AI FAQ

Respondents:

Tamiko Eto *(she/her/hers)*
Director of Research Operations - Human Research Protection Program Mayo Clinic

Mark Lifson *(he/him/his)*
Director - AI/ML Engineering
Center for Digital Health-AI Enablement

Nov 6, 2024

How this FAQ is set up

- This PowerPoint is set up in sections regarding FAQ's submitted.
  - HSR vs NHSR
  - FDA
  - Applications
  - Belmont Report
  - Applicability of the AI model to other regulatory criteria
  - Other discussion topics
  - Alternative Resources
  - References

Please note the following:

- The answers to the questions are from the speakers: Tamiko Eto, Mark Lifson, and the Clubhouse team. They do not represent any specific institution. They represent our experiences and expertise.
- This PowerPoint was set up with the questions asked and then the responses provided.
- We tried to make the information as accessible as possible for future use.
- Each question and answer is labeled and numbered with each section.
    - We welcome discussion and might not make it through all of the questions.

HSR vs NHSR Questions

HSR vs NHSR Question 1

- Most algorithm development at my institution occurs in a deidentified dataset, therefore, does not require IRB review.
  - How can IRB/regulatory experts influence this work?

HSR vs NHSR Answer 1

- What constitutes deidentification? (Common Rule vs HIPAA)
- If diagnostics, predictive models, and risk scores are involved, HIPAA is likely involved.
  - It is difficult to conduct AI research without identifiers (dates) so make sure deidentification meets HIPAA deidentification criteria (specific criteria/process for deidentification should be spelled out).
- NOTE: identifiability only matters under 45 CFR 46.
  - 21 CFR considers both identifiable and de-identified data as a human subject.

HSR vs NHSR Question 2

- How far out do you see that IRBs are from federal regulations specific to AI HSR?

HSR vs NHSR Answer 2

- Depends on the administration.
- **1 year** (~ish?) for establishing guidance.
  - **FDA:** November 2024 meeting → Public Comment Period → x-months to finalize
  - **OHRP:** 2023 Task Force
  - **FTC:** Standing laws. Example: using AI tools to trick, mislead, or defraud people is illegal
- **~2 years** after regulations are passed = enforcement.
- **REMEMBER:** Current mandatory laws ***do exist*** in several states and at the federal level (FDA). Keep an eye out for them– not just if your state is impacted. If you are collaborating with states that require it (e.g., single IRB, pooling data, etc.)

FDA-Related Questions

FDA Related Question 1

- Does it matter if its a free standing AI/ML algorithm that the researchers is trying to feed human data to predict risk scores to meet the definition of a medical device? Or to apply the FDA rules, does it have to be part of a software?

Answer FDA Related Question 1

- AI/ML is a software function. Software functions are products. Products are regulated.
- Software functions that meet CERTAIN criteria are considered “devices” under the FDA
  - *A software function (e.g., AI algorithm) that is designed for use in diagnosis, cure, mitigation, treatment or prevention of a disease is considered a medical device, regardless of what platform it’s being run on.*
- To be considered a “non-device” confirm the intended purpose/label

FDA Related Question 2

- Is IRB review required for AI testing of data that is not from subjects? Researchers input made up data that is similar to clinical data. Would FDA regulations apply to this scenario?

FDA Related Question 2

- It depends. FDA regs would apply if the software function meets the definition of a medical device.
- The FDA is responsible for protecting the public from harmful products.
- The FDA does not distinguish between identifiable data and deidentified data.
- Regardless, Synthetic data is not entirely “fake data” but based off human data (replicates the statistical properties of real-world data). Reverse engineering an issue?

FDA Related Question 3

- Does Mayo make device risk determinations for studies in the phase 1 algorithm development stage?

FDA Related Answer 3

- Likely “no”.
- If the application clearly defines the product as having a defined software function (has an indication or intended use) then it would likely not be considered "Phase 1".
- We would move to Phase 2 and make a device determination and then a risk determination.
  - It may be a CDS non-device, so the device determination comes first. Phase 1 should not have a well-defined software function/indication/intended use, so we would not make a device risk determination. Unless they combined a Phase 1 and Phase 2 study, then ***likely we would***.

FDA Related Question 4

- In the method you've laid out, is it correct that the research is considered FDA regulated but not subject to 812 until Phase 3?

FDA Related Answer 4

- No. 21 CFR 812 would apply from **Phase 2**, if medical device determinations are made.
  - NOTE: 21 CFR would apply from Phase 1 onward (GMLP, GCP, etc).

- **Phase 2** is when the software function is defined and has an “intended use” as defined by the FDA (“the general purpose of a product” as intended by the manufacturer)
  - “Manufacturer” = the entity/individual responsible for labeling the product

FDA Question 5

- Can you give an example of a function that is not designed to serve a medical purpose?

FDA Answer 5

- When you say "medical purpose" are you specifically trying to identify when something constitutes a "medical device"?
  - Some AI that is designed to serve a medical purpose may not be a "device" if it meets certain criteria (CDS non-device). See examples here: https://www.fda.gov/media/109618/download

- If the question is to identify AI software solutions that are not intended for use in the healthcare industry, there are tons. Facial recognition to open our phone, scheduling tools, Deep Learning and Reinforcement Learning to predict and recommend movies and content for music or movies/TV, etc.

- If we are looking more at AI technology in healthcare that has not typically been classified as a medical device, foundation models are one type. Clinical notes summary (without complex GenAI features) could be another.

FDA Question 6

- I heard the FDA wants to discontinue continuing review for NSR device determinations, how do you think this will impact the IRB's ability to monitor risks for a device that is consistently changing?

FDA Answer 6

- Did they say they "want to"? Or was that a public comment or general ambiguous statement made under a Q&A at one of their webinars?

- The need for CR is left up to the institution for Common Rule studies, so the same framework would apply under 21 CFR, if the FDA pursued such a determination.

FDA Question 7

- Who at FDA do you suggest PI's contact as a first step to see if their software idea is subject to FDA regs?

FDA Answer 7

- Requests for Feedback and Meetings for Medical Device Submissions: The Q-Submission Program Click here
- (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program).
- Email: DICE@fda.hhs.gov and DigitalHealth@fda.hhs.gov.
- Digital Health Frequently Asked Questions Click here
- (https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-frequently-asked-questions-faqs) and here (https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance).
- Digital Health Policy Navigator Click here https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-policy-navigator and Policy for Device Software Functions and Mobile Medical Applications Click here https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-device-software-functions-and-mobile-medical-applications

FDA Quesiton 8

- Should IRB review submissions as a device even though they aren’t a device?
- Isn’t that overreach?
- Also, there is plenty of risks introduced when using off the shelf AI systems but the presentation didn’t assess how sites can make the system fit for purpose.

FDA Answer 8

- Is that a rhetorical question? I do not believe we recommended reviewing all submissions as a device. If they are not a device, you wouldn't review them as such. Though, the same review framework works, and necessary risk controls apply.

- The presentation discussed the **development phase of software solutions** such as digital health tools that involve AI/ML, and when FDA regulations and standards are expected to be incorporated (as outlined by the FDA).

- As outlined in the presentation, if the project is in discovery (clinical association), we would not consider it a device if there lacks a clear indication or intended use. However, once the algorithm is developed and **clinical association is established** (there IS an intended use or indication). Then we would make a device determination (CDS non-device? device? non-medical device?).
    - If classified as a device, a risk determination would be made, and that determines what FDA regulations may be applicable.
- Agreed that off-the-shelf AI systems present the same issues, and each site will have to validate (with continuous monitoring) at the local level. Use it per intended use. NO AI system should be used outside of its intended use.

FDA Question 9

- In the event a researcher themselves are the source of the data used to code for the algorithm, how should the IRB assess the device?

FDA Answer 9

- If the researcher wants to use their own PII/PHI to train the AI system I don't think a model trained on one single individual will provide very meaningful results.

- If the question is asking whether the researcher is both the source of data and product developer, then COI management plans need to be in place (not letting them validate their own product).

Application Questions

Application Question 1

- What are your thoughts on the role of the scientific reviewer for these types of studies?

Application Question 1 Answer

- In general, I believe it is absolutely necessary to have proper expertise at the table (IRB membership or AI Ancillary Committee.

  "an IRB may, in its discretion, invite individuals with competence in special areas to assist in the review of issues that require expertise beyond or in addition to that available on the IRB. These individuals may not vote with the IRB."

- The IRB is charged with including persons knowledgeable in these areas, as required by federal regulations (45 CFR 46.107):

  "The Institutional Review Board (IRB) be sufficiently qualified through the experience and expertise of its members (professional competence) … The IRB shall be able to ascertain the acceptability of proposed research in terms of institutional commitments (including policies and resources) and regulations, applicable law, and standards of professional conduct and practice. The IRB shall therefore include persons knowledgeable in these areas."

Application Question 1 Answer Continued…

- **IRB Independence**

  These same regulations require that the IRB act independently of the institution:

  “The institutional authority under which the IRB is established and empowered, and the independence afforded the IRB to carry out its duties”

  …

  “No IRB may have a member participate in the IRB's initial or continuing review of any project in which the member has a conflicting interest, except to provide information requested by the IRB.”

Application Question 1 Answer Continued…Committee Considerations

- To address this need for expertise, some institutions add the necessary expertise as voting members. Other institutions establish ancillary committees.

- **Both approaches have benefits and disadvantages:**
  - ***Voting members*** add to the total number of members, thereby increasing the required attendance to meet quorum. It also results in limitations in perspective.
    - You can avoid this by positioning these individuals as “Scientific IRB Alternates”
  - ***Ancillary Committees*** broaden perspectives, and help ensure expert representation, but they can also lead to delays from IRB submission to approval because it requires the Ancillary Committee to convene and vote separately from the IRB
    - You can avoid this by having their review done in parallel with IRB review.

- My recommendation is to have appropriate experts assigned as alternate members (and/or committee members) to serve in this capacity. The IRB needs additional support and expertise, per the regulations for the complex scientific and technological review of artificial intelligence (AI) and machine learning (ML) studies.

Ethical Considerations
(Belmont Report, Declaration of Helsinki)

Belmont Consideration Questions

- As it relates to the Belmont Report, what are your thoughts on developing a device that is not inclusive of all populations at the moment. For example, racial bias.

Belmont Consideration Answer…continued 1

- AI models are inherently bias. Most often, that is not an intentional “feature” but rather a limitation.
- It's important to understand what "bias" means.
- One reason that AI systems produce biased or inaccurate responses is poor data quality. Many of these systems are trained on large datasets which may be full of misinformation, previously historic racist or discriminatory practices (CKD in black populations, and stroke prediction in women, for example), among other biases.

Belmont Consideration Answer…continued 2

Medical and health data are not free from bias, whether by race, ethnicity, ancestry, sex, gender identity or age. These systems are trained on health data and as a result often **encode such biases.** For example:

- Most data are collected in high-income settings and from English and Western English speakers.
- Genetic data tend to be collected disproportionately on people of European descent.
- These systems are often trained on electronic health records, which are full of errors and inaccurate information or rely on information obtained from physical examinations that may be inaccurate, thus affecting the output.
- Problems of data quality and bias affect all AI models in all sectors of society.

Belmont Consideration Answer… continued.. 3

- Another form of bias comes from automation bias.
  - E.g., when a clinician may overlook errors that should have been spotted by a human.

- HCPs might use these AI systems in making decisions for which there are competing ethical or moral considerations.
  - LMMs such as Chat GPT may be very inconsistent as moral advisers, although, as recent experiments indicated, they can influence users' moral judgement, even if users know that they are being advised by a chatbot.
  - Use of LMMs for moral judgments could lead to "moral de-skilling", as physicians become unable to make difficult judgements or decisions.

Belmont Consideration Answer continued.. 4

- To address specific "racial bias" (the question here) this can be mitigated by:
  - Conducting fairness auditing
  - Including a plan to check for differential performance between subgroups
  - Determining whether there is detectable bias, no bias outside of acceptable limits, or insufficient data to tell either way.
  - Unbiased algorithm should give patients with same needs the same calculation score or classification (but we still need to understand if that is appropriate).

- At the discovery phase (**Phase 1**), achieving AI fairness typically requires addressing and ***mitigating biases in data***, ***algorithms***, and the ***broader system design***.

- At the translation phase (**Phase 3**), the developed model should be evaluated as part of the study aims, to ensure AI fairness.

Belmont consideration continued answer 5…

What could go wrong if these issues are NOT addressed?

- Social discrimination
- Health inequality
- Legal consequences
- Loss of trust and credibility
- and more…

Applicability of the AI System/Model

Applicability of the model … Question 1

- I wonder if there are guidance from the Dept of Education that can be applied to studies of AI in the classroom?

Applicability of the model … Answer 1

- The framework (3-phased approach) is applicable regardless of FDA status.
- GMLP guidance
- 34971 Risk Assessment for ML models

Applicability of the model question 2

- Does the pre-development phase satisfy the requirement of a systematic investigation?

Applicability of the model… question 3

- Generalized “application” may not equal generalizable “knowledge.” Just a thought. Could be quality assurance or an ‘it depends’ situation.

Applicability of the model answer 3

- Please see SACHRP Guidance

  "It has been argued that AI validation activities (e.g., collecting data explicitly and only to train or subsequently validate an algorithm) is an activity that is not "designed to develop or contribute to generalizable knowledge. This ***argument is sound if the intended use of that algorithm is limited to its application to the original dataset****, but if the intent is to build a tool to be applied to a broader community or to data not-yet-collected, the situation is directly akin to the development of a diagnostic tool (in the broadest sense). It is the nature of research that diagnostic tools must be developed with data from a subset of the full population, hence the requirement for research participation. In this sense,* ***development of an AI/ML tool is not different from the development of an in vitro diagnostic device*** *and SACHRP takes the position that it should have the same degree of regulatory oversight."*

  (https://www.hhs.gov/ohrp/sachrp-committee/recommendations/irb-considerations-use-artificial-intelligence-human-subjects-research/index.html)

- "Generalized application" could be a sound argument if the intention of the QI is not to evaluate safety, effectiveness, performance, evaluation, testing, etc. The tool would already have to have established this.

Applicability of the model answer 3 continued…

Per OHRP (https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/quality-improvement-activities/index.html):

- Quality Improvement projects use **solutions that are either standard of care or evidence based** and therefore, the model's efficacy is NOT being questioned or tested.
- Quality Improvement activities are limited to projects that only involve improving clinical WORKFLOWS and **not examining how something will impact health outcomes**.
- The usefulness of the Quality Improvement project is limited to **something that would be immediately used to improve your own clinic** and not the field or your colleagues clinic.
- Models used in Quality project **can be developed by a licensed practitioner but can only be used in their immediate practice** and again, not by the bigger hospital or healthcare setting, nor by your colleagues in their practice.

- **Research**, on the other hand, compares one tool or method against another to test for impact on health outcomes.
- **Research** is needed to determine the safety or effectiveness of a device or product.
- **Research**, by its nature, develops, evaluates or validates a process that you hope can be used in other settings.
- **Research**, by nature, hopes to prove or answer a research question.
- **Research** typically (though not always) involves randomization or having control groups or something to compare one thing to another.

Applicability of the model answer 3 continued…

- Models developed with the hopes of making it generally available to other physicians would be considered “models under investigational research” and require FDA regulations as well as will trigger sponsor-investigator requirements.

- Projects that do NOT qualify as QA/QI are projects that are partially QA/QI but ALSO have Human Subjects Research components.

- If that’s the case, the entire project falls under Human Subject Research Regulations since we aren’t allowed to break it up into separate projects. They are not usually funded by outside organizations, and usually do not involve other sites. If you have an externally funded and/OR multi-site QA application, it would be prudent to understand why these external entities are invested in this project that shouldn’t necessarily apply to or benefit them. As a reminder: TERMINOLOGY MATTERS, and as such these QA projects cannot be called research at any point, So, If any of these items appear in your review, you’re likely dealing with Human Subjects Research and therefore IRB review is needed.

Applicability of the model Question 4

- Were all those risks as a result of the AI chart review conducted in a “live” environment? So if it wasn’t in a “live” environment, the risk would be only privacy and confidentiality?

Applicability of the model Answer 4

- For Phase 1, Yes… but…
- If not developed well, the largest AI risks are introduced (and directly impact all future risks), in **Phase 1** (discovery phase development of the algorithm). By the time it reaches deployment (**Phase 3**), these risks should have all been mitigated.
  - Many think breach is the only risk in AI studies. *Breach of PII/PHI use is not the only risk issue we should focus on* (assuming we have proper consent/authorization for such use and sharing).

- PHI can (and oftentimes ***must***) be used in these early discovery projects (Phase 1).
  - With proper guardrails, I would not escalate “risk” just because PHI is used or disclosed.
  - Med to high-risk applications are **higher risk because of their tendency to "drive" clinical decision making** as opposed to just supporting.
    - In discovery phase, it is too early to determine whether the project can or will drive or inform because its functionality is still not well-defined. Therefore, focus on clinical association.

- This risk consideration will come into play in **Phase 2** (Validation).
  - You say, "*early-stage feasibility evaluation*". Let’s not confuse that with Phase 1. Early-stage feasibility evaluation is Phase 2 (validation/pilot) which would be where those risks are.

- The AI IRB application should ask the study team to clearly articulate a decommission plan so that it cannot be used in clinical environments without going through proper regulatory pathways (FDA).

Possible Question to Discuss

Possible Question to discuss

Audience Member Comment from October Webinar:

“I would also note even under the common rule if the deidentified images have been collected specifically for the validation purposes you might satisfy the definition on human subject.”

Thoughts? How do other feel about this?

Alternative Resources

Alternative Resources

- https://research.uci.edu/human-research-protections/assessing-risks-and-benefits/privacy-and-confidentiality/artificial-intelligence-and-human-subject-research/

- https://techinhsr.com/wp-content/uploads/2022/08/AI-HSR-WHITE-PAPER-TechInHSR-08.2022-1.pdf

REFERENCES

https://www.hhs.gov/ohrp/sachrp-committee/recommendations/tab-c-the-protection-of-non-subjects-from-research-harm.html

https://www.congress.gov/118/meeting/house/116793/documents/HHRG-118-FD00-20240206-SD003.pdf

https://www.agencyiq.com/blog/digital-health-update-hhs-new-ai-office-fdas-report-on-ai-in-surveillance-and-coming-policy/

https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and

https://www.hhs.gov/sites/default/files/public-benefits-and-ai.pdf

https://www.hhs.gov/about/news/2024/07/25/hhs-reorganizes-technology-cybersecurity-data-artificial-intelligence-strategy-policy-functions.html

https://www.fda.gov/medical-devices/digital-health-center-excellence/fda-digital-health-advisory-committee

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation

https://www.fda.gov/medical-devices/digital-health-center-excellence/fda-digital-health-advisory-committee

https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=50

https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=56

https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-809/subpart-B/section-809.10

https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=820.30

https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.104

https://www.fda.gov/media/75414/download

https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-812

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(2)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(6)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(1)(i)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(3)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(5)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46/subpart-A/section-46.116

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(a)(7)

https://www.ecfr.gov/on/2018-07-19/title-45/subtitle-A/subchapter-A/part-46#p-46.111(b)

https://www.academia.edu/78265967/AI_HSR_Checklist_SHORT_w_Decision_Trees_Version_20220422l

https://education.primr.org/courses/43595/video_presentations/234866

https://about.citiprogram.org/course/artificial-intelligence-ai-and-human-subject-protections/

https://webstore.ansi.org/standards/bsi/bsaami349712023?srsltid=AfmBOookdFiyq4RltAUq7vgHbuJcQx5ZVrY10DUcrp9HSz9n6K0Wrb8V