2024 Annual Ethics Cases

The Committee on Scientific Conduct and Ethics (CSCE) has prepared three cases for 2024 that deal with some important topics relating to research with human subjects and using artificial intelligence in research. These include:

Case 1: IRB Protocol Deviation
Case 2: Using AI to Write a Manuscript
Case 3: Using AI to Analyze Research Data

Since it may not be possible to cover all three cases in the allotted time, we suggest that facilitators cover the cases that meet the needs and interests of their audience.

CONTENTS

Please note that the CSCE recommends using a gender-neutral honorific title (Mx., pronounced “mix” or “mux”) and “they/them/their” pronouns in the cases, to try to reduce bias in the case discussions. We encourage facilitators to take advantage of negative comments from participants about this practice to promote a positive discussion about reducing genderism (bias resulting from a gender binary view) and unconscious bias in science.

IRB Protocol Deviation (Case #1)

Mx. Fox is a research nurse at the NIH who recently started working with Dr. Bear on an IRB-approved Phase III clinical trial comparing three different FDA-approved medications for treating mild-to-moderate depression. The study has exclusion criteria pertaining to various health measures, such as blood pressure, kidney and liver function, depression score (based on two metrics), and body mass index (BMI). Participants' BMI must not be greater than 30 kg/m^2. One day, Mx. Fox was reviewing the records of new patients on the study and noticed that Dr. Bear had enrolled a patient with a BMI of 31, which is a protocol deviation. Mx. Fox asked Dr. Bear about this, but they shrugged and told Mx. Fox not to worry about it because, in their professional judgment, the patient was healthy enough to participate in the study. Not wanting to cause any trouble, Mx. Fox tried to forget about the incident, but Dr. Bear did the same thing the following week. This time when Mx. Fox asked about the deviation, Dr. Bear became angry, grabbed Mx. Fox's wrist and told Mx. Fox that it is none of their business. During lunch at the cafeteria, Mx. Fox told Mx. Badger, another research nurse, what Dr. Bear had done. Mx. Badger responded: "You better get used to it. Dr. Bear does not tolerate people questioning their judgment."

[Proceed to next page]

Questions for Case #1 discussion (with facilitator notes)

1. Should Mx. Fox follow the advice of Mx. Badger to "get used" to Dr. Bear's behavior? Why/Why not?

   Definitely not. Grabbing Mx. Fox's wrist and saying its "none of their business" violates NIH's policy against inappropriate conduct (https://hr.nih.gov/working-nih/civil/harassment-and-discrimination), which includes "inappropriate touching or any form of physical intimidation or aggression." It may also constitute harassment if the action is based on Mx. Fox's gender, race or other protected classification status. https://policymanual.nih.gov/1311. Touching a person in a threatening manner without their consent may constitute an assault and should be reported to the NIH police. https://ors.od.nih.gov/ser/dp/Pages/default.aspx .
   Mx. Fox should report the matter ASAP to the NIH Civil Program, which they may do anonymously if desired. https://hr.nih.gov/working-nih/civil. Mx. Fox may consult confidentially with the NIH Ombuds Office or with the NIH EAP (Employee Assistance Program) first for advice.

   It should also be noted that if Mx. Badger is a federal employee with any supervisory responsibilities (with Mx. Fox or anyone else reporting to them), they are mandated to report this incident once they become aware of it. Likewise, anyone else with supervisory responsibilities, who hears of the incident or is a bystander, is also required to report.

   Additionally, Dr. Bear is demanding that Mx. Fox not report protocol deviations and non-compliance that could place human research participants and the NIH at risk. The IRB approved the protocol under the condition that BMI is not greater than 30 kg/m^2. Enrolling patients with a BMI greater than 30 kg/m^2 is a deviation that must be reported to the IRB. Also, since the deviation appears to be intentional, it may be considered serious or continuing non-compliance, which also must be reported to the IRB https://policymanual.nih.gov/3014-801. In other words, Mx. Fox should report this protocol deviation to the IRB even if instructed not to do so by a superior. Any retaliation against Mx. Fox for reporting these deviations would be a violation of whistleblower protections.

2. What are the potential consequences of not addressing the protocol deviation?

   The Office of Human Research Protections (OHRP), which oversees NIH human subjects research, may take some form of corrective action (such as a warning or determination letter) against the NIH for not reporting this deviation or not reporting

it in a timely fashion. Deviations from well-designed inclusion/exclusion criteria risk compromising participant safety and/or data integrity.

Note: Some IRB-approved protocols leave room for "professional judgement" concerning whether a patient is healthy enough to qualify for a study. Although professional judgment is an important skill in clinical medicine, it can lead to biases that can undermine patient care or the validity of a study. Hence, IRBs are moving toward clearly-defined inclusion criteria that leave little room for professional judgment. The protocol described in this case has not left room for professional judgment in regard to BMI, however, there is a controversy over using BMI as a health measurement. https://www.yalemedicine.org/news/why-you-shouldnt-rely-on-bmi-alone Some argue that BMI is an invalid measure or is racially discriminatory. As with any inclusion/exclusion criteria, BMI should not be included as an inclusion/exclusion criterion unless there is a sound scientific or medical justification for requiring that BMI is not greater than 30 kg/m2. If Dr. Bear designed this study and included BMI as a historical measure they had typically used to assess general health but does not believe strict adherence to a BMI criterion is medically or scientifically necessary, they should submit an amendment to the protocol and present their reasoning to the IRB. If they are one PI on a multi-site phase III trial they did not design, they should discuss their concern with the investigator responsible for the trial design and argue for an amendment to be submitted for all sites. However, because the IRB has approved the protocol with explicit BMI exclusions, Dr. Bear should follow the protocol until it is modified, if ever.

3. What resources are available to support Mx. Fox in navigating this situation?

   Some resources include the: IRBO Office, Civil Program, Institute/Center Clinical Director, Clinical Center Bioethics Office, Ombudsman, and Employee Assistance Program. Mx. Fox should go to the IRBO Office first (and inform the Clinical Director) because of concerns about study integrity and patient safety. The Clinical Director could also be a good resource because they may have a supervisory role over Dr. Bear and can work with the IRB Office on this problem. Mx. Fox could (and probably should) also report the matter to the Civil Program because of Dr. Bear's inappropriate conduct. However, if they are not ready to take this step and want confidential advice, they should talk to the Ombudsman or Employee Assistance Program.

4. Is Mx. Fox being disloyal to the research team? How do they balance staff loyalty with ethical responsibilities to study participants and the scientific community?

   Reporting this situation to the proper institutional officials is not being disloyal. In fact, reporting is being loyal to the patients in the study and to the taxpayers, who are

funding the study. It is actually DISLOYAL to ignore clear violations of policies or laws. Loyalty requires one to communicate clearly with co-workers about issues and to work toward common goals. In the interests of loyalty, Mx. Fox could talk to Mx. Badger and other team members about the situation before making a report (if Mx. Fox is comfortable doing this).

5. What steps can be taken to ensure the safety and well-being of participants enrolled in the study and ensure the validity and reliability of the data collected?

   Reporting the situation to the NIH IRBO is the most important thing to do to protect the participants and data. The IRB will then decide what should be done once it receives the report. The IRB could suspend enrollment while it performs an investigation, withdraw patients from the study who were improperly enrolled, and inform affected participants about what happened and what is being done about it. The IRB could place Dr. Bear on probation and require them to take additional training in human subjects protections. The Clinical Director would also play a key role in this response and would work with the IRB.

6. How might this case impact the trust and confidence of participants in clinical research at the NIH?

   It could have a negative impact on trust and confidence if the OHRP writes a determination letter to NIH and the story is covered by the media. This bad publicity could make some participants more hesitant to enroll in NIH studies.

Ethical Concepts Relevant to this Case

Protocol Deviation: Dr. Bear's enrollment of a participant with a BMI exceeding the specified limit constitutes a protocol deviation. Deviations from the approved protocol compromise the integrity and validity of the clinical trial.

Professional Judgment vs. Protocol Adherence: Dr. Bear's reliance on professional judgment as a justification for the protocol deviation raises concerns about the balance between individual discretion and adherence to established research guidelines. Ethical research requires strict adherence to approved protocols to ensure participant safety and data reliability.

Unwelcoming Work Environment: Dr. Bear's dismissive response to Mx. Fox's inquiry and anger create an unwelcoming work environment. A culture that discourages questioning compromises ethical oversight and may contribute to non-compliance and inferior research.

Physical Intimidation: Dr. Bear's physical aggression towards Mx. Fox by grabbing their wrist is a clear breach of the NIH anti-harassment policy and cannot be tolerated. Physical intimidation is unacceptable in any professional setting and violates the respect and safety of the staff.

Whistleblower Retaliation: Mx. Badger's response indicating that Dr. Bear does not tolerate questioning raises concerns about the potential for possible retaliation against whistleblowers. A culture that discourages reporting unethical behavior hinders the identification and correction of deviations, risking the overall quality of the research.

Collaborative Decision-Making: A lack of collaborative decision-making and open communication within the research team is obvious. Ethical research requires a collaborative environment where concerns can be addressed with transparency and dissent is welcomed.

Professionalism and Accountability: Dr. Bear's unprofessional conduct and lack of accountability for protocol deviations undermine the principles of professionalism and responsibility in research.

[End of case study #1]

Please take the survey by either clicking on the link below or scanning the QR code on your hand-held device: https://www.surveymonkey.com/r/JTDK6JN

Using AI to Write a Manuscript (Case #2)

Dr. Blue is principal investigator at the NIH who specializes in cancer genotyping. A prestigious review journal has asked Dr. Blue to write an article reviewing the current state of the field. Dr. Blue is very busy with clinical, research, and administrative responsibilities, so they ask Dr. Green, a postdoctoral fellow working in the lab, to write the review. Without telling Dr. Blue, Dr. Green uses an artificial intelligence (AI) tool to summarize the literature on this topic and generate references. Dr. Blue reads the review and congratulates Dr. Green on a job well done. They submit the solicited review to the journal. The article lists Drs. Blue and Green as authors but does not acknowledge the use of the AI in preparing the article. Two months after publication, an anonymous critique of the article, appearing in a post-publication peer review blog, claims that two of the citations in the article are fake. The editors of the review journal inform Dr. Blue about this and ask them to submit a correction. Dr. Blue meets with Dr. Green about the issue and asks how the problem occurred. Dr. Green admits to using an AI tool to help write the article and says the tool must have made the mistakes. Dr. Blue is furious at Dr. Green for using this tool without consulting with them first. They both carefully examine the references and verify that the two references mentioned by the critic are indeed fake. They also discover that three additional references are inaccurate, three are irrelevant, and two sentences in the article are copied word-for-word from another article without quotation marks or attribution.

[Proceed to next page]

Questions for Case #2 discussion (with facilitator notes)

1. When Dr. Blue and Dr. Green submit their correction to the journal, should they also address the inaccurate and irrelevant references and the copied sentences and acknowledge the use of the AI tool?

   Yes, they should be completely honest and open with the journal about what happened. They want to avoid the embarrassment and possible repercussions, but they made mistakes and need to take responsibility for them. They should correct any inaccurate references and use their own scientific judgment to replace any references they consider to be irrelevant.

2. Should they explain how the problem occurred, i.e., that the AI tool made the mistakes?

   Yes, to the best of their ability. Again, honesty, transparency, and accountability demand this. They should communicate this to the editors and include it in their correction (or retraction) notice.

3. Should they retract the article?

   It is not clear whether they should retract the article. Committee of Publication Ethics guidelines https://publicationethics.org/retraction-guidelines recommend that editors can retract for fabrication, falsification, or plagiarism. The article includes some fake references, but is providing fake references the same as faking actual study data? The article includes a couple of plagiarized sentences—but is that enough to warrant a retraction? The authors should definitely submit a correction first and have a conversation with the editors about what would be most appropriate. The editors might insist on a retraction to send a message to other authors that they will not tolerate irresponsible AI use, but they could also allow the authors to make the text corrections.

4. Did they commit research misconduct, i.e., plagiarism?

   Plagiarism is defined as "appropriation of another person's ideas, processes, results, or words without giving appropriate credit (45 CFR 93.103)." For the NIH to make a finding of plagiarism against a respondent, the NIH must show by a preponderance of evidence that they acted intentionally, knowingly, or recklessly and significantly deviated from standard research practices. In this case, Drs. Blue and Green did not act intentionally or knowingly, but they may have acted recklessly if it was reckless to use the AI to summarize the literature without reviewing and verifying its results. However, it is not at all clear that their conduct was reckless. Certainly, it was negligent, but not so much so that it should be regarded as reckless.

5. What are the responsibilities of authors when using AI tools to review the literature?

   They are manifold and important:

   - **Carefully plan and refine the search and the prompts;**
   - **Review the results;**
   - **Check references for accuracy and relevance;**
   - **Check the text for plagiarism;**

- **Make sure that important facts or references are not omitted;**
- **In the article, disclose and describe use of the AI tool but don't name the tool as an author.**

For more, see:

- **NIH Guidelines for the Conduct of Research, https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-conduct_research.pdf**
- **ICMJE guidelines, https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html**

6. Did Dr. Blue meet NIH authorship criteria for this review article?

   Probably not. Dr. Blue assigned it to Dr. Green, who used an AI tool to write the article. It seems like Dr. Blue's contribution is minimal and to qualify as an author, they should have had a more active role, such as giving Dr. Green more explicit instructions on how to write the review, sharing important references with Dr. Green, or more thoroughly reviewing/revising the manuscript.

[End of case study #2]

Please take the survey by either clicking on the link below or scanning the QR code on your hand-held device: https://www.surveymonkey.com/r/JTDK6JN

Using AI to Analyze Data (Case #3)

Dr. Falcon, a postdoc in Dr. Hawk's research group, has struggled to analyze health survey and genomic data from a longitudinal NIH intramural research study with 10,000 human participants. Dr. Falcon wonders if they might be able to use artificial intelligence (AI) tools to help analyze the data. Dr. Falcon has an account for an NIH ChatGPT platform, but this version of ChatGPT does not have the functionality needed for this data analysis, so they sign up for a personal account with a commercial AI platform, HotBot1, which is able to analyze data from publicly accessible health databases that is similar to the IRP study data. Dr. Falcon uses HotBot1 to search for statistical relationships among dozens of variables from the public databases; however, Dr. Falcon soon realizes that to make significant progress, they need to supplement the publicly available data with additional, more detailed data. Fortunately, the IRP study includes the data needed to improve the analysis and HotBot1 allows users to upload data to the platform.

Dr. Falcon de-identifies the intramural study data so it includes no names or personal identifiers and uploads them to HotBot1. After several weeks of work, Dr. Falcon has some promising results, including a genetic association that could have important public health implications. Although the analysis appears to misrepresent findings for an underrepresented minority cohort of the data, Dr. Falcon is confident that the rest of the analysis is completely reliable. Dr. Falcon shares the results of this work with Dr. Hawk at their next regularly scheduled meeting and tells Dr. Hawk how HotBot1 was used to analyze both the public and intramural datasets together. While Dr. Hawk is not very familiar with AI tools, Dr. Hawk is excited about the new findings. They quickly draft a manuscript reporting the results of their data analysis and submit it for publication clearance review in their IC.

[Proceed to next page]

Questions for Case #3 discussion (with facilitator notes)

1. Has Dr. Falcon done anything wrong? If so, what actions should be taken to mitigate any mistakes?

   Yes, Dr. Falcon has uploaded IRP study data to a commercial server (HotBot1) in violation of NIH policies. Although Dr. Falcon removed personal identifiers to protect privacy, it cannot be determined from the facts given whether this step was sufficient to protect the privacy of the study participants, since the AI may be able to identify participants from the data provided or by linking it to other data, particularly genetic markers. Also, the security of this commercial server is unknown, so it is possible that other parties could hack into the server and gain access to private data. Additionally, once information is shared within AI, there is no control over how and where the data will be used. This could potentially violate study protocols and the informed consent researchers have received from the study participants. Even if the privacy of human participants has not been breached, there are still concerns about inadvertently sharing NIH data prematurely if outside parties are able to gain access to the data. Although NIH policies require sharing of data upon publication or completion of a project (with appropriate patient consent), they do not require data to be shared before such time because the data may need to be edited, cleaned, audited, reviewed, or validated. Also, the investigator's career/publication interests could be harmed (e.g., they could be scooped) if outside parties can get access to their data prior to publication.

   To mitigate these mistakes, this problem should be immediately reported to the NIH IT Service desk https://oma.od.nih.gov/DMS/Pages/Privacy-Program-Privacy-Incidents-and-Breach-Response.aspx , the NIH Institutional Review Board (IRB), and the AI company. It may be possible to withdraw or delete/scrub the data from the AI server before too much damage occurs.

2. Were the steps that Dr. Falcon took to protect NIH data sufficient? Has Dr. Falcon committed a data breach incident that should be reported?

   As discussed above, the steps were not sufficient to protect participant privacy. Yes, Dr. Falcon has committed a data breach that must be reported. NIH Personally Identifiable Information (PII) data must be securely stored on NIH devices/servers or on servers approved by NIH IT/security.

3. How can scientists balance the need to develop their research program quickly with their lack of formal education in emerging technologies?

   They can obtain more education by participating in workshops and training sessions offered at the NIH, by attending conferences and seminars, and by engaging in self-study. They can also collaborate with researchers who have expertise in emerging technologies. While it is

important to take advantage of new technologies, this should be done responsibly, with adequate knowledge and expertise.

4. How could HotBot1 have made an error in analyzing the underrepresented minority cohort of the population? What are the implications of using the entire dataset despite the concerns? How could this problem have been anticipated or prevented?

   Without more information, it is difficult to say how this happened, but the error could be due to biases in HotBott1's training data, biases in the data that were uploaded to HotBot1, or biases in the way that HotBot1's algorithms analyzed the data (e.g., looking for larger, more dominant patterns rather than patterns in any subpopulations). Using a dataset despite these concerns is problematic because it could lead to biased research and discrimination against the underrepresented minority group if the data are used, for example, to make health care decisions. This error might have been anticipated or prevented if the researchers had more carefully examined the training data, the uploaded data, the algorithms, and potential uses/applications of the data, which might have enabled them to identify potential biases. Researchers with the relevant expertise (such as in AI, biostatistics, genomics, and ethics) and representatives of the underrepresented minority group should be included in these discussions to make sure that scientific, technical, and social/ethical issues are properly addressed.

   Note: there is evidence that genomics databases have a European bias, which is an important issue for this case: https://www.healthaffairs.org/doi/10.1377/hlthaff.2017.1595.

5. In your opinion, is Dr. Hawk appropriately overseeing the research of Dr. Falcon? Should Dr. Hawk have been informed by Dr. Falcon that they were embarking on this exploratory path? Should Dr. Hawk delve more deeply into the work that Dr. Falcon did using HotBot1, or is it acceptable for Dr. Hawk to trust Dr. Falcon without independently verifying any of the analyses?

   It seems likely, based on the facts given in this case, that Dr. Hawk is not appropriately overseeing Dr. Falcon's work and training. Although Dr. Hawk should be able to trust Dr. Falcon to work independently, Dr. Hawk should be informed about what Dr. Falcon is doing and planning to do, especially when it comes to using a new technology like HotBot1. Dr. Hawk should have been having regular communication with Dr. Falcon about this project and reviewing the data and results.

6. Is it ever acceptable to use personal credentials instead of official credentials to set up an account using an NIH computer to analyze data? If so, under what circumstances?

   No, it is unacceptable. According to NIH policy, "Use of personal or non-NIH email systems or accounts for business purposes is prohibited."
   https://policymanual.nih.gov/chapter/export/1743-1/1?modelId=27af840b#:~:text=Use%20the%20NIH%20email%20system,for%20business%20purposes%20is%20prohibited.

7. More generally, what types of AI tools are permissible to use in your research?

   Large language models like ChatGPT; others?

   Additional Resources

   Conduct of Research – chapter that talks about AI use:

   - https://wiki.ocio.nih.gov/wiki/index.php/NIH_Artificial_Intelligence_(AI)_Cybersecurity_Guidance
   - https://wiki.nci.nih.gov/download/attachments/384206504/NIH%20IT%20General%20Rules%20of%20Behavior%20v2.0.pdf?version=1&modificationDate=1561042664000&api=v2
   - NIH Artificial Intelligence (AI) Cybersecurity Guidance - NIH InfoSec Wiki

   NIH OCIO Guidance

   External/public generative AI tools must only be utilized for public data (information already in the public domain) and cannot be used for Personally Identifiable Information (PII) or consented clinical research data (except as noted below). Therefore, NIH requires the following mitigations for the use of generative AI tools:

   1. Do not share PII: NIH personnel must never share PII (sensitive and non-sensitive), such as Social Security numbers, driver's license numbers, credit card information, or medical information through generative AI. This information must be relevant and necessary to accomplish an authorized purpose, transmitted only through secure channels to authorized personnel on a need-to-know basis, and for official purposes only. Information placed into these tools must be kept as generic as possible. For example, use synthetic data and/or de-identified data. Take action to sanitize PII and protect sensitive NIH data.

   2. Do not share consented clinical research data or controlled access data: Once information is shared within AI, there is no control over how and where the data will be used. This could potentially violate study protocols and the informed consent researchers have received from clinical research participants. One exception is if consent has been obtained from participants and stipulations from this guidance are adhered to.

   3. Do not share sensitive and nonpublic information: NIH personnel must never ask questions, type prompts, or upload material to a generative AI containing sensitive information such as health-related information, financial details, confidential topics, vendor proprietary information, grant sensitive data, evaluations, draft policy, or other proprietary and/or nonpublic data. This includes prohibiting NIH scientific peer reviewers from using generative AI technologies for analyzing and formulating peer review critiques for grant applications and research and development (R&D) contract proposals. Inputting sensitive information into a generative AI can lead to data being disclosed to others. Do not input material into a public generative AI that displays the pre-decision intentions of the U.S. Government. See the HHS policy for Securing Artificial Intelligence (AI) Technology for additional information.

4. Follow Department policies: Employees must always follow the Department of Health and Human Services (HHS) and NIH policies regarding PII and data protection. This includes policies related to data storage, transmission, and sharing. See the HHS policy for Securing Artificial Intelligence (AI) Technology.

5. Do not base decision-making or policymaking solely on data from AI services: When making important decisions, gathering reliable and accurate information from trustworthy sources is essential. Although generative AI tools can be helpful, they must only be used to supplement your research. It is not a replacement for thorough and independent research on a topic and can be subject to inherent bias and deliberately false information. It is not designed to substitute for internal governmental deliberation on topics. By consulting with experts, evaluating the quality of the information, and considering multiple perspectives, you can make more informed decisions and develop better policies.

6. Do not rely on the technology to be a software developer by proxy: All well-written code must adhere to security design and ethical principles. All code output needs to be reviewed for completeness, quality, efficiency, and, most of all, security. Leverage manual and automated validation tools and testing technologies to help ensure these factors. If you cannot identify or understand what a piece of AI generated code does, you should not use it.

7. Limit the sharing of research information to only post publication content: NIH personnel must not share or use the technology to help develop or review prepublication information that could be misused or cause the NIH, or any research subjects, any real or perceived harm. Remember, generative AI tools only learn by consuming more data. Assume that anything you share with AI chatbots could be misappropriated or compromised. Assume that others worldwide have likely shared incorrect data with AI chatbots, both unintentionally and intentionally.

[End of case study #3]

Please take the survey by either clicking on the link below or scanning the QR code on your hand-held device: https://www.surveymonkey.com/r/JTDK6JN