E&HR

When Is a Change Significant?
The Update Problem of Apps in Medical and Behavioral Research

Carmel Shachar, Sara Gerke, Walker Morrell, Aaron Kirby, I. Glenn Cohen, and Barbara E. Bierer

ABSTRACT Digital applications (apps) are commonly used across the research ecosystem. While apps are frequently updated in the course of clinical and behavioral research, there is limited guidance as to when an app update should trigger action related to human research participant protections and who should be responsible for monitoring and reviewing these updates. We term this the "update problem" and argue that, while it is the principal investigator's duty to track all relevant updates, the level of involvement and re-review by the institutional review board (IRB) of an approved research protocol should vary depending on whether the update may be classified as minor, not minor, or significant. Minor updates require at most annual notification of the IRB, updates that are not minor require prompt notification of the IRB, and significant updates may require full board re-review or another response. We also suggest how these policies might be implemented.

KEYWORDS human research ethics, digital applications, digital updates, digital research, researcher responsibilities, institutional review boards, IRBs

Shachar, C., S. Gerke, W. Morrell, A. Kirby, I. G. Cohen, and B. E. Bierer, "When Is a Change Significant? The Update Problem of Apps in Medical and Behavioral Research," *Ethics & Human Research* 44, no. 3 (2022): 2-11. DOI: 10.1002/eahr.500118

The use of digital applications (apps) has exploded over the last decade, and clinical and behavioral research is no exception to this trend. By "apps," for the purposes of this article, we mean all free-standing software used to engage in research, with the exception of operating systems themselves. Apps enter the research ecosystem in many ways: Apps may be developed within the research institution or be adapted from noninstitutional sources. Apps may be used as the focus of a research study or used by the research team to facilitate participant recruitment or manage data collection and security. Some of these apps also meet the definition of a medical device under section 201(h)(1) of the Federal Food, Drug, and Cosmetic Act and thus are regulated by the U.S. Food and Drug Administration (FDA).

As members of the research community consider using apps in their research studies, they face a persistent challenge that we call the "update problem." Unlike more analog tools, apps can be and are frequently updated during the course of their use. Some apps have clearly delineated updates—when a revision is released, for instance—whereas some programs are based on forms of artificial intelligence (AI) and machine learning (ML) that involve continuous learning and changing. Currently, there is limited guidance as to when an app update should trigger action related to human participants' protection and who should be responsible for monitoring and reviewing these updates, whether they manifest in clearly delineated or continuously changing updates. Should tracking and reviewing updates be viewed as akin to other data collection tasks and therefore be the responsibility of the principal investigator (PI) or the sponsor of the study? Do institutional review boards (IRBs) bear some of the responsibility to monitor and review such updates, and do they have the expertise

and resources to do so? What are the roles of other institutional officials, administrative departments or units such as information technology groups, the app developer, and other stakeholders? When outside vendors are involved, how should PIs and IRBs remain apprised and evaluate updates? How should the research community approach updates of app-adjacent programs and systems such as the iOS and Android operating systems?

This article will explore some of these challenges. We will first discuss the landscape of apps commonly used for research. We will then explain the update problem. The heart of our article will focus on providing guidance for when PIs and IRBs should take action in light of an update to an app in an approved research protocol. Lastly, we will discuss IRBs' responsibilities and capacities to evaluate constant changes to adaptive algorithms.

THE APP LANDSCAPE

In one sense, apps, as we use the term, are not a new phenomenon. Early digital measurement products, such as continuous glucose monitors, have been used in research for decades.[1] What is novel, though, is the extent to which researchers are currently using these technologies as part of their research studies. This explosion is unsurprising considering the utility these apps present to researchers as well as the expansion of digital connectivity, familiarity, and user experience over the last several decades. The use of apps allows researchers to collect more data with less disruption to participants in the study. For example, having study participants use wearable sensors can allow researchers to passively and, at times, continuously collect data while the participants are at home or their workplaces or are going about their daily routines, without requiring them to travel to a study location. The use of apps can also allow researchers to work with a greater number of participants with more geographic diversity, populations they might not otherwise reach. For example, a 2019 study used a smartphone app to monitor 419,297 patients for a median monitoring time of 117 days for atrial fibrillation (A-fib),[2] generating an immense amount of data for a large group of participants that would have been significantly more difficult to compile without use of digital technology.

Overall, the use of apps in research has many benefits in allowing research teams to gather more and different types of data. However, it also raises several challenges, one of which is the focus of this paper: the update problem.

THE UPDATE PROBLEM

Apps are different from many other technologies and devices used in research because of the frequency and ease of updates, as well as the challenge of knowing when the updates occur and to what extent they are material to human research protections. Some app updates are essential to ensure the safety and reliability of these products. An app may be updated when its developers discover a security flaw that threatens users' privacy, when an operating system is upgraded to ensure compatibility, or when a new functionality becomes available. Of course, apps may be updated for other reasons, such as incorporating new content, general performance improvement, or fixing bugs.[3] Frequent updates may also be an inherent part of an app's design, such as in some ML technologies. ML, a subset of AI, often uses many adaptive algorithms that can learn and alter the performance and behaviors of the algorithms. In the case of adaptive algorithms, preventing updates (i.e., "locking" the algorithm) would defeat some of the value of using these technologies.[4]

> In determining whether an app update is significant, the first question one should ask is, does this change implicate a subject matter that the IRB considered as part of its initial review of the protocol?

Nevertheless, the capacity and need for updates when using apps present a significant challenge to PIs seeking to use these products in their studies and to the IRBs that review the study protocols. Requiring IRBs to conduct a re-review of the app and attendant research protocol after each update might quickly overwhelm

IRBs' capacities and delay the implementation of necessary updates. Additionally, the option to withhold implementation of the app until after IRB review may not be available if app updates are not under the control of researchers. Allowing apps to update without any oversight, however, would create blind spots. Therefore, it is crucial to determine when an update can be implemented without any notification to the IRB, when it is sufficient for the PI to notify the IRB that an update has occurred, when an update should trigger a new review by the PI or the overseeing IRB, and what standards should be used in evaluating updates to drive these different processes.

It is essential to clarify the responsibilities of each of the parties involved—the IRB, the PI, and the app developer. Such an analysis raises many questions we address in this article, such as the following: Who is responsible for tracking updates? Who is responsible for reviewing updates? What responsibilities can be realistically delegated to these stakeholders, given their differing levels of involvement in research?

THREE KINDS OF UPDATES

To understand how to manage the update problem, we need to recognize that updates differ in content, intent, risk, and significance. Practically, from the perspective of the governance of human participant research, we recommend separating updates into three general categories: minor updates, more significant than minor updates, and significant changes.

The first category consists of updates that are so minor that no notification or re-review is necessary. To use an extreme example, when an app changes the background color of its home screen in an update to a new version, that is not something that requires notification to the IRB or IRB re-review of the study protocol. While the exact border of this category is fuzzy, in general, we think of these as updates that are either so trivial or so orthogonal to the research use as to make the change largely irrelevant. These changes could be fixes to "buggy" code but are minor enough that the user, the IRB, and/or even the PI of the study might not be aware of the change, nor should they care. An example might be fixing a bug that causes the app in question to crash when a particular other app is loaded. Other types of updates that would fall in this category are changes in appearance that do not impact accessibility and changes in coding to improve app performance, such as running more smoothly, using less battery power, or speeding up certain processes. However, whenever there is any doubt whether the update falls in this category, PIs should assume that some notification of the IRB is required. All updates that fall (or, by the PI's judgment, appear to fall) in the first category should be tracked and be included as part of continuing review for those studies that require annual review. Ideally, app developers would cooperate with researchers to maintain a change log or release notes and to track updates for submission at annual review. Ideally, this update tracking would be findable by PIs and IRBs, perhaps by being publicly posted in venues such as the Apple and Google Play stores. Realistically, however, most app developers have little incentive to take on this task, and so tracking minor updates must fall to the PI (see below for a further discussion). In some cases and with advance notice, the applicable IRB may consider waiving the tracking requirement for minor updates.

All other updates require at least prompt notification by the PI to the IRB (second category). Some updates will be more significant than trivial changes; these include, for example, requests for additional information or design and/or navigation changes that could affect the readability or accessibility of the app for some research participants. Some of these kinds of updates will include notable features that are not used as part of the research design but that participants may stumble across as they use the app. Examples of such updates would be adding or removing a language not used by the participants or adding a location feature when it is not used in the research protocol. It is important for the IRB to have a sense of the available functionalities of the apps that researchers use in their studies, but changes in this category may not require re-review by the IRB because the updates in question do not impact the actual conduct of the research or increase potential risk to the participants.

Whether such updates also require IRB re-review (third category) will depend on whether they constitute a significant change. Only updates that qualify as a significant change should be re-reviewed, while other updates should likely result in notification by the PI to the IRB. As we discuss below in more depth, ultimately,

it is for the IRB and *not the PI* to determine whether the change is of such significance to warrant re-review.

SIGNIFICANT CHANGE

The FDA defines "significant change" in the context of medical devices to constitute a major change in the intended use of the marketed device, or a change that could significantly affect its safety and effectiveness (21.C.F.R. 807.81[a][3]), or a major change to the algorithm.[5] This is a helpful standard that could be applied more broadly to the way IRBs think about updates to non-FDA-regulated apps.

In determining whether a change is significant, the first question one should ask is, does this update implicate a subject matter that the IRB considered as part of its initial review of the protocol? IRBs are tasked with a broad but ultimately limited set of concerns when they evaluate research protocols to determine if the research is approvable, with a focus on protecting participants. These requirements include (a) respect for enrolled participants, including protecting their privacy, supporting their well-being, and allowing opportunities to withdraw from the study when necessary; (b) the value or merit of the research; (c) the scientific validity or methodological rigor of the study design; (d) fair subject selection, including equitable distribution of risks and benefits and avoiding exploitation of vulnerable populations; (e) a favorable risk-benefit ratio in which risks are minimized and are proportionate to the potential benefits to individuals and society; (f) independent review of the research protocol before it is implemented; and (g) informed consent for participants involved.[6] Updates that result in a change from what was initially approved and implicate any of these concerns should lead an IRB to consider re-review.

If an update touches upon one of these concerns, one should then ask this question: Does this update negatively impact one of the key concerns that the IRB evaluated in its initial review? For example, does it implicate the privacy of the participants? This is the directional component of the update evaluation. For example, an update that increases privacy protections for participants should trigger a different response than an update that decreases such protections. Only updates that weaken protections in an area of concern for IRBs should trigger re-review. Improvements should not trigger re-review because we know that the research protocol was already considered acceptable in its previous incarnation. Depending on the technological complexity of the update concerned, IRBs may want to consider if they need to consult with additional technological experts to ensure that they understand the potential impact the change will have on the study protocol.

A notable exception here should be when a PI seeks to alter a protocol in response to an improvement, such as omitting language in the informed consent document regarding a risk of reidentification because of a privacy improvement. When such changes are made in the protocol and/or informed consent document in response to an update, the IRB should review those proposed changes, just as it would were those changes made without an underlying update. If possible, until those changes are reviewed and approved by the IRB, the research should continue using the originally approved study materials even if they warn of risks that no longer apply—because it is better to err in this direction.

THE ROLES AND RESPONSIBILITIES OF THE PARTIES INVOLVED

Ultimately, PIs should be responsible for notifying the cognizant IRBs when an update is released; this is similar to other tracking and notification responsibilities they execute in the course of their research. This notification to the IRB from the PI should include information on (a) the nature of the update (including a technical specification from the app designers if available), (b) whether the update alters any of the key subject matter described above, and, if so, (c) whether the changes are improvements or diminutions and how. A PI's failure to notify the IRB of an update that is not minor and is also negatively impactful should be treated as noncompliance, with continued failure to report potentially rising to the level of "serious or continuing noncompliance."

Once notified, the IRB should decide whether an update meets the subject matter and directional criteria for a significant change. The IRB chair or designee can review the notification to determine whether re-review is necessary. IRBs may want to consider which apps are used in a significant number of their approved protocols and, as with evaluations of operation systems, centrally monitor and evaluate any updates to those apps or, at a

minimum, use the information to alert other PIs to any significant changes. IRBs will likely want to add questions regarding app updates during their continuing review of protocols, giving PIs an opportunity to report trivial, minor updates that do not merit notification.

Just as with other significant changes to research protocols, in some instances, these types of updates may merit notifying participants. It is for the IRB and not the PI to decide if and when the PI should reach out to participants to notify them of significant changes or updates to the app, and if in some instances reconsent is necessary, just as the IRB would with other changes to a protocol. To determine the correct course of action, the IRB should consider contextual questions such as how significant the change is, whether it is the kind of thing the average participant will care or want to know about, and how often the research team communicates with participants. Reconsent would likely be necessary if the significant change might influence the participant's willingness to remain enrolled in the research, as it might if, for example, the change introduces new risks to their privacy and the security of their data. Some IRBs may elect to set a presumption that significant changes will require, at a minimum, notification of participants, if not reconsent.

One potentially notable distinction in the types of apps used in research concerns who develops an app and whether the app in question was designed specifically for research. In the case of the A-fib study mentioned above, the app was developed and disseminated by Apple, although the research was conducted by a team from Stanford University.[7] At the time, this app was not an FDA-authorized device. While the study was silent on whether the app used was updated during the study, the app was available to the public through the Apple App Store. This suggests that Apple's first priority would be to maintain the app for general use, rather than to respond specifically to the concerns of researchers. When apps are developed for research or as medical devices (see the FDA guidance discussed above) subject to regulatory oversight, then app developers may be more likely to be responsive to the needs of researchers and IRBs. An incidental use of an app in research or an app intended for a broader audience, however, may involve developers who are unconcerned by the needs and responsibilities of the research community as a small subset of their users.

Our recommended approach to monitoring, evaluating, and categorizing app updates relies on the PI and IRB members to take action. We acknowledge that this may be a challenge for PIs, especially those using apps not specifically designed for research or to which participants have access independent of the research (e.g., on an Apple watch). PIs may want to consider asking participants to alert the PI if they are asked to update or change an app used in the research. Another best practice for monitoring would be for the PI or another member of the research team to maintain the app on their own device, if possible, to monitor any updates pushed from the developer. PIs should also become accustomed to reporting all updates and changes at their continuing or annual IRB review of their protocols. This will allow IRBs to understand all the changes that have occurred over the last year, even if they do not merit immediate notification to the IRB.

In an ideal world, the app developers would bear at least some responsibility in categorizing, notifying, and explaining their updates to the appropriate researchers, IRBs, and participants (when appropriate). We are concerned, however, that many app developers are not engaged in or incentivized or required to participate in communications regarding research. For example, the makers of a fitness track app may be focused on marketing their app to consumers who exercise. If a research team then uses this app as part of their research protocol to monitor activity of patients following surgery, the app developer has no reporting obligations or incentive to engage in the research. We do not want researchers to be blocked from using apps because developers do not participate in research reporting obligations. Because the participation of app developers cannot be relied on (with some exceptions when the app is being developed for a medical application and tested for such), we have argued for a system that does not depend on developers in order to trigger review of relevant updates.

SOME SPECIAL CASES

Until now, we have considered the update problem in the context of individual, relatively straightforward apps, in other words, those that are not operating systems, that are accessed via mobile devices and

not embedded in any medical device, and that do not use ML. The rest of this article will flag some additional considerations for software that does not meet the criteria of a "standard" app.

The update problem as applied to operating systems. Our discussion thus far has focused on app updates, but what should be done about operating-system updates, for example, changes to the underlying iOS or Android systems? One version of this problem is when participants are running research-oriented apps on their own phones and they seek to update their phones' operating systems on their own. This raises the question of when operating-system updates should be encouraged or discouraged, a problem familiar to many information technology (IT) departments across the world. It may be necessary to support multiple versions of the operating system, at least when an update is new. Even though such operating-system updates may make a difference as to privacy or other issues within the IRB's purview, our view is that it is better to handle this outside the research ethics apparatus. This is because operating-system updates will affect the use of all devices and apps in the institution; such an update can raise the question, for instance, whether clinicians should be allowed to update their devices used to access electronic health records. Therefore, it is likely that an evaluation of operating-system updates is already being conducted at the institutional level, and decisions must be balanced across different users. The better approach to avoid duplication of effort is for an institutional official or administrator, perhaps working with IT experts or third-party consultants, to send out an alert to all PIs in cases of significant changes to major operating systems, such as iOS and Android. The PIs should then consult with the IRB to determine whether, when, and how participants should be informed of these particular changes.

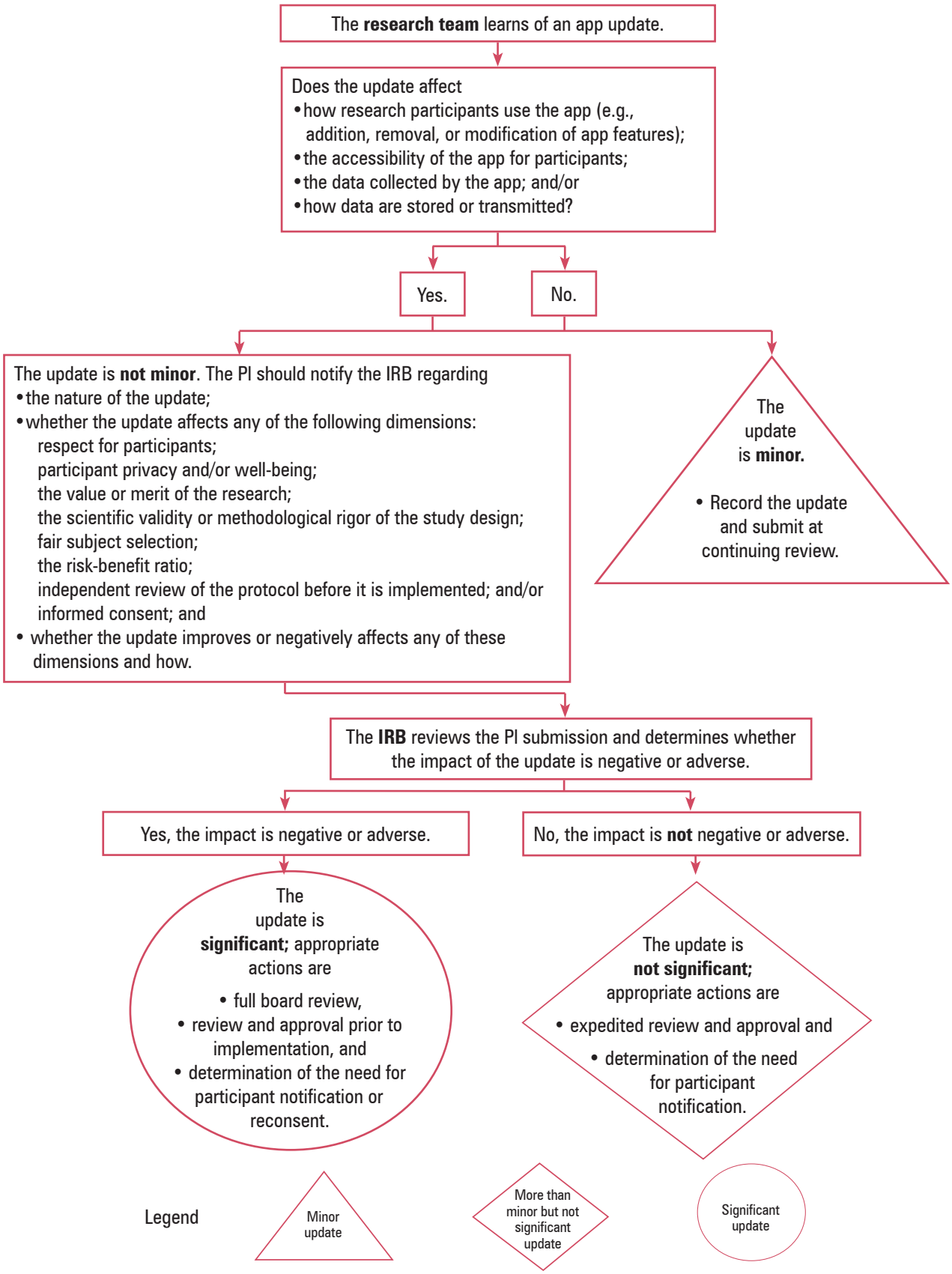
The update problem as applied to participants' own devices. As mentioned above, there are additional complexities when participants use their own devices to run research-related apps. They may choose to update their operating systems before the research team is able to evaluate the update in question and to send it for any necessary review. Participants using their own devices may also be more likely to allow updates to research-related apps or may allow updates to be installed without knowing that the updates will affect the research-related app, perhaps because they have set all apps on their device to automatically update. During enrollment and the informed consent process, the research team should flag issues with accepting potential updates with participants and discuss why it is better to hold off on installing updates until the team has explicitly approved the updates in question. Participants should also be urged to contact the research team when they have installed updates to the operating system or to the research app in question.

The update problem as applied to embedded algorithms and software, including adaptive algorithms. When apps are embedded within a medical device, there are some additional considerations for IRBs. If the app meets the medical device definition under the Federal Food, Drug, and Cosmetic Act,[8] as with Apple's A-fib app, and is exclusively for use in research involving human participants, PIs conducting such research are exempt from listing and registering the app with the FDA, but they may need an approval from the FDA for an investigational device exemption.[9] IRBs and research institutions may also consider the FDA's guidance on deciding when to submit a 510(k) premarket notification for a change to an existing medical device.[10] Under 21.C.F.R. 807.81(a)(3), a significant change of marketed devices requires a new premarket notification.

A good starting point for addressing the update problem is to distinguish between minor changes (which require, at most, annual notification), changes that are not minor (which require prompt notification), and changes that are significant (which may require full board re-review or another response).

Apps that use adaptive algorithms, loosely defined as ML that learns on "the go" as new data is encountered rather than being "locked," are especially difficult for

Figure 1. App-Update Decision Tree



IRBs to evaluate because they are constantly changing as new data is inputted. Additionally, these changes are often opaque to outside observers. The key question in evaluating shifts in adaptive algorithms should be, does this evolution or series of changes lead to significant new risks in the relevant AI/ML system? When in the course of adaptation should the PI or the IRB stop the research to reevaluate? The ultimate goal is to determine whether the change(s) significantly impacts the safety, efficacy, and reliability of the app. Concept drift in ML, occurring when the true relation between inputs and outputs changes, would constitute a significant change meriting re-review.[11] An example of concept drift might be an AI/ML system trained to evaluate skin lesions as benign or malignant that further incorporates more diverse images of race and skin type into its analysis to improve its own accuracy and efficacy. Of course, using independent data can help expose underlying embedded biases in the original training or test dataset, which demonstrates the challenge of evaluating changes in ML systems from a participant protection standpoint. Another significant change would be covariate shift, or when the input distribution of new data is different from the input distribution of the original data used to train and/or test the adaptive algorithm.[12] For example, an algorithm to identify different types of cancers trained mainly on images from men may experience a covariate shift when applied to female patients. Changes to the stability of an algorithm are also a particular concern for AI/ML apps, as these systems are unfortunately especially vulnerable to instability.[13] Stable algorithms produce similar predictions when given slight variations in inputs.[14] In the context of medical AI, stable clinical decision-making algorithms would produce similar diagnoses for medically similar patients. An increase in instability in an AI/ML system should be flagged as a significant change because its predictions will be less consistent from subject to subject.

Admittedly, evaluating concept drift, covariate shift, and stability is challenging, especially to PIs and IRBs that may not have the expertise to do so. Additionally, some adaptive algorithms developed in research institutions are subject to quality-improvement review rather than research oversight, suggesting that responsibility falls to the developer of these algorithms (although this is not the case when researchers participate in the research and validation of adaptive algorithms developed by outside companies). Nevertheless, leadership at research institutions that employ adaptive algorithms developed within the institution should consider whether their IRBs are equipped to perform oversight of these algorithms when appropriate (that is, when they are not viewed as quality improvement), including the administrative burden involved with such an analysis, whether the IRB has the content expertise to review adaptive algorithms, or if there are other individuals, groups, or departments who might be better positioned to do so. This content expertise could be found in a bioinformatics department or outside computer science consultants. IRBs tasked with reviewing adaptive algorithms should consider engaging with individuals with additional technical expertise to ensure they understand what changes are occurring and the significance of those changes.

Regulatory agencies, such as the FDA, may also be a source of guidance for IRBs and research institutions struggling with the update problem in regard to their evaluation of products that incorporate adaptive algorithms. For example, the FDA is currently developing a framework for modifications of AI/ML-based software as a medical device (so-called SaMD),[15] including a recently released action plan outlining next steps.[16] IRBs may be able to learn from some of the rules the FDA sets and consider applying them to apps that are not subject to FDA regulation.

CONCLUSION

With the significant increase in the use of apps in biomedical and behavioral research, PIs and IRBs are increasingly confronted by the challenge of evaluating app updates. Best practice standards and guidance for the supervision and responsibility of app updates used in approved research protocols are needed for IRBs. A good starting point for addressing the update problem is to distinguish between minor changes (which require, at most, annual notification), changes that are not minor (which require prompt notification), and changes that are significant (which may require full board re-review or another response). Minor changes can be reported by the PI during annual review. All other changes should be promptly reported by the PI to the relevant IRB, with some substantive description included in the report. See figure 1 for an overview. While

the PI bears tracking and notification responsibilities, the IRB should be the entity to determine if a nonminor change is significant. To minimize the workload of IRBs and ensure that updates that truly merit consideration are given appropriate attention, only significant-change updates, or those that negatively affect subject matter that the IRB considered in its initial review of the protocol, should result in re-review. In the case of adaptive algorithms, which by their nature involve constant and often obscure flux, IRBs should consider upon initial review and approval what changes would require or prompt reporting or re-review. Lastly, IRBs should create relationships and communications with technical experts to help them make determinations related to updates, including when a change is significant and when an adaptive algorithm should be reviewed. ♦

Carmel Shachar, JD, MPH, *is the executive director of the Petrie-Flom Center at Harvard Law School;* **Sara Gerke, Dipl-Jur Univ, MA,** *is an assistant professor of law at Penn State Dickinson Law and a leadership team member of the Project on Precision Medicine, Artificial Intelligence, and the Law (PMAIL) at the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School;* **Walker Morrell, MBE,** *is a project manager at the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard;* **Aaron Kirby, MSc,** *is the director of regulatory affairs operations at Harvard Catalyst;* **I. Glenn Cohen, JD,** *is the deputy dean, a James A. Attwood and Leslie Williams Professor of Law, and the faculty director of the Petrie-Flom Center at Harvard Law School; and* **Barbara E. Bierer, MD,** *is a professor of medicine at Harvard Medical School and Brigham and Women's Hospital, the faculty director of the Multi-Regional Clinical Trials Center, and the program director of regulatory foundations, ethics, and the law at Harvard Catalyst.*

ACKNOWLEDGMENTS

This work was conducted with support from Harvard Catalyst, The Harvard Clinical and Translational Science Center (through grant UL1 TR002541 from the National Center for Advancing Translational Sciences at the National Institutes of Health), and financial contributions from Harvard University and its affiliated academic health care centers. The contributions of I. Glenn Cohen, Carmel Shachar, and Sara Gerke were also supported by a grant from the Collaborative Research Program for Biomedical Innovation Law, a scientifically independent collaborative research program supported by a Novo Nordisk Foundation grant (NNF17SA0027784).

DISCLAIMER

The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, or the National Institutes of Health.

REFERENCES

1. Marra, C., et al., "Quantifying the Use of Connected Digital Products in Clinical Research," *NPJ Digital Medicine* 3 (2020): article 50.

2. Perez, M. V., et al., "Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation," *New England Journal of Medicine* 381 (2019): 1909-17.

3. McIlroy, S., N. Ali, and A. E. Hassan, "Fresh Apps: An Empirical Study of Frequently-Updated Mobile Apps in the Google Play Store," *Empirical Software Engineering* 21 no. 3 (2015): 1346-70.

4. Babic, B., et al., "Algorithms on Regulatory Lockdown in Medicine," *Science* 366 (2019): 1202-4.

5. U.S. Food and Drug Administration, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)," discussion paper and request for feedback, 2019, https://www.fda.gov/media/122535/download.

6. Emanuel, E. J., D. Wendler, and C. Grady, "What Makes Clinical Research Ethical?," *Journal of the American Medical Association* 283 (2000): 2701-11.

7. Perez, "Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation," 1910.

8. U.S. Food and Drug Administration, "De Novo Classification Request for Irregular Rhythm Notification Feature," 2018, https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180042.pdf.

9. U.S. Food and Drug Administration, *Policy for Device Software Functions and Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff* (Silver Spring, MD: U.S. Food and Drug Administration, Center for Biologics Evaluation and Research, 2019), https://www.fda.gov/media/80958/download.

10. U.S. Food and Drug Administration, *Deciding When to Submit a 510(k) for a Change to an Existing Device: Guidance for Industry and Food and Drug Administration Staff* (Silver Spring, MD: U.S. Food and Drug Administration, Center for Biologics Evaluation and Research, 2017), https://www.fda.gov/media/99812/download.

11. Babic et al., "Algorithms on Regulatory Lockdown in Medicine," 1203.

12. Ioffe, S., and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Proceedings of the 32nd International Conference on Machine Learning* 37 (2015) 448-56.

13. Goodfellow, I., J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," arXiv, submitted December 20, 2014, doi:10.48550/arXiv.1412.6572.

14. Dwork, C., et al., "Fairness through Awareness," *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (2012): 214-26.

15. U.S. Food and Drug Administration, "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)."

16. Center for Devices and Radiological Health, *Artificial Intelligence and Machine Learning (AI/ML) in Software as a Medical Device Action Plan* (U.S. Food and Drug Administration, January 2021), at https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.