case, these ethical considerations have implications for experimental design since they arise from the push for homogeneity. By more clearly considering the ethical implications of experimental design at the preclinical and clinical stage, we might change it and improve translation. These points highlight the need for review committees to exercise their authority to ask for improvements in representation, which would help reduce current levels of homogeneity. Reprioritizing the values already contained within their guiding principles is necessary for review committees to exercise that authority.

There is currently momentum on the clinical side to prioritize diversity and inclusion by requiring IRBs to establish reasonable expectations regarding these values as a condition for study approval.[13] IACUCs should capitalize on this momentum and set similar goals and expectations for preclinical studies. Achieving better representation on the preclinical side may be more challenging in some respects, but easier in others. For instance, IRBs may have an easier

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

time setting clear goals for improving representation in clinical studies. If a treatment is supposed to be available for the general human population and a disease isn't more likely to affect some population of humans more than another, the ideal set of participants in a clinical trial should be representative of the population as a whole. Hence, if non-Hispanic white populations comprise some portion of the U.S. population, they should make up a similar portion of participants in U.S. trial sites. Currently, that's not the case since non-Hispanic white populations tend to be overrepresented.[32] Thus, reducing the number of non-Hispanic white participants in clinical trials is one specific goal that can help improve representation.

In contrast, setting clear goals for improving representation on the preclinical side is challenging. While increasing the number of female or older mice used may be reasonable, categories like race, ethnicity, and "lived experience of structural and interpersonal racism, lower socioeconomic status, and lower educational attainment"[18, p. 24] have no rodent equivalents, even though they are associated with specific health risks in humans (e.g., elevated blood pressure and cardiovascular risk [see 41, 42]). Aiming to build those representational frameworks into preclinical trials is thus very challenging, if not impossible. This suggests that the objective of improving representation in preclinical trials should not be, and is not, about matching mice to target human populations with their own idiosyncratic categories. Rather, improving representation and diversity in preclinical trials is about heterogeneity in the experimental subjects and their environments to better reflect biological reality. Since rodent lives differ significantly from human lives, achieving trial parity is unlikely. Nonetheless, if we can broaden the inference space with added variation, the findings from preclinical trials can inform our expectations of clinical trials.

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

Adding variation in preclinical research is an easy step to take since there are lots of opportunities to add variation as experiments have been progressively homogenized in the name of standardization over the years. Even so, researchers may not always know what factors impact the norm of reaction of a treatment effect, which might make it more challenging to improve representation in animal trials than in human trials. Nonetheless, researchers have options for introducing variation, including sex, age, environmental conditions, genetic strain, among others.[43] Furthermore, avenues of variation can largely be controlled at the preclinical stage by, for example, using several inbred strains. Indeed, to minimize the number of animals used, in accordance with the reduction principle, researchers can divide an experiment into several mini-experiments and add heterogeneity in small, controlled increments.[see 24, 26, 44]

There is still much to learn about heterogenization and its impact on treatment outcomes, so mandating a checklist to govern researchers in their experimental design is not the solution. Instead, researchers should be asked to discuss their choices regarding the variation they will introduce, or will not introduce, and how it applies to the intended inference space. As Korrina Duffy and colleagues explain, asking researchers to justify certain choices as part of ethical review does not increase the regulatory burden of review committees but forces researchers to consider the generalizability of their findings early in the research process.[45] Thus, for example, even if researchers are allowed to provide a compelling justification for the lack of representation in their sample, providing that justification forces them to confront the effects of a homogenized study sample on the generalizability of their findings (and hopefully nudges them to adjust their study design in a way that will improve translation). Similarly, when David Strauss and colleagues argue for improved representation in human trials, they are not interested in "overly prescriptive approaches…, specific mandates, or the application of quotas to study

This article has been accepted for publication in the *Journal of Medical Ethics* (July 2023) following peer review, and the Version of Record can be accessed online at doi: 10.1136/jme-2023-109166

samples."[13, p. 1211] Instead, they think much can be achieved if IRBs simply draw attention to diversity and inclusion as a goal and set reasonable expectations as conditions for study approval.

I mentioned earlier that some steps needed to improve representation on the preclinical side will be easier than on the clinical side. Setting specific goals and reasonable expectations might be harder on the preclinical side, simply because we still don't know which factors to vary and to what degree in order to improve translation. Despite this fact, whatever goals are set, achieving them will be easier in preclinical research because researchers have complete control over non-voluntary animal research. They can decide on the strain, sex, age, environment, etc. to use in their research without relying on volunteers.[8] In contrast, clinical researchers depend on volunteers, and some populations are understandably hesitant to participate due to historical abuses in clinical research. Unfortunately, these same populations are often underrepresented in clinical research. While researchers can take steps to encourage participation and retention, they cannot force or coerce people to participate or remain in a research study. Therefore, in non-voluntary animal research, improvements to diversity and inclusion should be easier to implement.

But regardless of the difficulty involved, we must not lose sight of the bigger picture. Without improvements in diversity and inclusion, we risk the continued waste of animal lives and unnecessary animal distress for the sake of flawed research. One way to move in the right direction is for IACUCs to exercise their authority to evaluate representation in preclinical trials.

References

[8] Although researchers conducting preclinical trials are still constrained by their ability to breed or obtain a relevantly diverse colony of research subjects. Thanks to an anonymous reviewer for bringing this to my attention.

1. Akhtar A. The flaws and human harms of animal experimentation. Cambridge quarterly of healthcare ethics: CQ: the international journal of healthcare ethics committees. 2015;24(4):407-19. https://doi.org/10.1017/S0963180115000079
2. Perel P, Roberts I, Sena E, et al. Comparison of treatment effects between animal experiments and clinical trials: Systematic review. Proceedings of the National Academy of Sciences. 2015;112(2):536-541. doi: 10.1073/pnas.1416164112
3. Angell EL, Bryman A, Ashcroft RE, et al. An analysis of decision letters by research ethics committees: the ethics/scientific quality boundary examined. Quality & safety in health care. 2008;17(2):131-6. https://doi.org/10.1136/qshc.2007.022756
4. Friesen P, Gelinas L, Kirby A et al (2022). IRBs and the Protection-Inclusion Dilemma: Finding a Balance The American journal of bioethics : AJOB [Internet]. [cited];1–14.
5. Van Norman GA (2019). Limitations of animal studies for predicting toxicity in clinical trials: is it time to rethink our current approach? [Internet]. JACC: Basic to Translational Science;4(7):845-854.
6. Garner JP (2014). The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it? ILAR journal;55(3):438–456.
7. Landi M, Everitt J, Berridge B. Bioethical, Reproducibility, and Translational Challenges of Animal Models. ILAR journal. 2021;62(1-2):60–65. https://doi.org/10.1093/ilar/ilaa027
8. Eggel M, Würbel H. Internal consistency and compatibility of the 3Rs and 3Vs principles for project evaluation of animal research. Laboratory animals. 2021;55(3):233–243.
9. Strech D, Dirnagl U (2019). 3Rs missing: animal research without scientific value is unethical [Internet]. BMJ open science;3(1):bmjos-2018-000048 [cited 2023 May 30]. Available from: https://doi.org/10.1136/bmjos-2018-000048
10. Curzer HJ, Perry G, Wallace MC, Perry D. The Three Rs of Animal Research: What they Mean for the Institutional Animal Care and Use Committee and Why. Science and engineering ethics. 2016;22(2):549–565.
11. Eggel M, Neuhaus CP, Grimm H. Reevaluating Benefits in the Moral Justification of Animal Research: A Comment on “Necessary Conditions for Morally Responsible Animal Research”. Cambridge quarterly of healthcare ethics : CQ : the international journal of healthcare ethics committees. 2020;29(1):131–143.
12. Kimmelman J Federico C (2017). Consider drug efficacy before first-in-human trials Nature;542(7639):25–27.
13. Strauss DH, White SA, Bierer BE. Justice, diversity, and research ethics review. Science (New York, N.Y.). 2021;371(6535):1209-1211. doi:10.1126/science.abf2170
14. Spanagel R (2022). Ten Points to Improve Reproducibility and Translation of Animal Research [Internet]. Frontiers in behavioral neuroscience;16:869511 [cited 2023 May 30]. Available from: https://doi.org/10.3389/fnbeh.2022.869511
15. Heller J (1972). Syphilis patients died untreated. Washington Star.
16. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research [Internet]. 1979 [cited 2023 May 30]. Available from: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

17. Drozda K, Wong S, Patel SR, et al. Poor warfarin dose prediction with pharmacogenetic algorithms that exclude genotypes important for African Americans. Pharmacogenetics and genomics. 2015;25(2):73–81.
18. Emanuel EJ, Wendler D, Grady C. The ethics of research involving women and underrepresented groups In: Emanuel EJ, Wendler D, Grady C (eds). Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups Oxford University Press; 2020:23-44.
19. Liu KA, Mager NA. Women's involvement in clinical trials: historical perspective and future implications. Pharmacy practice. 2016;14(1):708. https://doi.org/10.18549/PharmPract.2016.01.708
20. Bluhm R, Borgerson K. An Epistemic Argument for Research-Practice Integration in Medicine. The Journal of medicine and philosophy. 2018;43(4):469–484.
21. Engber D (2011). The trouble with Black-6 [Internet]. Slate [cited 16 December 2021]. Available from: http://www.slate.com/articles/health_and_science/the_mouse_trap/2011/11/black_6_lab_mice_and_the_history_of_biomedical_research.html
22. Clayton JA. Studying both sexes: a guiding principle for biomedicine. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2016;30(2):519–524.
23. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science (New York, N.Y.). 1999;284(5420):1670–1672.
24. Voelkl B, Altman NS, Forsman A, et al. Reproducibility of animal research in light of biological variation. Nature reviews. Neuroscience. 2020;21(7):384–393.
25. Lewontin RC. The analysis of variance and the analysis of causes. American Journal of Human Genetics. 1984;36(5):1288-1292.
26. von Kortzfleisch VT, Ambrée O, Karp NA, et al. Do multiple experimenters improve the reproducibility of animal studies?. PLoS biology. 2022;20(5):e3001564. https://doi.org/10.1371/journal.pbio.3001564
27. Russell WMS, Burch RL (as reprinted 1992). The principles of humane experimental technique [Internet]. Wheathampstead (UK): Universities Federation for Animal Welfare; [cited 2023 May 30].
28. Animal Welfare Act, 7 U.S.C. §§ 2131-2159 (1985).
29. Garner JP (2021). The mouse in the room: The critical distinction between regulations and ethics In: Beauchamp TL, DeGrazia D (eds). The Oxford Handbook of Animal Ethics Oxford University Press;2021:448-458.
30. Waltz M, Fisher JA, Walker RL (2023). Mission Creep or Mission Lapse? Scientific Review in Research Oversight [Internet]. AJOB empirical bioethics;14(1):38–49 [cited 2023 May 30]. Available from: https://doi.org/10.1080/23294515.2022.2123868
31. Varma T, Jones CP, Oladele C, et al. Diversity in clinical research: public health and social justice imperatives. Journal of Medical Ethics. 2023;49:200-203.
32. U.S. Food and Drug Administration (2020). Summary of clinical trials of drugs conducted from 2015 to 2019 (Publication No.FDA-2019-N-5123-0002) [Internet]. U.S. Department of Health and Human Services.
33. Tannenbaum J,Bennett BT (2015). Russell and Burch's 3Rs then and now: the need for clarity in definition and purpose [Internet]. Journal of the American Association for Laboratory Animal Science: JAALAS;54(2):120–132.

34. Vialard F, Olivier M. Thermoneutrality and Immunity: How Does Cold Stress Affect Disease?. Frontiers in immunology. 2020;11:588387. https://doi.org/10.3389/fimmu.2020.588387
35. Russell WM. The Progress of Humane Experimental Technique: The First Annual FRAME Lecture, presented at the Royal Society of Medicine, London, on 24 September 1999. Alternatives to Laboratory Animals. 2009;37(3):277-283.
36. Russell WMS. The Three Rs: past, present and future. Animal Welfare. 2005;14(4):279-286.
37. Tolstoy L, Pevear R, Volokhonsky L (2002). Anna Karenina: a novel in eight parts [Internet]. New York(NY), Penguin Books.
38. Everitt JI, Berridge BR (2017). The role of the IACUC in the design and conduct of animal experiments that contribute to translational success ILAR journal;58(1):129-134.
39. National Centre for the Replacement, Refinement, and Reduction of Animals in Research. ARRIVE guidelines [Internet]. 2010 [cited 2023 May 30]. Available from: https://www.nc3rs.org.uk/arrive-guidelines
40. Animals in Science Regulation Unit. PREPARE guidelines for planning animal research and testing. National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs). 2014.
41. Hamad R et al (2020). Association of low socioeconomic status with premature coronary heart disease in US adults JAMA Cardiology;5(8):899–908.
42. King JB et al (2021). Multiple social vulnerabilities to health disparities and hypertension and death in the regards study Hypertension;79(1):196–206.
43. Josef Van Der Staay F et al (2010). The standardization–generalization dilemma: a way out Genes Brain and Behavior;9(8):849-855.
44. Suman PR, Lino de Oliveira C (2022) Systematic heterogenisation to improve reproducibility in animal studies [Internet]. PLoS Biol;20(5): e3001629 [cited 2023 May 30]. Available from: https://doi.org/10.1371/journal.pbio.3001629
45. Duffy KA, Ziolek TA, Epperson CN. Filling the Regulatory Gap: Potential Role of Institutional Review Boards in Promoting Consideration of Sex as a Biological Variable. Journal of women's health (2002). 2020;29(6):868–875.

Acknowledgments: I am indebted to Ilona Piotrowska and Ted Piotrowski for making the memorable hiking trips in Patagonia possible, where the initial idea for this paper took shape. In addition, many thanks to Damian Zuloaga for his scientific expertise and Matt Mosdell for his outstanding editorial skills.