



December 2024

NATIONAL INSTITUTES OF HEALTH

Assessing Efforts to Improve Animal Research Could Lead to Greater Human Health Benefits

Reissued on January 29, 2026, to reflect corrected information from the National Institutes of Health about the agency's spending on animal research on the Highlights page and pages 2 and 34.

Assessing Efforts to Improve Animal Research Could Lead to Greater Human Health Benefits

Why GAO Did This Study

NIH, within the Department of Health and Human Services, spent billions of dollars in fiscal year 2024 on research, including on projects that involve animals. NIH supports research conducted by its own institutes and by external entities such as universities. In recent years, researchers have reported low success rates in reproducing and translating the results of animal experiments. Reproducibility of a study can reflect how reliable its results are, and translatability enables animal research to benefit human health.

The Consolidated Appropriations Act, 2023 includes a provision for GAO to report on aspects of animal welfare, reproducibility, and translatability in research NIH conducts or supports. This report addresses (1) steps NIH takes to ensure animal welfare in research it conducts and supports, (2) challenges that limit the reproducibility and translatability of this research, and (3) steps NIH has taken to enhance reproducibility and translatability in animal research.

GAO analyzed NIH documents and data; reviewed relevant scientific publications; conducted two site visits to labs that conduct animal research; and interviewed NIH officials, scientific researchers, and representatives of nongovernmental organizations.

What GAO Recommends

GAO is making two recommendations to NIH to define short-term goals and collect evidence to assess its efforts to enhance the reproducibility and translatability of animal research. The agency concurred with both recommendations.

View GAO-25-107140. For more information, contact Steve Morris at (202) 512-3841 or morriss@gao.gov or Candice N. Wright at (202) 512-6888 or wrightc@gao.gov.

What GAO Found

Research involving animals has contributed to a better understanding of human health and new treatments for diseases; for example, it has helped scientists to develop a COVID vaccine. However, concerns have been raised about the welfare of laboratory animals. The National Institutes of Health (NIH) has a policy that addresses the welfare of animals in research conducted by its institutes and external institutions. NIH takes various steps to oversee this research. These include requiring institutions to report and address noncompliance with the policy at the institutional and project levels and monitoring compliance with the policy.

Examples of Types of Animals Used in Animal Research



Source: HYUNGKEUN/stock.adobe.com (left photo); New Africa/stock.adobe.com (right photo). | GAO-25-107140

Concerns have also been raised about whether the results of animal research can consistently be reproduced in subsequent animal studies (reproducibility) or translated into similar results for humans (translatability). Multiple challenges contribute to reported low success rates in reproducibility and translatability, according to NIH, scientific publications, and scientific researchers GAO interviewed. For example, researchers sometimes do not publish enough details about their study design and use of animals. As a result, other researchers attempting to reproduce the research may obtain different or inconsistent results. Also, animals' responses to experimental treatments and drugs do not necessarily resemble those of humans because of inherent biological differences.

NIH has taken steps to enhance the reproducibility and translatability of animal research it conducts and supports. For example, NIH issued a policy in 2015 on enhancing reproducibility through rigor and transparency, issued guidelines, and provided resources for researchers. However, NIH has not fully implemented key practices that can help agencies assess whether their efforts have led to measurable improvements. For example, NIH has not defined short-term goals to help track progress toward improving reproducibility and translatability or collected evidence that would help the agency assess the effectiveness of its efforts. Such evidence could include agency-wide information on whether grant applicants are following its 2015 policy. Agency officials said they had not done so because variability among different fields of study would require setting field-specific goals and measures. However, GAO has reported that, in cases like this, agencies can set specific targets and time frames for different areas and assess the contributions of each area to an agency's long-term goals.

Contents

| | | |
|--|---|----|
| Letter | | 1 |
| Background | | 5 |
| NIH Takes Various Steps to Help Ensure Animal Welfare in Research It Conducts or Supports | | 7 |
| Multiple Challenges Limit the Reproducibility and Translatability of Animal Research | | 15 |
| NIH Has Taken Steps Intended to Enhance Reproducibility and Translatability but Has Not Assessed Its Progress | | 25 |
| Conclusions | | 34 |
| Recommendations for Executive Action | | 35 |
| Agency Comments and Our Evaluation | | 35 |
| <hr/> Appendix I | Categories of Noncompliance with the Public Health Service Policy | 37 |
| <hr/> Appendix II | Recommendations from NIH Advisory Committee to the Director to Improve Animal Research | 39 |
| <hr/> Appendix III | Comments from the Department Health and Human Services | 43 |
| <hr/> Appendix IV | GAO Contacts and Staff Acknowledgments | 45 |
| <hr/> Tables | | |
| Table 1: Data on Annual Reports the National Institutes of Health (NIH) Received from Institutions with Approved Animal Welfare Assurances, 2018–2023 | | 8 |
| Table 2: Categories of Noncompliance with the Public Health Service (PHS) Policy Most Frequently Reported to the National Institutes of Health (NIH) for Cases Closed, 2021–2023 | | 14 |
| Table 3: National Institutes of Health's (NIH) Four Areas of Rigor for Grant Applications | | 26 |
| Table 4: Categories of Noncompliance with the Public Health Service Policy Reported to the National Institutes of Health (NIH) for Cases Closed, 2021–2023 | | 37 |

| | |
|--|-----------|
| Table 5: Recommendations from the National Institutes of Health (NIH) Advisory Committee to the Director (ACD) Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research, 2021 | 40 |
|--|-----------|

Figures

| | |
|---|-----------|
| Figure 1: Sources of Reports of Noncompliance with Public Health Service (PHS) Policy for Research Involving Animals at Intramural and Extramural Institutions, 2023 | 12 |
| Figure 2: Actions NIH Can Take in Response to Noncompliance with the Public Health Service (PHS) Policy for Research Involving Animals | 15 |
| Figure 3: Challenges That Limit the Reproducibility and Translatability of Animal Research | 18 |
| Figure 4: Different Arrangements for Placement of Animal Cages | 21 |
| Figure 5: Examples of Controllable and Less Controllable Extrinsic Factors That Can Influence Outcomes in Animal Research | 24 |
| Figure 6: Examples of Challenge Areas, Recommendations, and Steps NIH Has Taken Intended to Enhance Reproducibility and Translatability of Animal Research | 29 |

Abbreviations

| | |
|------|---|
| ACD | Advisory Committee to the Director |
| HHS | Department of Health and Human Services |
| NIH | National Institutes of Health |
| OLAW | Office of Laboratory Animal Welfare |
| PHS | Public Health Service |

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



U.S. GOVERNMENT ACCOUNTABILITY OFFICE

441 G St. N.W.
Washington, DC 20548

December 19, 2024

The Honorable Bernard Sanders
Chair
The Honorable Bill Cassidy, M.D.
Ranking Member
Committee on Health, Education, Labor, and Pensions
United States Senate

The Honorable Cathy McMorris Rodgers
Chair
The Honorable Frank Pallone, Jr.
Ranking Member
Committee on Energy and Commerce
House of Representatives

Scientific research involving animals—referred to as animal research—has contributed to a better understanding of human health and new treatments for diseases, such as medications and vaccines.¹ For example, such research helped scientists to develop a COVID vaccine. However, concerns have been raised about the welfare of laboratory animals, including the suitability of the animals' living conditions and pain they may experience. Concerns have also been raised following a number of recent studies that reported unsuccessful attempts to reproduce the results of animal experiments by using similar methods (reproducibility) and to translate the results of these experiments into similar experiments and results for humans (translatability).² Reproducibility helps ensure that results of animal research are reliable,

¹For the purposes of our report, we use the definition of "animal" from the Public Health Service Policy on Humane Care and Use of Laboratory Animals, which includes any live, vertebrate animal used or intended for use in research, research training, experimentation, or biological testing or for related purposes.

²See, for example, Emma Wilson et al., "Designing, conducting, and reporting reproducible animal experiments," *Journal of Endocrinology*, vol. 258, no. 1 (2023); Duxin Sun et al., "Why 90% of clinical drug development fails and how to improve it?," *Acta Pharmaceutica Sinica B*, vol. 12, no. 7 (2022): 3,049–3,062; Benjamin Ineichen et al., "Analysis of animal-to-human translation shows that only 5% of animal-tested therapeutic interventions obtain regulatory approval for human applications," *PLOS Biology*, vol. 22, no. 6 (2024): e3002667.

and translatability enables animal research to benefit human health.³ Results that cannot be reproduced or translated to humans can slow medical progress, waste resources, and decrease public trust in scientific research.

The National Institutes of Health (NIH), within the Department of Health and Human Services (HHS), conducts projects at its own research institutes (intramural research) and supports projects conducted by external institutions such as universities, medical schools, and companies (extramural research). NIH spent about \$47.4 billion in fiscal year 2024 to support research projects generally, including about \$5.2 billion on intramural research and \$37.5 billion on extramural research. In recent years, about 41 percent of intramural research projects have involved animals and about 25 percent of extramural research awards have involved animals.⁴ Animal research conducted or supported by NIH addresses a wide range of scientific questions, from basic research aimed at understanding biological and physiological processes to testing new drugs or treatments that, if successful, could lead to human clinical trials. Examples of animals used in research include mice, rats, rabbits, zebrafish, and guinea pigs.

NIH's Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS Policy) addresses the welfare of animals in research conducted and supported by NIH and other components of HHS's Public Health Service.⁵ NIH research institutes and domestic external institutions receiving support from NIH are responsible for

³In our prior reports, we have differentiated between reproducibility (achieving similar results using similar or same methods and data used in a prior experiment) and replicability (achieving similar results using similar methodology but different data). These terms are sometimes used interchangeably. For the purposes of this report, we use the term reproducibility to refer to the concept of successfully reproducing the results of an experiment using methods similar to those used in that experiment, regardless of the data.

⁴According to NIH officials, it is not possible to accurately capture the annual amount of NIH funding that is allocated to research involving animals because animal activities associated with intramural and extramural research are integral to the research question being investigated and cannot be defined as a standalone project and budget item. For intramural research projects, contracts are not counted separately from projects; the count of extramural research awards includes contracts, according to NIH officials.

⁵The document known as the PHS Policy was published in 1986, revised in 2002, and revised to its current edition in 2015. National Institutes of Health, Office of Laboratory Animal Welfare, *Public Health Service Policy on Humane Care and Use of Laboratory Animals* (Bethesda, Md.: 2015).

complying with the PHS Policy.⁶ NIH offices in headquarters oversee its institutes' and funding recipients' compliance with the policy. We have previously reported on related issues, including NIH's oversight of animal research conducted in foreign facilities and the reliability of federally funded research.⁷

The Consolidated Appropriations Act, 2023 includes a provision for us to report on the extent to which animal research NIH conducts or supports meets federal requirements for animal welfare, as well as NIH's processes for ensuring that this research may be reasonably anticipated to be reproducible and translatable.⁸ This report (1) describes the steps NIH takes to help ensure animal welfare in research that it conducts and supports, (2) describes challenges that limit the reproducibility and translatability of animal research, and (3) evaluates steps NIH has taken to enhance reproducibility and translatability in animal research and the effectiveness of these steps.

To describe the steps NIH takes to help ensure animal welfare in research it conducts and supports, we reviewed NIH documents and data and interviewed agency officials and other stakeholders. Specifically, we reviewed the PHS Policy and other NIH policies and guidance for animal research conducted by NIH and other institutions. We also reviewed NIH data on funding for animal research, and oversight of animal research—including data on annual reports and noncompliance reports from funding recipients—from 2018 through 2023. We assessed the reliability of these data, which included screening for omissions and anomalies, obtaining written responses from agency officials to questions about the data's reliability, and reviewing technical documentation. We determined that the

⁶Foreign facilities performing research for domestic award recipients also generally must comply with the PHS Policy unless otherwise specified, and foreign facilities performing research for foreign award recipients must either comply with the PHS Policy or provide evidence that acceptable standards for the humane care and use of animals will be met. Research conducted at foreign facilities is outside the scope of this report. For information about NIH's oversight of animal research conducted at foreign facilities, see GAO, *Animal Use in Research: NIH Should Strengthen Oversight of Projects It Funds at Foreign Facilities*, GAO-23-105736 (Washington, D.C.: Mar. 30, 2023).

⁷See GAO-23-105736 and *Research Reliability: Federal Actions Needed to Promote Stronger Research Practices*, GAO-22-104411 (Washington, D.C.: July 28, 2022). In our 2022 report, we recommended that NIH collect information on indicators of rigor to assess the research projects it funds, and implement steps, as needed, to promote strong research practices in future work. NIH concurred with our recommendation but had not implemented it as of August 2024.

⁸Pub. L. No. 117-328, § 2331(b), 136 Stat. 4459, 5781 (2022).

data were sufficiently reliable for providing descriptive information about the steps NIH takes to help ensure animal welfare in the research it conducts and supports.

We also interviewed NIH officials, scientific researchers, and representatives from institutions receiving NIH support and nongovernmental organizations. For these interviews, we selected entities and individuals knowledgeable about conducting or overseeing animal research, using the results from such research, or understanding the welfare of animals used in research. In addition, we visited two laboratories—an NIH lab and a university lab—that conduct animal research to observe animal care practices related to the PHS Policy. Because we focused our review on intramural and extramural animal research conducted by domestic institutions, we selected one site that conducted intramural research and one that conducted extramural research. We visited these sites so that we could directly observe and interview knowledgeable individuals about practices used in laboratories that conduct significant amounts of research with different types of animals. The site visits were not intended to generate findings representative of all entities or individuals conducting or overseeing animal research.

To identify challenges that limit the reproducibility and translatability of animal research, we reviewed a range of documents and interviewed scientific researchers. Specifically, we reviewed 38 relevant scientific publications from 2012 through January 2024, two National Academies reports, an NIH working group report, NIH written responses, and our relevant work.⁹ We also interviewed 10 scientific researchers about their views on challenges and potential root causes, and we discussed challenges during our visits to the NIH and university laboratories. We identified scientific researchers to interview through our review of the literature, referrals from interviewees, and our prior work on related topics. We selected scientific researchers who represent academia, industry, and nongovernmental organizations and are knowledgeable about challenges that limit reproducibility and translatability. We reviewed the scientific publications and our interview notes and identified eight interrelated challenges, which we consolidated into the three challenge areas described in this report. Although the challenges and challenge

⁹For the 38 scientific publications, we initially searched for papers related to challenges of reproducibility and translatability without setting a range of dates. The most relevant papers we found were published during these years.

areas we identified are not necessarily mutually exclusive or exhaustive, they were confirmed across multiple sources.

To evaluate the steps NIH has taken to enhance reproducibility and translatability in animal research and the effectiveness of these steps, we reviewed NIH's Strategic Plan and gathered information from NIH about the steps it has taken from 2015 through July 2024, such as implementing policies and providing funding opportunities. Specifically, we reviewed NIH policies, notices, reports, and other documents, as well as written responses from NIH officials. We compared NIH's steps to relevant aspects of GAO's *Key Practices for Evidence-Based Policymaking*, which includes steps that federal agencies can take to assess the results of federal efforts.¹⁰

We conducted this performance audit from October 2023 to December 2024 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Background

Requirements Governing Animal Research Conducted and Supported by NIH

The Health Research Extension Act of 1985 required the Director of NIH to establish guidelines for the proper care and treatment of animals used in biomedical and behavioral research.¹¹ In response, NIH established the PHS Policy.¹² Before institutions can receive PHS support from NIH, they must describe how they will comply with the PHS Policy. The policy requires institutions to establish and maintain proper measures to ensure the appropriate care and use of animals involved in research, research training, and biological testing activities—collectively referred to as “activities”—conducted or supported by the Public Health Service,

¹⁰GAO, *Evidence-Based Policymaking: Practices to Help Manage and Assess the Results of Federal Efforts*, GAO-23-105460 (Washington, D.C.: July 12, 2023).

¹¹42 U.S.C. § 289d(a).

¹²NIH formalized, revised, and expanded animal welfare policies beginning in the 1950s. In 1986, to implement the Health Research Extension Act of 1985, NIH issued a new edition of its policy bearing the current title.

including NIH and other HHS agencies.¹³ The PHS Policy also requires institutions to use the *Guide for the Care and Use of Laboratory Animals* as a basis for developing and implementing an institutional program for activities involving animals.¹⁴

| | |
|--|---|
| NIH Entities' Roles and Responsibilities Related to Animal Research | <p>Several NIH entities have roles and responsibilities related to animal research.</p> <ul style="list-style-type: none">• NIH has 27 institutes and centers, most of which conduct and support animal research. The institutes and centers often focus on particular diseases or body systems. For example, the National Cancer Institute and the National Heart, Lung, and Blood Institute are named for their respective focus areas.• The Office of Intramural Research includes the Office of Animal Care and Use, which helps ensure NIH research programs and facilities for animal care and use are in compliance with the PHS Policy and other regulatory requirements.• The Office of Laboratory Animal Welfare (OLAW) provides guidance and interpretation of the PHS Policy and monitors compliance with the policy to help ensure the humane care and use of animals in intramural and extramural research. |
| NIH Review and Funding Processes for Proposed and Ongoing Animal Research | <p>NIH has separate review and funding processes for ongoing intramural research and proposed extramural research:</p> <ul style="list-style-type: none">• Intramural research. Boards of Scientific Counselors, composed of non-NIH scientists, evaluate the performance of NIH researchers and the quality of their research programs. These evaluations inform NIH decisions about tenure and funding.• Extramural research. Scientists' organizations submit applications to NIH for grant funding. Groups of primarily non-federal scientists with expertise in the field—known as peer reviewers—score the grant applications for their scientific and technical merit, and NIH advisory |

¹³The PHS Policy also applies to activities supported or conducted by entities that have a memorandum of understanding with NIH, including the National Science Foundation, National Aeronautics and Space Administration, and Department of Veterans Affairs.

¹⁴Institute for Laboratory Animal Research, National Research Council of the National Academies, *Guide for the Care and Use of Laboratory Animals: Eighth Edition* (Washington, D.C.: National Academies Press, 2011).

councils review the applications for mission relevance. NIH directors of institutes and centers make final funding decisions.

NIH Takes Various Steps to Help Ensure Animal Welfare in Research It Conducts or Supports

NIH's Office of Laboratory Animal Welfare Monitors Compliance with PHS Policy Requirements at the Institutional Level

NIH takes various steps to help ensure animal welfare at the institutional and research project levels for animal research it conducts or supports. NIH also requires institutions to report and address noncompliance with the PHS Policy at both the institutional and project levels.

OLAW interprets and monitors compliance with PHS Policy requirements for extramural and intramural research at the institutional level. OLAW oversees institutions' compliance with the requirements under the PHS Policy, which include the following:

- **Animal care and use committee.** The PHS Policy calls for intramural and extramural institutions to have committees that oversee animal care and use programs, as well as inspect animal facilities. Specifically, each animal care and use committee is required to, at least semiannually, review the institution's program for the humane care and use of animals; inspect the animal facilities; and prepare reports of the committee's evaluations, which the institution is to maintain and make available to OLAW upon request. Each committee is also required to review concerns involving the care and use of animals at the institution.¹⁵
- **Animal welfare assurance.** Before they can receive PHS funding from NIH, intramural and extramural institutions are required to have an animal welfare assurance on file with OLAW. Each assurance is to describe the institution's program for animal care and use. Such description must include the membership list of the animal care and use committee, the procedures the committee will follow to fulfill the PHS Policy's requirements, and a synopsis of relevant training or instruction offered to scientists, animal technicians, and other personnel. OLAW reviews all assurance documents, and the

¹⁵Committee members are appointed by the chief executive officer of their institution, and committees may, after following certain procedural requirements, suspend projects if they determine that the projects are not being conducted in accordance with applicable requirements. Each committee must have at least five members, including at least one veterinarian, one practicing scientist experienced in animal research, one member whose primary concerns are in a nonscientific area, and one member who is not affiliated with the institution other than as a member of the committee.

assurance is generally approved for up to 4 years. In 2023, approximately 1,200 domestic institutions had approved assurances.

- **Annual report.** The PHS Policy requires annual reports from each institution's animal care and use committee.¹⁶ In the annual report, the committee must report any changes in the description of the institution's program of animal care and use or facilities in its assurance, or its committee membership. The committee must also report, among other information, the dates that it conducted its semiannual evaluations of the institution's program and facilities.¹⁷ According to NIH officials, OLAW reviews all annual reports for completeness and forwards a portion of the facilities' reports to an OLAW assurance officer for additional review if an institution is out of compliance (for example, for not completing required animal facility inspections) or for other reasons such as program changes. In some cases, OLAW provides guidance to institutions. From 2018 through 2023, OLAW provided guidance or explanations—for example reminding them about PHS Policy requirements—to about 10 percent of institutions that submitted annual reports (see table 1).

Table 1: Data on Annual Reports the National Institutes of Health (NIH) Received from Institutions with Approved Animal Welfare Assurances, 2018–2023

| Reporting period | Total number of annual reports from domestic institutions | Accepted, no issues | Additional review, guidance provided |
|------------------|---|-------------------------------|--------------------------------------|
| 2018 | 890 | 828 (93%) | 62 (7%) |
| 2019 | 882 | 802 (91%) | 80 (9%) |
| 2020 | 872 | 768 (88%) | 104 (12%) |
| 2021 | 885 | 779 (88%) | 106 (12%) |
| 2022 | 876 | 784 (89%) | 92 (11%) |
| 2023 | 871 | 632 of 696 (91%) ^a | 64 of 696 (9%) ^a |

Source: GAO analysis of NIH data. | GAO-25-107140

Note: In 2020, NIH transitioned from reporting calendar years to reporting fiscal years to harmonize reporting periods with USDA as recommended by a working group convened in response to the 21st

¹⁶According to NIH officials, some assurances cover multiple entities within institutions, multiple institutions, or both. The officials told us a single annual report is submitted per assurance, so the number of annual reports is less than the number of institutions with approved assurances.

¹⁷The PHS Policy also requires the institution's animal care and use committee to report any changes in the institution's program or facilities that would place the institution in a different status category than specified in its assurance (i.e., category 1, accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International, or category 2, evaluated by the institution's committee).

Century Cures Act. Data for 2018 and 2019 are for the calendar year; data for 2020 are for January 1, 2020, through September 30, 2020; and data for 2021 through 2023 are for the fiscal year.

^aAs of September 2024, NIH had reviewed 696 of the 871 annual reports it received, and the remaining 175 were pending completion, according to NIH officials. The officials told us they plan to complete the reviews by December 2024.

Under the PHS Policy, OLAW is also responsible for conducting site visits to selected intramural and extramural institutions that have animal welfare assurances. Of the 1,200 such institutions, OLAW generally visited between eight and 12 institutions' animal housing and procedure facilities each year from 2018 through 2023.¹⁸ According to NIH officials, the purpose of site visits is to evaluate compliance with the PHS Policy, and OLAW conducts the site visits "for cause" if the established compliance procedures need additional in-person evaluation, or for other reasons such as requests from Congress or NIH leadership. OLAW also considers institution size, amount of funding, and region of the country in selecting sites to visit, according to NIH officials.

NIH Oversight at the Project Level Includes Guidance for Reviewing Research Proposals and Progress Reports

In addition to monitoring compliance with requirements and other practices that apply at the institutional level, NIH conducts oversight of intramural and extramural animal research at the individual project level, including by issuing guidance to animal care and use committees, agency officials, researchers, and peer reviewers.

For example, NIH has issued guidance for animal care and use committees' reviews of projects involving animals, which are required under the PHS Policy. Under the policy, these committees must review and approve animal protocols in proposed projects or significant changes to protocols in ongoing research to ensure they are in accordance with the policy. These requirements include avoiding or minimizing discomfort, distress, and pain to animals and generally using appropriate sedation or anesthesia for procedures that may cause more than momentary or slight pain.

NIH's guidance for committee review of projects involving animals includes the following:

- **Weighing benefits.** As the impact of the proposed procedures on the animal's well-being increases, the committee must decide if the study's benefits to medicine and science outweigh the costs to the

¹⁸OLAW completed three site visits in 2020 and none in 2021 due to the COVID-19 pandemic, according to NIH officials.

animal's well-being, according to the guidance.¹⁹ The committee is to conduct this analysis by using the project proposal's explanation of procedural alternatives that have been considered, number and justification of animals required, and other factors.

- **Conducting reviews.** For intramural and extramural research, animal care and use committees must conduct continuing reviews of previously approved, ongoing animal research activities at appropriate intervals determined by the committee, including a complete review at least once every 3 years.²⁰

NIH has also issued guidance for agency officials on reviewing the periodic reports that extramural researchers submit to NIH on their research involving animals.²¹ NIH requires these researchers to prepare progress reports each year for ongoing research projects and submit them for review. Progress reports are to document the project's accomplishments and status. NIH officials are required to review all progress reports that they receive, including to determine whether there are animal welfare issues or concerns.²²

For extramural research specifically, NIH has also issued policy and guidance related to grant requirements, the grant application process, and review of grant applications. The NIH Grants Policy Statement discusses, for example, grant application peer review, which is required

¹⁹National Institutes of Health, Animal Research Advisory Committee, "Guidelines for Review and Approval of Animal Study Proposals and Significant Changes" (Washington, D.C.: May 24, 2023), accessed October 30, 2024, https://oacu.oir.nih.gov/system/files/media/file/2023-05/C6_Review-ASP-Significant-Changes.pdf.

²⁰See National Institutes of Health, *NIH Policy Manual*, 3040-2 - Animal Care and Use in the Intramural Research Program (Washington, D.C.: Apr. 13, 2023) and National Institutes of Health, Office of Laboratory Animal Welfare, *Public Health Service Policy on Humane Care and Use of Laboratory Animals*.

²¹See, for example, National Institutes of Health, *NIH Policy Manual*, 54444 - Evaluation of Grant Progress Reports by Program Officials (Washington, D.C.: Oct. 1, 2001).

²²According to NIH officials, for some extramural progress reports, program officials conduct follow-up with institutions, which may involve institutions submitting corrected or additional materials, or resolve issues via email correspondence.

under federal law and NIH regulation.²³ For any research involving live vertebrate animals, OLAW provides a checklist to peer reviewers. The checklist guides reviewers to look for certain items in the vertebrate animals section in grant applications, including a description of procedures to be used that involve animals, justification that the species are appropriate for the proposed research, a description of the interventions to minimize pain and distress, and information regarding the method of euthanasia.

After peer review, OLAW reviews approximately 5 percent of vertebrate animals sections of funded proposals, according to NIH officials. OLAW reviews these sections for various reasons, such as peer reviewers raising animal welfare concerns or the need for a new animal welfare assurance to be issued to an institution that has applied to perform animal research, according to OLAW officials. For example, in one case, peer reviewers raised concerns because the animals section did not describe how researchers would determine whether anesthesia they administered to mice was working or whether the mice were in distress during prolonged restraint. According to OLAW officials, in this case, they asked the research institution to address the peer reviewer comments, revise the animals section, and provide additional information. OLAW officials reviewed the revised animals section and accepted it.

OLAW Requires Institutions to Self-Report and Address Noncompliance

For both intramural and extramural research, the PHS Policy requires animal care and use committees to report any serious or continuing noncompliance with the PHS Policy, any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or any suspension of an activity by the committees.²⁴ Examples of reportable situations include conditions that jeopardize the health or well-being of animals, the failure to adhere to protocols approved by the institution's animal care and use

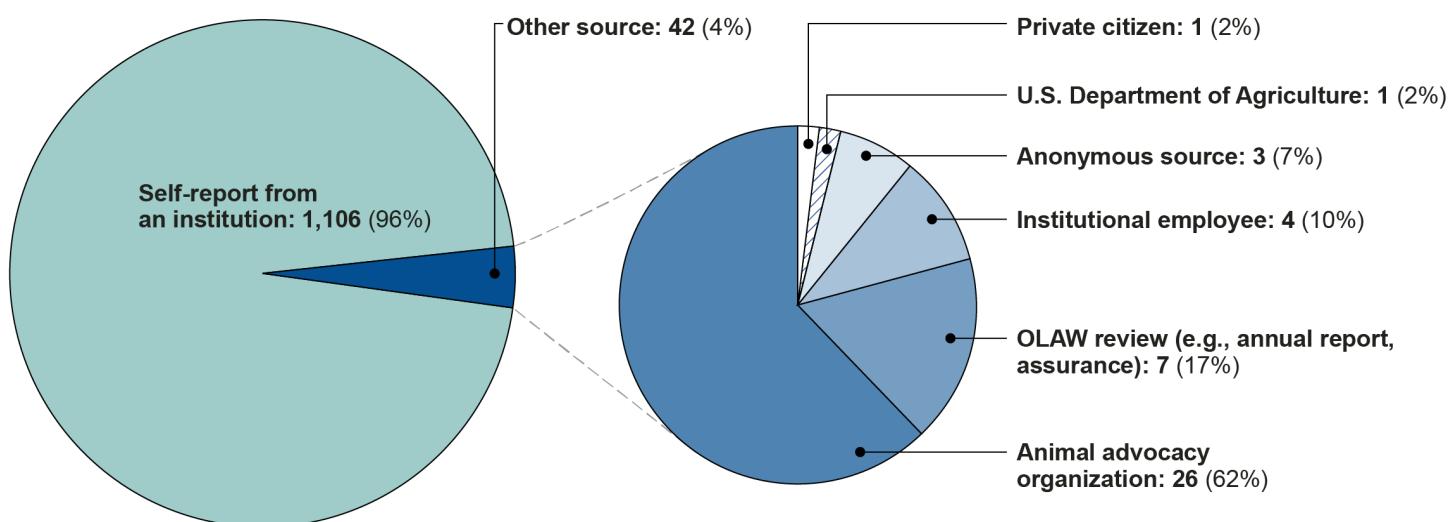
²³National Institutes of Health, *NIH Grants Policy Statement* (Washington, D.C.: 2024). NIH regulations implementing the Public Health Service Act, as amended, require peer review of applications for grants and cooperative agreements for biomedical and behavioral research before grants may be awarded. 42 U.S.C. § 289a(a); 42 C.F.R. §§ 52h.1(a)(1); 52h.7(a). In addition to applications for grants and cooperative agreements, peer review is also required for contract proposals. 42 U.S.C. § 289a(a); 42 C.F.R. §§ 52h.1(a)(2), 52h.9(a), 52h.10(a).

²⁴Specifically, the PHS Policy states that animal care and use committees are to promptly provide OLAW with a full explanation of the circumstances and actions taken with respect to (1) any serious or continuing noncompliance with the PHS Policy, (2) any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or (3) any suspension of an activity by the committee. In this report, we use the term "noncompliance" to refer to all three types of reportable situations.

committee, and equipment failure. OLAW also requires institutions to describe the corrective and preventative actions they took to address the situation. Most noncompliance reports are self-reports from the research institutions themselves, and NIH officials told us that OLAW reviews all noncompliance reports it receives (see fig. 1).

Figure 1: Sources of Reports of Noncompliance with Public Health Service (PHS) Policy for Research Involving Animals at Intramural and Extramural Institutions, 2023

Number of reports: 1,148



OLAW = National Institutes of Health's Office of Laboratory Animal Welfare

Source: GAO analysis of National Institutes of Health information. | GAO-25-107140

Note: For the purposes of this report, noncompliance includes (1) any serious or continuing noncompliance with the PHS Policy, (2) any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or (3) any suspension of an activity by the animal care and use committee.

In recent years, NIH has tracked the categories of noncompliance identified in each report. NIH closed about 3,150 noncompliance cases from 2021 through 2023, the most recent years for which NIH had complete data.²⁵ Some of these cases involved more than one category of noncompliance. The two categories of noncompliance cited in the most cases were related to not following animal study protocols (about 1,210 cases) and not following policies and procedures (about 1,110 cases). For example, one institution reported tail-tipping five mice (removing a

²⁵NIH did not have complete data for 2020 when we requested it in August 2024 because the agency has a 4-year retention policy for these data, according to NIH documentation.

small portion of the tail for genetic analysis) without anesthesia or analgesia, which was not included in the researcher's approved animal study protocols and resulted in complications in one mouse.²⁶ The institution reported taking corrective actions including counseling the researcher and reviewing with all laboratory members the animal care and use committee's policy on genotyping of laboratory mice.

The third most frequent category reported was neglect/abuse (about 590 cases). In one case that NIH recorded as neglect/abuse, two mice were found alive in a freezer after failed euthanasia by research staff.²⁷ The animal care and use committee determined the incident was a continuation of a pattern of failure to employ proper animal use procedures and lack of attention to animal welfare. Accordingly, the committee voted to suspend the researcher's protocol. The committee required that the institution take corrective actions including creating a monitoring form for evaluating mice, hiring a laboratory manager to oversee mouse care and use, and requiring various personnel, including researchers, to complete training, according to the noncompliance report. In another neglect/abuse case, approximately 100 mice died or were euthanized due to dehydration during the implementation of a new animal watering system.²⁸ The animal care and use committee voted to require a written plan for implementation of the new watering system to prevent a similar incident from occurring in the future, as well as other plans and training. Table 2 shows the 10 categories reported most frequently in cases closed during this time frame.

²⁶This case included the following categories of noncompliance, according to NIH data: failure to follow animal study protocols, failure to follow institutional policies/procedures, anesthesia/analgesia, and neglect/abuse.

²⁷This case included the following categories of noncompliance, according to NIH data: failed euthanasia, surgical/post-op failures, and neglect/abuse. The incident involving mice found in the freezer was one of several issues identified in this case.

²⁸This case included the following categories of noncompliance, according to NIH data: institutional animal care and use committee-specific issues, food/water issues-husbandry, human error, neglect/abuse, and training failure.

Table 2: Categories of Noncompliance with the Public Health Service (PHS) Policy Most Frequently Reported to the National Institutes of Health (NIH) for Cases Closed, 2021–2023

| Noncompliance category | Approximate number of cases ^a that include category | Approximate percentage of cases ^a that include category |
|---|--|--|
| Failure to follow animal study protocols | 1,210 | 38 |
| Failure to follow institutional policies/procedures | 1,110 | 35 |
| Neglect/abuse | 590 | 19 |
| Significant change without approval | 490 | 16 |
| Human error | 480 | 15 |
| Food/water issues ^b | 400 | 13 |
| Anesthesia/analgesia | 390 | 12 |
| Other husbandry deviation | 290 | 9 |
| Surgical/post-op care failures | 210 | 7 |
| Equipment failure | 170 | 5 |

Source: GAO analysis of NIH data. | GAO-25-107140

Note: We rounded numbers of noncompliance cases to the nearest 10 to show approximate numbers of cases because NIH data on noncompliance cases could not be analyzed electronically without modifying the data, potentially resulting in small discrepancies in counts. For the purposes of this report, noncompliance includes (1) any serious or continuing noncompliance with the PHS Policy, (2) any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or (3) any suspension of an activity by the animal care and use committee.

^aNIH closed about 3,150 noncompliance cases from 2021 through 2023. Some cases include more than one category of noncompliance, so the total adds up to more than 3,150 cases and more than 100 percent.

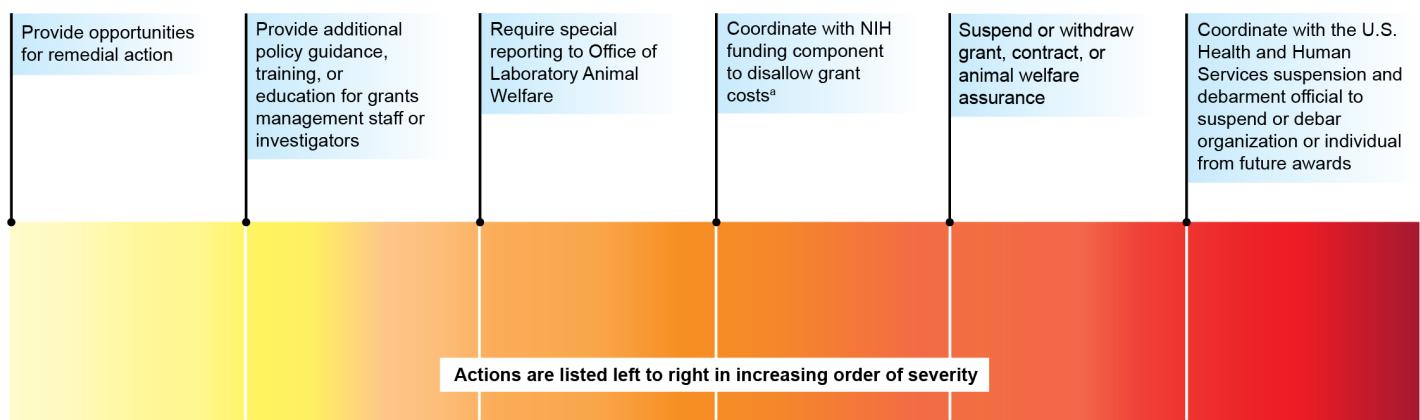
^bFood/water issues includes the following two NIH categories: food/water restriction issues and food/water issues – husbandry.

Other types of noncompliance that NIH recorded include accidents, such as cage flooding, out-of-date drugs, and failed euthanasia. Additional categories of noncompliance cases that NIH tracked for 2021 through 2023 are listed in appendix I.

Figure 2 shows actions NIH can take in response to noncompliance. From 2020 through 2023, NIH did not suspend grants or take more severe actions for any domestic institutions in response to noncompliance, according to NIH officials. In addition, the Health Research Extension Act of 1985, as amended, requires that institutions be given a reasonable opportunity to take corrective action before NIH is

to suspend or revoke a grant or contract.²⁹ Officials told us that they provide opportunities for corrective action before taking more severe actions, which they indicated is consistent with the act.

Figure 2: Actions NIH Can Take in Response to Noncompliance with the Public Health Service (PHS) Policy for Research Involving Animals



Source: GAO analysis of National Institutes of Health (NIH) information. | GAO-25-107140

Note: For the purposes of this report, noncompliance includes (1) any serious or continuing noncompliance with the PHS Policy, (2) any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or (3) any suspension of an activity by the animal care and use committee.

^aThe NIH funding component is the NIH institute or center that funded the grant or contract (i.e., the award).

Multiple Challenges Limit the Reproducibility and Translatability of Animal Research

While animal studies have contributed to the development of treatments and cures for human diseases, the majority of animal research does not result in an approved drug or treatment that would benefit human health,

²⁹42 U.S.C. § 289d(d)(2).

in part because of low success rates in reproducibility and translation.³⁰ In our review of the literature and interviews with scientific researchers, we identified a number of challenges that limit the reproducibility and translatability of animal research. Those challenges fall into three areas: modeling human biology, study design and data analysis, and reporting methodologies and results.

Low Success Rates

Researchers have reported low success rates when analyzing previously published animal research for reproducibility and translation. Specifically, researchers have estimated that 10 to 30 percent of animal research is reproducible and that 5 to 10 percent of therapies tested in animals result in approved medical treatments for humans.³¹ These measures are significant for several reasons. For example, if the results of an animal research study are successfully reproduced, this indicates that the results of this study are more likely to be reliable. In addition, many preclinical animal studies are done with the intent to translate the treatments to humans. For example, scientists may use an animal study to test a new drug with the expectation that if it is successful, they will test the drug in humans.

Reproducibility and translation of animal research cannot be expected to be successful all the time, according to literature we reviewed and scientific researchers we interviewed. For example, early exploratory research is not necessarily designed to be reproducible and can be considered an important part of the scientific discovery process, scientific researchers told us. Failures to reproduce and translate some animal research to humans can also serve as learning opportunities to better

³⁰NIH has acknowledged concerns over what some researchers call a “reproducibility crisis” in animal research. For example, see National Institutes of Health, Advisory Committee to the Director, *ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research Final Report* (Bethesda, Md.: June 11, 2021). Scientific researchers have also reported observations of low success rates of reproducibility and translation in animal research. For example, see Thomas S. Reichlin, Lucile Vogt, and Hanno Würbel, “The Researchers’ View of Scientific Rigor—Survey on the Conduct and Reporting of In Vivo Research,” *PLOS One*, vol. 11, no. 12 (2016): e0165999; Stacy L. Pritt and Robert E. Hammer, “The Interplay of Ethics, Animal Welfare, and IACUC Oversight on the Reproducibility of Animal Studies,” *Comparative Medicine*, vol. 67, no. 2 (2017): 101–105; Duxin Sun et al., “Why 90% of clinical drug development fails”; and Lindsay J. Marshall et al., “Poor Translatability of Biomedical Research Using Animals—A Narrative Review,” *Alternatives to Laboratory Animals*, vol. 51, no. 2 (2023): 102–135.

³¹Asher Mullard, “Parsing clinical success rates,” *Nature Reviews Drug Discovery*, vol. 15, no. 7 (2016): 447; Duxin Sun et al., “Why 90% of clinical drug development fails”; Benjamin Ineichen et al., “Analysis of animal-to-human translation”; and Emma Wilson et al., “Designing, conducting, and reporting reproducible animal experiments.”

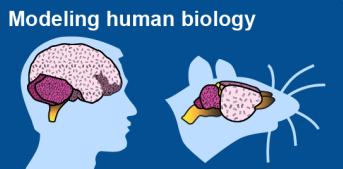
address the causes of the failures. For example, analyzing why certain studies did not reproduce or translate can help researchers improve the selection of animal models,³² identify and address challenges in study design, and improve the reporting of methodologies and results in scientific publications.

However, low rates of reproducibility and translation may reflect study designs that result in considerable financial costs and numbers of animals used for research that never translates to clinical trials, and in some cases, harmful effects to humans in clinical trials. For example, one drug designed to treat multiple sclerosis showed promise in animal models but caused severe adverse reactions in human trials. Another drug showed substantial promise in treating stroke in animals but failed in human trials. Failures like these can lead researchers down unproductive lines of scientific inquiry and can set back progress by years or even decades, according to literature we reviewed.

These low rates of reproducibility and translation have led the scientific community to examine the reasons for these failures, better understand the challenges that limit reproducibility and translatability, and take steps intended to address them. Through our review of the literature and interviews with scientific researchers, we identified eight challenges, which we organized into three key areas as shown in figure 3 and which we discuss in more detail below.

³²An animal model is a nonhuman species used in biomedical research because it can mimic aspects of a biological process or disease found in humans.

Figure 3: Challenges That Limit the Reproducibility and Translatability of Animal Research

| Challenge area | Challenge | Description |
|---|---|---|
| Modeling human biology  | Biology of animal models does not always resemble human biology | Animal models' responses to experimental treatments or drugs do not always resemble a human response because of biological differences. Consequently, a treatment that is successful in animals may not be in humans. |
| Study design and data analysis  | Lack of blinding Lack of randomization Lack of inclusion and exclusion criteria Small sample sizes Inappropriate application and interpretation of statistics | Researchers sometimes do not design the study to conceal which animals receive which treatments from the individuals working on the study. As a result, those individuals' expectations and unconscious biases may influence how they handle animals and interpret results. Researchers sometimes do not randomly decide which animals go into which group in an experiment. Study results may then be incorrectly attributed to experimental treatments rather than other factors. Researchers sometimes do not define how animals are selected for the study before the experiment starts and data are collected. Without this, it may be difficult to detect whether the results are caused by the experimental treatment or other factors. Researchers sometimes use sample sizes that are too small to detect the true effect of a treatment. As a result, treatments may appear effective in animals but fail when tried in other groups of animals and humans. Researchers sometimes do not correctly apply statistical tests or interpret statistical outcomes. For example, researchers may interpret statistical significance to mean a study is reproducible when it is not. |
| Reporting  | Incomplete reporting of methodologies Incomplete reporting of results | Researchers sometimes do not include enough details in their published studies about the methods they use and how animals were cared for. As a result, other researchers may not have enough information to successfully reproduce these studies. Researchers sometimes prioritize publishing positive results that show a treatment has the desired effect and choose not to publish negative results. Without information about both positive and negative results, subsequent researchers may have difficulty determining whether they successfully reproduced or translated the results. |

Sources: GAO analysis of scientific literature, NIH documents and responses to GAO's questions, and interviews with scientific researchers; GAO (illustrations). | GAO-25-107140

Modeling Human Biology

Because the biology of animal models does not always resemble human biology, drugs and therapeutics that are successful in animals often are not successful in humans. Animals and humans have inherent biological differences, which include how certain body systems function, how diseases manifest and progress, and how treatments interact with the body. For example, the field of pain research, in which rodents are commonly used, has produced almost no new approved treatments for chronic pain for decades. This has been attributed in part to the differences in how rodent and human bodies sense and react to pain.

In addition, treatments for human diseases including stroke, multiple sclerosis, and Alzheimer's disease have shown promise in animals but

have generally not translated to successful human treatments. For example, according to a study published in 2017, the low success rate of translating Alzheimer's disease treatments from animal models to humans was largely because certain mouse models only resembled some aspects of Alzheimer's disease in humans.³³ However, recent advances led to mouse models that more closely resemble human Alzheimer's disease, leading in 2024 to one of the first U.S. Food and Drug Administration approved treatments.

Some animal models are more effective than others at imitating certain aspects of human biology. For example, nonhuman primates, such as monkeys and apes, are often used to study neurological diseases because of their highly developed brains, while pigs are used for cardiovascular research because their heart anatomy closely resembles that of humans. However, researchers do not always select animal models on the basis of how similar they are to humans. Researchers may sometimes base their selection on other factors such as cost, availability, ease of handling, available expertise, and tradition (e.g., a researcher whose lab has always used mice may continue to use mice).

Study Design and Data Analysis

Challenges related to study design and data analysis can limit reproducibility by introducing bias into experimental design and can cause results to skew toward certain outcomes, such as showing that a treatment is effective regardless of the actual effectiveness of the treatment, according to literature we reviewed and scientific researchers we interviewed.

The five challenges we identified in this area relate to (1) blinding, (2) randomization, (3) inclusion and exclusion criteria, (4) small sample sizes, and (5) applying and interpreting statistics.³⁴

- **Lack of blinding.** Without blinding, a researcher's expectations or unconscious biases may influence how they handle animals and how they interpret animal behavior or results. For example, if a researcher

³³Eleanor Drummond and Thomas Wisniewski, "Alzheimer's Disease: Experimental Models and Reality," *Acta Neuropathologica*, vol. 133, no. 2 (2017): 155–175.

³⁴In addition to these challenges, other questionable research practices related to study design can limit reproducibility and translatability. For example, according to academics who study research, it can be tempting for some researchers to develop a hypothesis after they have collected and analyzed data. This seemingly innocuous practice results in a greater likelihood that the study will report spurious results. This practice can come in the form of HARKing (hypothesizing after results are known) or "p-hacking" (manipulating data analyses to enable favored results to be presented as statistically significant).

knows which animals will receive surgery and which will not, the researcher may handle the animals differently, potentially affecting study outcomes. Another researcher who tries to reproduce an unblinded study may have different expectations or unconscious biases, or no biases, leading to different results. A study published in 2017 by a team of independent researchers assessed almost 3,400 preclinical peer-reviewed cardiovascular studies for characteristics that promote reproducibility.³⁵ This team found blinding in about 33 percent of studies, concluding that flawed study design was prevalent in this field and had not improved over the prior 10 years, potentially hindering progress in the field of cardiovascular medicine.³⁶

- **Lack of randomization.** Without randomization, researchers may choose animals for certain treatment groups on the basis of preexisting characteristics or arrange treatment groups based on convenience. This can introduce variation that may lead to biased study outcomes. For example, researchers might place cages for an experimental group at eye level for convenience (see fig. 4), but placement of cages on a rack can influence outcomes due to extrinsic factors. This is because different locations on the rack might experience varying environmental conditions such as light. Randomizing the placement of treatment animals in cages on a rack reduces the likelihood that extrinsic factors will bias the results. The 2017 study mentioned above found randomization in about 22 percent of the studies it assessed.³⁷

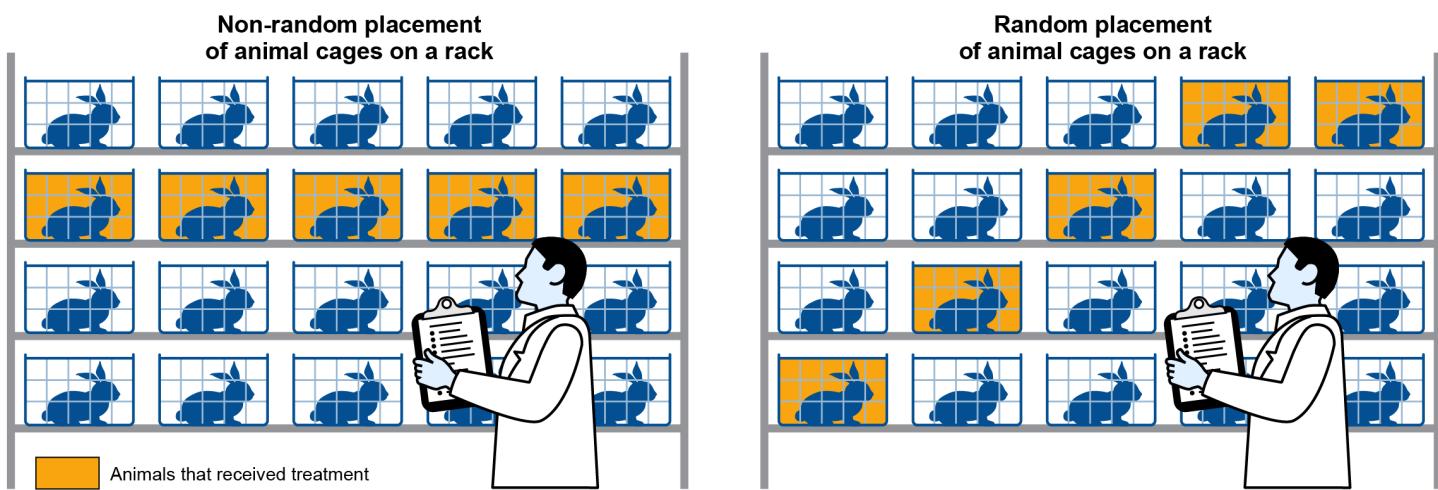
³⁵F. Daniel Ramirez et al., “Methodological Rigor in Preclinical Cardiovascular Studies: Targets to Enhance Reproducibility and Promote Research Translation,” *Circulation Research*, vol. 120, no. 12 (2017): 1,916–1,926.

³⁶F. Daniel Ramirez et al., “Methodological rigor in preclinical cardiovascular studies.”

³⁷F. Daniel Ramirez et al., “Methodological Rigor in Preclinical Cardiovascular Studies.”

Figure 4: Different Arrangements for Placement of Animal Cages

Animal researchers may arrange animal cages for convenience (left); however, the random arrangement of animal cages on a rack (right) is considered better study design.



Sources: GAO analysis of scientific literature and interviews with scientific researchers; GAO (illustrations). | GAO-25-107140

- **Lack of inclusion and exclusion criteria.** By not applying or reporting these criteria, researchers may introduce bias that can affect study outcomes. For example, an animal's bodyweight may need to fall within a certain range in order to be included for a specific procedure, and those that do not should be excluded. Without this type of information, other researchers may not be able to reproduce a study. A study published in 2016 reported that only 4 percent (2 out of 47) of the studies they examined described inclusion and exclusion criteria.³⁸
- **Small sample sizes.** When researchers attempt to reproduce a study that used small sample sizes, they are less likely to obtain the same observations because the results of underpowered studies may reflect chance rather than true effects.³⁹ Researchers sometimes are pressured to reduce the number of animals they use (i.e., sample size) for ethical reasons or to reduce costs, according to literature we

³⁸Marc T. Avey et al., "The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research," *PLOS One*, vol. 11, no. 11 (2016): e0166733.

³⁹Studies with low statistical power, also called underpowered studies, are those where a statistical test has a low chance of detecting a true effect. Insufficient sample size is one cause of low statistical power.

Different Types of Bedding Used in Mouse Cages

One example of an extrinsic factor that can influence research outcomes is the type of bedding used for laboratory mice, which can cause changes in the mice that may be interpreted as changes resulting from experimental treatments. The type of bedding used can affect mice's respiratory systems, immune systems, and body weight, among other things. In studies where the type of bedding is not reported, the extent to which variations in bedding may have contributed to observed differences among experimental groups may not be clear. The photo below provides examples of bedding types in bedded cages.



The four bedding materials are (A) shaved aspen, (B) 1/4-in. corncobs, (C) 1/8-in. pelleted cellulose, and (D) refined virgin diced cellulose. The photo below shows mice with bedding in a cage.



Sources: GAO analysis of scientific literature, National Institutes of Health (NIH) working group report, and an interview with a scientific researcher (text); American Association for Laboratory Animal Science and NIH (photos). | GAO-25-107140

reviewed and scientific researchers we interviewed. Researchers we interviewed said it can be challenging to balance the competing direction they receive to reduce the number of animals against using sample sizes that are large enough to obtain reliable results. For example, U.S. government principles provide that the animals selected for a procedure should be of an "appropriate species and quality and the minimum number required to obtain valid results."⁴⁰ A study published in 2018 of 410 neuroscience experiments using rodents showed that 88 percent of experiments did not use sample

⁴⁰The U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training were incorporated into the PHS Policy in 1986 and continue to provide a framework for conducting research in accordance with the PHS Policy. The principles are supplemented and implemented by the PHS Policy and the *Guide for the Care and Use of Laboratory Animals*.

sizes large enough to detect the true effects of treatments.⁴¹ As a result, researchers aiming to reproduce these experiments would likely find different results.

- **Inappropriate application and interpretation of statistics.** Animal researchers sometimes do not correctly apply the statistical tests or analyze and interpret their data correctly. This can result in misleading research publications that may not be reproducible or translatable. For example, researchers may calculate statistical significance to help determine whether a study has successfully reproduced previous results. However, this interpretation can be problematic because statistical significance does not always mean that a study's result will be reproducible. Researchers may choose small sample sizes or inappropriately apply statistics if they do not have sufficient statistical training, good collaborations with statisticians, or the resources to hire statistical consultants, according to literature we reviewed and scientific researchers to whom we spoke.

Reporting

Incomplete reporting of methodologies or results can limit the reproducibility and translatability of animal research.

- **Incomplete reporting of methodologies.** When researchers do not publish certain aspects of their methodology, other researchers trying to reproduce an experiment may follow a different approach and may obtain different or inconsistent results, according to literature we reviewed and scientific researchers we interviewed. Researchers sometimes do not report aspects of their methodology related to intrinsic factors, such as the proportion of male or female animals used in their study, the ages of animals used, or genetic type (i.e., strain). These details may significantly affect outcomes, and without this information, other researchers may not be able to reproduce a published study. A study published in 2020 estimated that basic animal characteristics (e.g., sex and age) are reported in fewer than 10 percent of research publications.⁴²

Researchers also sometimes do not report extrinsic factors that are under their control or the control of animal care staff, according to our analysis (see fig. 5). Some of these factors maybe be controllable,

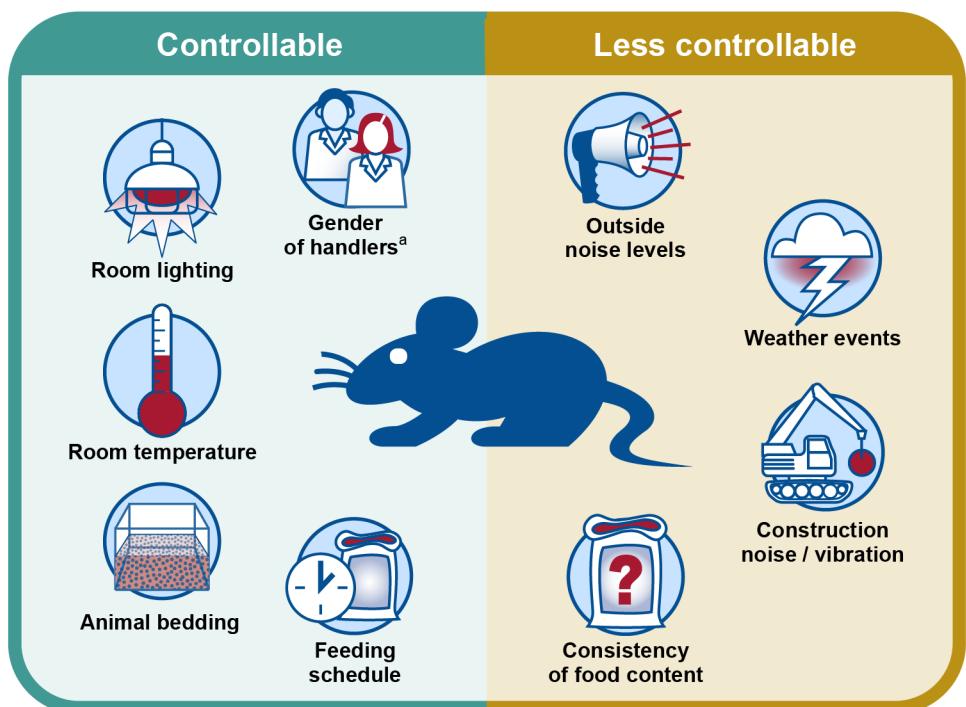
⁴¹Clarissa F. D. Carneiro et al., "Effect size and statistical power in the rodent fear conditioning literature – A systematic review," *PLOS One*, vol. 13, no. 4 (2018): e0196258. This study reviewed 122 articles.

⁴²Natalie Percie du Sert et al., "The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research," *PLOS Biology*, vol. 18, no. 7 (2020).

such as the type of bedding used, room temperature and lighting, and feeding schedule. Other extrinsic factors may be less controllable or unknowable such as noise from weather and construction and unexpected changes in the ingredients of animal feed. Except for temperature and humidity, most of the key extrinsic variables present in animal housing spaces and research laboratories are either not reported or are subjectively evaluated, according to a review published in 2024.⁴³

Figure 5: Examples of Controllable and Less Controllable Extrinsic Factors That Can Influence Outcomes in Animal Research

When publishing results of animal research, researchers sometimes do not include extrinsic factors that can influence study outcomes. Some of these factors are controllable by the researchers and animal care staff, while others are less controllable.



Sources: GAO analysis of scientific literature, NIH report and responses to GAO's questions, and interviews with scientific researchers; GAO (illustrations). | GAO-25-107140

^aRecent studies have found that animals may experience different levels of stress depending on the gender of their handler. See Polymnia Georgiou et al., "Experimenters' sex modulates mouse behaviors and neural responses to ketamine via corticotropin releasing factor," *Nature Neuroscience*,

⁴³Jeremy G. Turner et al., "Extrinsic Environmental Variables: The Umwelt of Research Animals and the Implications for the 3Rs and Study Reproducibility," *Journal of the American Association for Laboratory Animal Science*, vol. 63, no. 2 (2024): 106.

vol. 25, no. 9 (2022): 1,191–1,200; and Alicia S. Zumbusch et al., “Normative preclinical algesiometry data on the von Frey and radiant heat paw-withdrawal tests: an analysis of data from more than 8,000 mice over 20 years,” *The Journal of Pain*, vol. 25, no. 7 (2024).

- **Incomplete reporting of results.** Animal researchers sometimes selectively publish positive results (i.e., results that indicate that a treatment had the desired effect) and not negative results (i.e., results that indicate that a treatment did not have the desired effect). This selective reporting, known as publication bias, may be a result of pressure from institutions and the broader scientific community to report results that are interesting, novel, and statistically significant, according to literature we reviewed. Negative results are sometimes not published.

The effect of publication bias on animal research is difficult to measure, but some researchers have stated that pressure to publish and selective reporting of results have a negative impact on reproducibility. A 2012 survey of 454 animal researchers estimated that 50 percent of animal experiments were not published, and the top cause was “lack of statistically significant differences (‘negative’ findings).”⁴⁴ Some researchers have suggested that this publication bias may also contribute to the high failure rate in translation of animal research to human clinical trials. When researchers prioritize reporting positive results, this can make it appear that certain treatments are more effective than they actually are. Subsequent researchers may then find it difficult to reproduce and translate published results.

NIH Has Taken Steps Intended to Enhance Reproducibility and Translatability but Has Not Assessed Its Progress

NIH has taken steps intended to enhance reproducibility and translatability in animal research it conducts and supports, such as implementing a rigor and transparency policy and establishing a working group to identify opportunities for improvement. However, NIH has not determined whether these steps have helped the agency make progress toward these goals.

⁴⁴Gerben ter Riet et al., “Publication Bias in Laboratory Animal Research: A Survey on Magnitude, Drivers, Consequences and Potential Solutions,” *PLOS One*, vol. 7, no. 9 (2012).

NIH Implemented a Rigor and Transparency Policy Aimed at Enhancing Reproducibility of Extramural Research

NIH's approach to addressing challenges that limit reproducibility and translatability in animal research is to take steps aimed at increasing the rigor and transparency of the research it funds, according to agency officials. For example, in 2015, NIH issued a policy entitled *Enhancing Reproducibility through Rigor and Transparency*, which applies to extramural research grants, including animal research grants, and includes guidance and resources for extramural applicants and grantees.⁴⁵ This policy instructs applicants to address four areas of rigor in grant applications and directs peer reviewers to evaluate the same areas when scoring applications. See table 3 for more information about the four areas of rigor.

Table 3: National Institutes of Health's (NIH) Four Areas of Rigor for Grant Applications

| Area of rigor | Instructions to applicants |
|--|--|
| Rigor of the prior research ^a | Applicants should describe the strengths and weaknesses in the rigor of the prior research that applicants use as the key support for the proposed research project. Applicants should also describe their plans to address the identified weaknesses in the prior research. |
| Scientific rigor | Applicants should describe how the experimental design and proposed methods will achieve robust and unbiased results. They should also describe plans to reduce bias, such as using randomization. |
| Consideration of biological variables | Applicants should explain how they will factor biological variables such as the sex and age of the animal into the research design, analysis, and reporting. Applicants must provide a strong justification for applications proposing to study only one sex. |
| Authentication of key resources | Applicants should describe methods to ensure the identity and validity of key biological resources (e.g., antibodies) and chemical resources (e.g., specialty chemicals) used in the proposed studies. These key resources may differ over time or between laboratories. These differences may affect the outcomes of proposed studies, so applicants should take steps to verify that these resources are authentic throughout their studies. |

Source: GAO summary of NIH policy and guidance to applicants and reviewers. | GAO-25-107140

^aIn 2018, NIH updated its application instructions to replace the term "scientific premise" with "rigor of the prior research."

As part of the implementation of the 2015 policy, NIH also issued updated guidance for researchers on completing annual progress reports or grant

⁴⁵The policy applies to grant applications but not intramural research conducted by NIH scientists. For intramural research, Boards of Scientific Counselors conduct reviews of intramural researchers, which are primarily retrospective and based on scientific accomplishments since the last review.

close-out reports for their research projects.⁴⁶ This guidance states that researchers should describe the approaches they took to ensure robust and unbiased results in both the past year and the upcoming reporting period, if applicable. NIH program officials are to review all progress reports to determine whether the researchers provided sufficient information to address these questions, according to NIH officials. For example, the officials told us that in one NIH-funded research project, a grant application proposed using mice of both sexes, but the progress report described studies using only male mice. The NIH program official reminded the grantee of the importance of using both sexes and asked for an updated progress report. The grantee submitted a progress report stating that future studies would use both male and female mice.

NIH Has Taken Additional Steps Intended to Enhance Reproducibility and Translatability

NIH has taken additional steps intended to enhance the reproducibility and translatability of research it supports, including animal research in particular. For example, in 2019, the Director of NIH established a working group to examine issues such as translatability in animal research.⁴⁷ According to a 2021 statement announcing the working group's findings, the Director created the working group because he believed improving animal research required additional attention from NIH. The working group's charge included identifying gaps and opportunities to improve rigor, transparency, reproducibility, and translatability in animal research and evaluating how to improve the use of animal models. The working group's 2021 report included 19 recommendations and associated sub-recommendations on steps NIH could take to enhance the reproducibility and translatability of animal research.⁴⁸ (See appendix II for a full list of the recommendations.)

During the course of our review, we asked NIH officials to provide us with information on the agency's efforts to implement the working group's

⁴⁶National Institutes of Health and Agency for Healthcare Research and Quality, *Updates to NIH & AHRQ Research Performance Progress Reports (RPPR) to Address Rigor and Transparency*, NOT-OD-16-031 (Washington, D.C.: Dec. 15, 2015). NIH's Grants Policy Statement describes when grant recipients are required to submit these progress reports and grant close-out reports. See National Institutes of Health, *NIH Grants Policy Statement*.

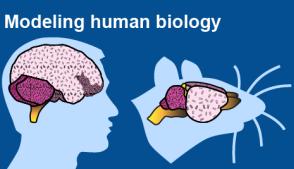
⁴⁷The working group had two co-chairs, one from NIH and one from a university. Members included NIH leaders in intramural and extramural research and representatives from other parts of the U.S. government, academia, industry, and scientific journals.

⁴⁸National Institutes of Health, Advisory Committee to the Director, *ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research Final Report*.

recommendations.⁴⁹ NIH officials told us that it would be labor intensive and time consuming to provide a status for each recommendation. Instead, NIH provided examples of steps it has taken that are consistent with the recommendations (see fig. 6). These steps could help address the challenges we describe earlier in this report.

⁴⁹This working group is under NIH's Advisory Committee to the Director. The Advisory Committee to the Director is governed by the provisions of the Federal Advisory Committee Act, as amended, under which advisory committees are to be utilized solely for advisory functions. See 5 U.S.C. § 1008(b). The act further provides that solely the President or an officer of the federal government is to make determinations of actions to be taken and policy to be expressed with respect to matters upon which an advisory committee reports or makes recommendations. *Id.*

Figure 6: Examples of Challenge Areas, Recommendations, and Steps NIH Has Taken Intended to Enhance Reproducibility and Translatability of Animal Research

| Challenge area | Working group recommendation ^a | Step NIH took | How the step could address the challenge area |
|---|--|--|---|
| Modeling human biology  | NIH should work to improve the design of animal models through the funding of focused research programs that enhance understanding of comparative human–animal biology. | Funding to study animal models. NIH released a funding notice in December 2021 to support projects aimed at developing technologies, tools, and resources for validating animal models. No applications were funded under this funding notice. The agency released a similar funding notice in 2022 that is open for applications until July 2025. | Funding to better validate animal models can enhance the rigor and reproducibility of animal models and could help researchers understand differences between human and animal biology. |
| Study design and data analysis  | NIH should develop statistical resources specifically for animal researchers. | Statistical resources. NIH published web pages in 2023 for intramural and extramural researchers with statistical resources, including resources for calculating sample sizes. According to NIH guidance, intramural and extramural researchers are expected to provide a statistical justification for the sample size in proposed research studies. | These resources could prompt researchers to use sample sizes—an element of study design—that are large enough to achieve sufficient statistical power. |
| Reporting  | NIH should expect all animal research to include 10 key elements of study design, procedures, and results in publications. ^b | Reporting guidelines. NIH issued a notice in February 2023 encouraging grantees to report the 10 key elements in NIH-funded animal research publications. The 10 elements include randomization, blinding, inclusion and exclusion criteria, and sample size. The agency also issued a notice to intramural researchers in 2024 endorsing these reporting guidelines. | The guidelines could encourage researchers to report their methodologies more fully, and to include important elements of study design, such as randomization, in the planning stage for their experiments. |
| Reporting  | NIH should encourage and support work to better understand, monitor, record, and report important extrinsic factors related to animal care that may impact research results. | Funding to study extrinsic factors. NIH released a funding notice in March 2024 to support projects aimed at studying the roles of critical extrinsic factors in biological, behavioral, and treatment studies using animal models. | Studying extrinsic factors could help researchers determine what factors can influence research outcomes and thus should be reported in publications so that others can try to reproduce the results. |

Sources: GAO analysis of National Institutes of Health (NIH) documents; GAO (illustrations). | GAO-25-107140

Note: We reviewed scientific publications and interviewed scientific researchers to identify challenges that limit the reproducibility and translatability of animal research. We identified eight challenges, which we grouped into three challenge areas and discuss in more detail in our report. In reviewing the NIH working group's recommendations, we determined which of our challenge areas these recommendations addressed. We then determined steps that NIH took that were consistent with these recommendations.

^aRecommendations are from National Institutes of Health, *Advisory Committee to the Director, ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research Final Report* (Bethesda, Md.: June 11, 2021). For a full list of the recommendations, see appendix II.

^bFor additional information on these 10 key elements, see National Institutes of Health, *NIH Encourages the Use of the ARRIVE Essential 10 Checklist in All Publications Reporting on the Results of Vertebrate Animal and Cephalopod Research* (Bethesda, Md.: Feb. 10, 2023). The ARRIVE (Animal Research: Reporting of In Vivo Experiments) Essential 10 guidelines were developed by an international working group with support from the National Centre for the Replacement, Refinement & Reduction of Animals in Research. The guidelines describe 10 minimum elements of study design, procedures, and results that researchers should report in publications so that readers and reviewers can assess the reliability of the research findings.

In July 2024, NIH announced a replication initiative to reproduce significant lines of research. According to the initiative's website, the effort will explore if, and under what conditions, directly reproducing certain research studies is an effective approach for improving reproducibility.⁵⁰

Individual NIH institutes have also taken steps to enhance reproducibility and translatability, including for animal research. For example, the National Institute on Aging and the NIH Library developed the Alzheimer's Disease Preclinical Efficacy Database, a public database of preclinical studies on Alzheimer's disease that tracks whether published studies report specific elements of study design, such as randomization.⁵¹ Also, the National Institute of Neurological Disorders and Stroke published a list of elements of rigor, in addition to those in the NIH-wide policy, that researchers should consider when applying for funding from the institute.⁵²

NIH Has Not Assessed Whether It Has Made Progress in Enhancing Reproducibility and Translatability of Animal Research

While NIH's strategic plan includes goals related to enhancing reproducibility and translatability, the agency has not assessed whether the steps it has taken have led to progress toward these goals as they relate to animal research. The agency's strategic plan includes an objective to enhance reproducibility through rigorous and transparent research. The strategic plan also states that to achieve its mission, NIH strives to support research aimed at improving human health. In the case of animal research, doing so generally relies on researchers successfully translating the results of animal research to clinical trials in humans. NIH has taken steps intended to achieve these goals, including through its 2015 policy on enhancing reproducibility and additional steps described above, but the agency has not determined the results of these steps. When we asked for evidence of the effectiveness of NIH's efforts, NIH officials identified steps the agency has taken but did not provide evidence of the effectiveness of these steps in enhancing reproducibility and translatability.

⁵⁰National Institutes of Health, "Replication to Enhance Research Impact Initiative," accessed October 9, 2024, <https://commonfund.nih.gov/replication-initiative>.

⁵¹National Institute on Aging, "Alzheimer's Disease Preclinical Efficacy Database," *AlzPED*, accessed September 16, 2024, <https://alzped.nia.nih.gov/>.

⁵²National Institute of Neurological Disorders and Stroke, "Rigorous Study Design and Transparent Reporting," accessed September 18, 2024, <https://www.ninds.nih.gov/funding/Preparing-your-application/Preparing-research-plan/Rigorous-study-design-and-transparent-reporting>.

Federal decision makers need evidence about whether federal programs and activities are achieving intended results so they can set priorities and identify ways to improve programs, as we have previously reported.

Specifically, GAO's *Key Practices for Evidence-Based Policymaking* describes key practices that can help agencies use evidence to assess the results of federal efforts.⁵³ Selected practices include the following:

- **Defining goals.** An agency identifies long-term goals for how the agency will advance its mission and short-term goals with targets and time frames against which an agency can measure performance.
- **Building evidence and assessing results.** An agency collects new evidence to help understand whether it is making progress toward its goals.⁵⁴
- **Making decisions.** An agency uses the evidence it has collected to inform decisions such as changes to policies or funding.

Adopting these practices can help agencies define what they are trying to achieve, determine how well they are doing, and identify steps needed to improve their efforts.⁵⁵

NIH's actions to date are consistent with some aspects of these practices, such as by establishing long-term goals and collecting some evidence. However, NIH has not fully implemented the above mentioned three practices that would help it assess its progress toward enhancing reproducibility and translatability of animal research:

Defining goals. While NIH has established long-term goals, it has not developed short-term goals with targets and time frames. As we described above, NIH's strategic plan includes a strategic objective to enhance reproducibility and a goal to advance human health, which relies on translatability. Establishing long-term goals like this is an important step in the process for assessing the agency's results. However, NIH does not have short-term goals with targets and time frames against which the agency could measure its progress toward this long-term goal. Such short-term goals could include, for example, setting a target for the percentage of applications that follow its 2015 policy on enhancing

⁵³[GAO-23-105460](#).

⁵⁴This includes identifying new evidence needs and the types of evidence that would address these needs as well as taking steps to ensure the quality of the evidence that the agency collects.

⁵⁵[GAO-23-105460](#).

reproducibility or the percentage of publications that include the 10 key elements of research design and methods described above, as well as identifying when the agency aims to achieve these targets. Without measurable short-term goals, it will be difficult for NIH to assess whether it is making progress in this area, particularly given the complexity of assessing reproducibility and translatability.

Building evidence and assessing results. NIH collected limited evidence related to its 2015 policy but has not collected evidence it would need to assess whether its efforts are resulting in progress toward its goals. In 2016, NIH engaged with a contractor to conduct a pilot project evaluating the extent to which grant applicants were following its 2015 policy. However, in the pilot project, some agency staff reviewing the applications had different interpretations about whether applicants followed the policy because the staff found some parts of the policy to be unclear, according to documentation of the pilot project.⁵⁶ NIH ended the evaluation in 2017 without obtaining data on how many applicants were following the policy agencywide.⁵⁷ In addition, during a 2018 public meeting, the Director of NIH raised questions about whether the 2015 policy was being rigorously enforced during application reviews, and an NIH official said the policy was not being rigorously enforced at the time.⁵⁸ Collecting and assessing information on the impacts of its policy would help ensure NIH leadership is aware of how it is implementing and enforcing the policy, and position it to make informed decisions about addressing such issues.

Individual institutes have collected information that could help inform subsequent NIH evaluations. For example, in 2017, one NIH institute

⁵⁶RTI International, *NIH Rigor and Reproducibility Evaluation Interim Pilot Test Report* (Washington, D.C.: 2017).

⁵⁷NIH officials said that the agency issued a stop work order on the project in 2017. Officials said it would take a high level of effort to evaluate the 2015 policy using a similar approach to the one above of review by agency staff. The officials told us that the agency is developing a different approach to evaluate its policy. Specifically, agency officials told us about an ongoing project that will use artificial intelligence to develop indicators of rigor in grant applications. While this project could be a positive step in the process of collecting evidence, NIH did not provide details of how the project would assess progress toward its goals of enhancing reproducibility and translatability.

⁵⁸The purpose of this meeting was to discuss updates and recommendations from NIH's Advisory Committee to the Director and its working groups. See National Institutes of Health, "Advisory Committee to the Director – June 2018 (Day 1)" (Bethesda, Md.: June 14, 2018), accessed September 19, 2024, <https://videocast.nih.gov/Summary.asp?File=23957&bhcp=1>.

assessed its applicants' inclusion and interpretation of certain criteria in the 2015 policy. The institute found that both applicants and reviewers inadequately addressed these criteria.⁵⁹ This assessment recommended that NIH revise a portion of its application instructions, which the agency did in 2018. However, as of September 2024, NIH had not built on this effort by collecting agencywide information on what percentage of applicants were following the 2015 policy or evaluating the effects of the 2018 change on applicants' compliance with the policy, according to NIH officials.

Collecting additional evidence would help NIH assess whether the steps it has taken are helping enhance reproducibility and translatability in the research it conducts and supports, as well as determine whether any changes are needed. Such evidence could include

- high-quality information on whether grant applicants are following NIH's 2015 policy and whether these applicants' research results are becoming more rigorous and reproducible;⁶⁰
- building on results from a 2023 institute-level analysis that found that after NIH implemented its 2015 policy, authors of NIH-funded publications on Alzheimer's disease reported certain elements of rigor more frequently; and⁶¹

⁵⁹Specifically, the evaluation looked at how 84 applications addressed the strengths and weaknesses of the prior research—including its methodological rigor—and the rigor of the proposed research. For example, regarding the methodological rigor of prior studies, the evaluation found that many applications (68 percent) and reviewers' statements (74 percent) did not discuss it.

⁶⁰In 2022, we previously recommended that NIH collect information on relevant indicators of rigor to assess the research projects the agency funds and implement steps, as needed, to promote strong research practices in future work. See [GAO-22-104411](#). While NIH concurred with our recommendation, the agency had not implemented it as of August 2024. Implementing our prior recommendation could also help the agency determine whether researchers are complying with its 2015 policy on enhancing reproducibility.

⁶¹National Institutes of Health, National Institute on Aging, *From Mouse to Medicine: Optimizing the Predictive Value of Preclinical Research* (Washington, D.C.: 2023).

- analyses of attempts to reproduce NIH-funded studies or information from projects to reproduce significant lines of research, such as the replication initiative we described earlier in this report.⁶²

Making decisions. Because NIH has not set short-term goals or collected usable evidence on its progress, it is not able to use this evidence to inform its decisions about policy or funding changes. For example, evidence on the effectiveness of NIH's current efforts could inform decisions about whether additional revisions to the 2015 policy are needed or about how the agency allocates funding among animal research and other types of research.

NIH officials told us they have not defined short-term goals or collected evidence that would enable the agency to measure the effectiveness of its efforts because there is variability among different fields of study that would require specific goals and measures for each field. For example, such variability includes the animal model being used, research methodology, and outcomes being measured, among other factors, according to agency officials. However, where factors differ across fields, we have previously reported that agencies can set specific targets and time frames for different areas and assess the contributions of each area to an agency's long-term goals.⁶³ Also, some factors that affect reproducibility and translatability—such as sample size and use of randomization—are similar across different fields of study and could be measured broadly.

Defining short-term goals and collecting relevant evidence would help NIH better assess whether its efforts are helping enhance reproducibility and translatability in animal research—in turn increasing its benefits to human health. These practices would also help congressional and agency decision-makers to make better-informed decisions about animal research while considering resource constraints and challenges.

Conclusions

NIH spent billions of dollars in fiscal year 2024 on research, including on projects that involve animals. Animal research has contributed to important advances in treatments to benefit human health. However, such

⁶²Previous studies have used this type of meta-analysis to study rates of reproducibility and translation in animal research by reviewing multiple published studies. For an example involving reproducibility, see Bernhard Voelkl et al., "Reproducibility of preclinical animal research improves with heterogeneity of study samples," *PLOS Biology*, vol. 16, no. 2 (2018). For an example involving translation, see Benjamin Ineichen et al., "Analysis of animal-to-human translation."

⁶³[GAO-23-105460](#).

advances depend on researchers being able to reproduce and translate the results of animal research to humans. Multiple challenges, such as differences between human and animal biology and flawed study design, limit researchers' ability to reproduce and translate the results of animal research. In part because of these challenges, many treatments that researchers find to be successful in animals cannot be reproduced or translated to humans.

NIH has taken steps intended to enhance the reproducibility and translatability of the animal research it conducts and supports but has not assessed whether the agency has made progress toward its goals. Specifically, the agency has not developed short-term goals or collected evidence it could use to assess its efforts and inform its decisions—practices we have identified in prior work as effective for assessing the results of federal efforts. Defining short-term goals and collecting relevant evidence would help NIH to better assess whether its efforts are helping enhance reproducibility and translatability in animal research—in turn increasing the benefits to human health. These practices would also help congressional and agency decision-makers to make better-informed decisions about animal research while considering resource constraints and challenges.

Recommendations for Executive Action

We are making the following two recommendations to NIH:

The Director of NIH should define short-term goals with measurable targets and time frames related to enhancing reproducibility and translatability in animal research that the agency conducts and supports. For example, some initial goals could include targets for the number of NIH-funded publications that report certain factors that affect reproducibility and translatability, such as randomization and appropriate sample sizes. (Recommendation 1)

The Director of NIH should collect evidence needed to assess NIH's efforts to enhance reproducibility and translatability in animal research. This could include steps such as (1) analyzing attempts to reproduce NIH-funded studies or (2) collecting information from projects that attempt to reproduce significant lines of research. (Recommendation 2)

Agency Comments and Our Evaluation

We provided a draft of this report to HHS for review and comment. In its written comments, reproduced in appendix III, HHS concurred with both of our recommendations. HHS also provided technical comments, which we incorporated as appropriate.

With regard to recommendation 1, HHS said NIH was evaluating potential methods for developing indicators of rigor and reproducibility. Once it develops these indicators, NIH can use them to assess publications that result from NIH-funded research for their adherence to NIH's 2015 policy on enhancing reproducibility, according to HHS. HHS also said NIH would develop appropriate, measurable targets for this type of analysis. Regarding recommendation 2, HHS said NIH would develop plans to use these indicators of rigor to evaluate projects that attempt to reproduce NIH-funded research.

We will evaluate the responsiveness of NIH's actions once they are completed.

We are sending copies of this report to the appropriate congressional committees, the Secretary of Health and Human Services, and the Director of NIH. In addition, the report is available at no charge on the GAO website at <https://www.gao.gov>.

If you or your staff have any questions about this report, please contact Steve Morris at (202) 512-3841 or MorrisS@gao.gov or Candice N. Wright at (202) 512-6888 or WrightC@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix IV.

//SIGNED//

Steve Morris
Director, Natural Resources and Environment

//SIGNED//

Candice N. Wright
Director, Science, Technology Assessment, and Analytics

Appendix I: Categories of Noncompliance with the Public Health Service Policy

Table 4 shows categories of noncompliance with the Public Health Service Policy reported to the National Institutes of Health for cases closed from 2021 through 2023.

Table 4: Categories of Noncompliance with the Public Health Service Policy Reported to the National Institutes of Health (NIH) for Cases Closed, 2021–2023

| Noncompliance category | Approximate number of cases ^a that include category | Approximate percentage of cases ^a that include category |
|---|--|--|
| Failure to follow animal study protocols | 1,210 | 38 |
| Failure to follow institutional policies/procedures | 1,110 | 35 |
| Neglect/abuse | 590 | 19 |
| Significant change without approval | 490 | 16 |
| Human error | 480 | 15 |
| Food/water issues ^b | 400 | 13 |
| Anesthesia/analgesia | 390 | 12 |
| Other husbandry deviation | 290 | 9 |
| Surgical/post-op care failures | 210 | 7 |
| Equipment failure | 170 | 5 |
| Work begun before approval (i.e., unauthorized) | 150 | 5 |
| Unauthorized/unqualified personnel | 140 | 5 |
| Inadequate ID/record keeping | 140 | 4 |
| Training failure | 130 | 4 |
| Accident (e.g., cage flooding) | 120 | 4 |
| Out-of-date drugs | 120 | 4 |
| Institutional animal care and use committee-specific issues | 80 | 2 |
| Vet care issues ^c | 70 | 2 |
| HVAC-related issues | 70 | 2 |
| Inadequate animal study protocol oversight | 60 | 2 |
| Other | 60 | 2 |
| Other physical plant issues | 60 | 2 |
| Failed euthanasia | 60 | 2 |
| Escaped animals | 50 | 2 |
| Work under expired animal study protocol | 40 | 1 |
| Performance site not covered | 40 | 1 |

**Appendix I: Categories of Noncompliance with
the Public Health Service Policy**

| Noncompliance category | Approximate number of cases^a that include category | Approximate percentage of cases^a that include category |
|--|--|--|
| Conducted prohibited procedure | 30 | 1 |
| Space/overcrowding | 30 | 1 |
| Failure to do semiannual and/or follow-up | 30 | 1 |
| Occupational safety and health program issues | 20 | 1 |
| Emergency power/lighting | 20 | 1 |
| Natural disaster | 20 | 1 |
| Sanitation ^d | 20 | 1 |
| Construction/maintenance issues | <10 | <1 |
| Failure to report to Office of Laboratory Animal Welfare | <10 | <1 |
| Social enrichment/exercise | <10 | <1 |
| Break-in | <10 | <1 |
| Dysfunctional program | <10 | <1 |
| Theft | <10 | <1 |
| Arson | <10 | <1 |
| Storage facilities | <10 | <1 |

Source: GAO analysis of NIH data. | GAO-25-107140

Notes: We rounded numbers of noncompliance cases to the nearest 10 to show approximate numbers of cases because NIH data on noncompliance cases could not be analyzed electronically without modifying the data, potentially resulting in small discrepancies in counts. For the purposes of this report, noncompliance includes (1) any serious or continuing noncompliance with the PHS Policy, (2) any serious deviation from the *Guide for the Care and Use of Laboratory Animals*, or (3) any suspension of an activity by the animal care and use committee.

^aNIH closed approximately 3,150 noncompliance cases from 2021 through 2023. Some cases include more than one category of noncompliance, so the total adds up to more than 3,150 cases and more than 100 percent.

^bThis category includes the following two NIH categories: food/water restriction issues and food/water issues – husbandry.

^cThis category includes the following two NIH categories: vet care (surv/diag/trt/control) and vet care (procure/quar/prev med).

^dThis category includes the following two NIH categories: sanitation facilities and sanitation failures.

Appendix II: Recommendations from NIH Advisory Committee to the Director to Improve Animal Research

NIH charged the Advisory Committee to the Director (ACD) Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research with several tasks:¹

- identify gaps and opportunities to improve the rigor, reproducibility, translatability, and transparency of studies involving animal models;
- evaluate how animal models of human disease are currently developed, validated, and accepted into routine use, and how this process could be improved;
- assess the current state of science for validating alternative models to animal research;
- consider the benefits and burdens of registering animal studies that aim to lead to research in humans; and
- model the financial implications of potential changes in the average costs of grants using animal models, the number of studies funded, or the need to develop multi-lab organizations to achieve appropriate statistical power.

The working group's final report, released in June 2021, included 19 recommendations and associated sub-recommendations to NIH. The report organizes the recommendations into five themes. This working group is under NIH's Advisory Committee to the Director. Governed by the Federal Advisory Committee Act, the Advisory Committee to the Director is to be utilized solely for advisory functions, along with its working groups.²

¹The working group had two co-chairs, one from NIH and one from a university. Members included NIH leaders in intramural and extramural research and representatives from other parts of the U.S. government, academia, industry, and scientific journals.

²See 5 U.S.C. § 1008(b). The act further provides that solely the President or an officer of the federal government is to make determinations of actions to be taken and policy to be expressed with respect to matters upon which an advisory committee reports or makes recommendations. *Id.*

**Appendix II: Recommendations from NIH
Advisory Committee to the Director to Improve
Animal Research**

Table 5: Recommendations from the National Institutes of Health (NIH) Advisory Committee to the Director (ACD) Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research, 2021

| Theme | Recommendations |
|--|--|
| Improve study design and data analysis | <ol style="list-style-type: none"> 1. NIH should improve and expand statistical training for animal researchers. <ul style="list-style-type: none"> ▪ NIH should partner with other organizations to develop modern and innovative statistics curricula relevant to animal researchers. ▪ NIH should develop statistical resources specifically for animal researchers. ▪ NIH should require statistical training for trainees conducting animal research and strongly encourage it for team members involved in study design and data analysis. 2. NIH should facilitate collaboration between statisticians and animal researchers. <ul style="list-style-type: none"> ▪ NIH should expand research collaborations between statisticians and animal researchers. ▪ NIH should fund training for statisticians on domain-specific subject matter and on challenges faced by animal researchers. ▪ NIH should increase animal researchers' access to statistical consulting through funding opportunities. ▪ NIH should incentivize research in statistical methods for animal study design and analysis. 3. NIH should add a single page to the NIH grant application research strategy section that is solely dedicated to the description of critical elements of study design, including inclusion/exclusion criteria, sample size estimation, data analysis plan, blinding, and randomization, to reduce the risk of bias and chance observations. This page would be in addition to the current research strategy page limit and would apply to vertebrate and cephalopod studies. 4. NIH should evaluate where in the pre-study research process experts could assess the quality of study design and data plans, then implement pilot studies of assessment at the most plausible stage(s). |
| Address incomplete reporting and questionable research practices | <ol style="list-style-type: none"> 5. NIH should launch a campaign to raise awareness and understanding of prospectively documenting study design and analysis plans. 6. NIH should develop and implement a pilot program to generate data on and evaluate the effects of solutions that involve the prospective documentation of study design and analysis plans in preclinical animal studies. <ul style="list-style-type: none"> ▪ NIH should develop and incentivize projects that generate data on the impact of prospective registration and registered reports. ▪ NIH should set up a dedicated program to evaluate the data generated from the projects on prospective registration and registered reports and guide future adoption of prospective registration practices in preclinical animal studies. |
| Improve selection, design, and relevance of animal models | <ol style="list-style-type: none"> 7. NIH should establish a framework for rationalizing the scientific and, when appropriate, translational (human) relevance of an animal model and its selection. This framework should be employed as part of the justification for animal uses in grant applications and included in ethical review processes and in journal reports. 8. NIH should establish or identify venues for the exchange of information related to animal model design and characterization, study design, and general best practices. 9. NIH should work to improve the design of animal models through the funding of focused research programs that enhance understanding of comparative human–animal biology. |

**Appendix II: Recommendations from NIH
Advisory Committee to the Director to Improve
Animal Research**

| Theme | Recommendations |
|--|--|
| | <ol style="list-style-type: none">10. NIH should provide adequate research support for larger and long-lived non-rodent species when justified.<ul style="list-style-type: none">▪ NIH should create policy to accommodate longer time frames and higher budgets for larger and long-lived non-rodent species.▪ NIH should continue to develop national resources to produce larger and long-lived animals.11. NIH should educate the public on the value of animal research, including the important roles of long-lived, non-rodent mammals for translation to improved human health and disease.12. NIH should charter a high-level working group on non-animal modeling systems in biomedical research to complement the activities and recommendations of this ACD working group. |
| Improve methodological documentation and results reporting | <ol style="list-style-type: none">13. NIH should expect that key supporting data reported on animal research submitted in support of grant applications will include measures of quality and uncertainty for reported estimates and an interpretation of effect sizes within the context of the field.14. NIH should expect all vertebrate and cephalopod animal research to include the ARRIVE 2.0 Essential 10 at the publication stage.³15. NIH should encourage and support work to better understand, monitor, record, and report important extrinsic factors (such as temperature and lighting) related to animal care that may impact research results.<ul style="list-style-type: none">▪ NIH should provide education about the importance of extrinsic factors to the research community, provide a method to report such factors, and incentivize pilot studies to further identify which extrinsic factors are impactful to reproducibility.▪ NIH should establish a task force to implement the cataloging of extrinsic factors as data from pilot studies are gathered.▪ NIH should dedicate funds for controlled randomized trials to test the effect of potentially high-value extrinsic factors identified from pilot studies and task force recommendations.16. NIH should provide support for documenting larger and longer-lived animals' longitudinal experimental, medical, and husbandry histories.<ul style="list-style-type: none">▪ NIH should formalize funding mechanisms to longitudinally record and manage animal-level experimental, medical, and husbandry history data for larger and longer-lived animals.▪ NIH should identify minimal animal-level experimental, medical, and husbandry history data that would be longitudinally recorded.▪ NIH should encourage the sharing of animal-level experimental, medical, and husbandry history. |

³The ARRIVE (Animal Research: Reporting of In Vivo Experiments) Essential 10 guidelines were developed by an international working group with support from the National Centre for the Replacement, Refinement & Reduction of Animals in Research. The guidelines describe 10 minimum elements of study design, procedures, and results that researchers should report in publications so that readers and reviewers can assess the reliability of the research findings.

**Appendix II: Recommendations from NIH
Advisory Committee to the Director to Improve
Animal Research**

| Theme | Recommendations |
|---|---|
| Measure the costs and effectiveness of efforts to improve rigor, transparency, reproducibility, and translatability | <p>17. NIH should externally support and internally conduct analyses on elements of rigor and transparency in grant applications and publications to examine their financial costs, opportunity costs, and impact on portfolio balance.</p> <ul style="list-style-type: none">▪ NIH should identify and collect computationally extractable information from grant proposals and reports on potentially important variables, including publication metrics, methodological rigor, funding, investigator career stage, involvement of statisticians, experimental design descriptions, and numbers and species of animals and conduct extensive analyses of these data.▪ NIH should allow applicants to include text in the budget justification section on how projected animal budgets are linked to efforts to enhance transparency, rigor, and reproducibility.▪ NIH should identify scientists who demonstrate the highest levels of transparency and rigor to help define enterprise best practices. <p>18. NIH should develop an evaluation program to assess the progress in implementing the report recommendations, their effects on NIH and the research community, and challenges that arise in implementing recommendations.</p> <p>19. NIH should develop an evaluation program to assess the progress in implementing the report recommendations, their effects on NIH and the research community, and challenges that arise in implementing recommendations.</p> |

Legend: ▪ = sub-recommendation

Source: NIH documents. | GAO-25-107140

Note: Information in this table is from National Institutes of Health, Advisory Committee to the Director, *ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research Final Report* (Bethesda, Md.: June 11, 2021).

Appendix III: Comments from the Department Health and Human Services



DEPARTMENT OF HEALTH & HUMAN SERVICES

OFFICE OF THE SECRETARY

Assistant Secretary for Legislation
Washington, DC 20548

December 5, 2024

Steve Morris
Director, Natural Resources and Environment
U.S. Government Accountability Office
441 G Street NW
Washington, DC 20548

Dear Mr. Morris:

Attached are comments on the U.S. Government Accountability Office's (GAO) report entitled, **“National Institutes of Health: Assessing Efforts to Improve Animal Research Could Lead to Greater Human Health Benefits” (GAO-25-107140).**

The Department appreciates the opportunity to review this report prior to publication.

Sincerely,

Melanie Anne Egorin

Melanie Anne Egorin, PhD
Assistant Secretary for Legislation

Attachment

**GENERAL COMMENTS FROM THE DEPARTMENT OF HEALTH & HUMAN
SERVICES ON THE GOVERNMENT ACCOUNTABILITY OFFICE'S DRAFT
REPORT - NATIONAL INSTITUTES OF HEALTH: ASSESSING EFFORTS TO
IMPROVE ANIMAL RESEARCH COULD LEAD TO GREATER HUMAN HEALTH
BENEFITS (GAO-25-107140)**

The U.S. Department of Health & Human Services appreciates the opportunity from the Government Accountability Office to review and comment on this draft report.

GAO Recommendation 1:

The Director of NIH should define short-term goals with measurable targets and timeframes related to enhancing reproducibility and translatability in animal research that the agency conducts and supports. For example, some initial goals could include targets for the number of NIH-funded publications that report certain factors that affect reproducibility and translatability, such as randomization and appropriate sample sizes.

HHS Response:

NIH concurs with GAO's finding and corresponding recommendation about developing short-term goals related to enhancing reproducibility and translatability in NIH-funded, animal research. As reported in status updates for the GAO report "Research Reliability: Federal Actions Needed to Promote Stronger Research Practices," NIH is evaluating potential methods for developing indicators of rigor and reproducibility across the broad landscape of biomedical, behavioral and social sciences research literature. Once developed, NIH could use these indicators of rigor and reproducibility to assess publications resulting from NIH-funded research for their adherence to NIH's Enhancing Reproducibility through Rigor and Transparency policy. NIH will develop appropriate measurable, targets for this type of analysis. NIH will provide a target completion date in our 180-day letter response to Congress.

GAO Recommendation 2:

The Director of NIH should collect evidence needed to assess NIH's efforts to enhance reproducibility and translatability in animal research. This could include steps such as (1) analyzing attempts to reproduce NIH-funded studies or (2) collecting information from projects that attempt to reproduce significant lines of research.

HHS Response:

NIH concurs with GAO's finding and corresponding recommendation about collecting evidence needed to assess efforts to enhance reproducibility and translatability in NIH-funded animal research. NIH will develop plans to use the indicators of rigor described in the response above to evaluate projects that attempt to reproduce NIH-funded research. NIH will provide a target completion date in our 180-day letter response to Congress.

Appendix IV: GAO Contacts and Staff Acknowledgments

GAO Contacts

Steve Morris at (202) 512-3841 or MorrisS@gao.gov or Candice N. Wright at (202) 512-6888 or WrightC@gao.gov

Staff Acknowledgments

In addition to the contacts named above, Nico Sloss (Assistant Director), Hayden Huang (Assistant Director), Christy Feehan (Analyst in Charge), Tara Congdon, Adriana Derksen, Mollie Lemon, Rob Letzler, Serena Lo, Tricia Moye, Danny Royer, and Craig Starger made key contributions to this report.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

Order by Phone

The fastest and easiest way to obtain copies of GAO documents at no cost is through our website. Each weekday afternoon, GAO posts on its [website](#) newly released reports, testimony, and correspondence. You can also [subscribe](#) to GAO's email updates to receive notification of newly posted products.

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, <https://www.gao.gov/ordering.htm>.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on [Facebook](#), [Flickr](#), [X](#), and [YouTube](#).
Subscribe to our [RSS Feeds](#) or [Email Updates](#). Listen to our [Podcasts](#).
Visit GAO on the web at <https://www.gao.gov>.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact FraudNet:

Website: <https://www.gao.gov/about/what-gao-does/fraudnet>

Automated answering system: (800) 424-5454 or (202) 512-7700

Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Sarah Kaczmarek, Managing Director, KaczmarekS@gao.gov, (202) 512-4800, U.S. Government Accountability Office, 441 G Street NW, Room 7149 Washington, DC 20548

Strategic Planning and External Liaison

Stephen J. Sanford, Managing Director, spel@gao.gov, (202) 512-4707 U.S. Government Accountability Office, 441 G Street NW, Room 7814, Washington, DC 20548



Please Print on Recycled Paper.