Patient-Reported Outcomes: A Privacy-Centric and Federated Approach to Machine Learning

Sara Jordan [*1] **Rachele Hendricks-Sturrup** [*1]

Abstract

Machine learning developers, healthcare practitioners, ethicists, and privacy professionals can all agree: health data requires special protections. Should certain types of health data warrant specific protections? In this scientific position paper, we answer this question affirmatively and stake the position that patient reported outcomes measures (PROM) data requires specific protections, particularly when it is used in or informed by machine learning regimes. We posit that a federated learning architecture is appropriate, both when model components are sent or gradients received, for ensuring privacy of users' data. However, use of a federated learning approach for privacy without attention to security dimensions of private data management may compromise users' data unexpectedly.

1. Patient-Reported Outcomes

Calls for protection of healthcare data are commonplace in discussions of ethical uses of machine learning and artificial intelligence (Kent, 2020). The conversation around protection of health data is becoming more nuanced as scholars and advocates realize that the requirements for effective protection of genetic data differ from protection of electronic health record (EHR) data (Fernández-Alemán et al., 2013). In this position paper, we advance the argument that specific forms of health data require specific protections by stating the claim that patient reported outcomes measures (PROM) data should be protected in specific ways. Our central position is that, when collected through user's mobile devices, PROM data should be protected by use of federated data architecture, while ensuring high standards of quality in reporting and managing patient data and consent to privacy policy terms and conditions (World Economic Forum, 2019; Xu and Wang, 2019). When such data is part of machine learning enabled PROM applications, a federated approach to machine learning can be essential to protect users' privacy, to protect the investments in valid PROM instruments, and to protect investment in model development.

1.1. PROs and PROMs

PROs are broadly defined by the United States (US) Food and Drug Administration (FDA) and the National Institute of Health (NIH) as "a measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else" (FDA and NIH, 2016). PROs, being a raw form of patient expression and feedback, help clinicians, researchers, medical device and drug manufacturers, and governmental stakeholders overseeing medical device and drug development, distribution, and safety monitor, understand, and document, in a readable format, patients' symptoms, preferences, complaints, and/or experiences following a clinical intervention. PROs are often context-driven and specific to disease, indication, and context, allowing PROs to precisely inform clinical practice and regulatory decisions by capturing very specific endpoints related to quality of life (e.g. energy level and nausea/vomiting; FDA, n.d.). PROs also inform the development of broader initiatives like value-based payment reform and health system-level quality improvement initiatives (Squitieri et al., 2017).

Through an assortment of paper and digital questionnaires, surveys, diaries, and digital reporting tools, PROs are captured and measured during doctor visits using 'paper-and-pencil' strategies or, more recently, digital and/or mobile device tools that capture PROs *in situ* or in real time to and identify possible changes in patient experience and feedback longitudinally. PRO measures (PROMs) are embedded within these instruments or tools to gather quantitative and qualitative data reflecting the health status of a patient, allowing for the data to be used to correlate or predict serious adverse events like hospitalizations for acute and life-threatening symptoms (FDA, 2019). Mobile health tools can be reliable as a time sampling strategy for re-

[*] Equal contribution [1] Future of Privacy Forum, Washington DC, USA. Correspondence to: Sara Jordan <sjordan@fpf.org>, Rachele Hendricks-Sturrup <rhendrickssturrup@fpf.org>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020. Copyright 2020 by the author(s).

mote multidimensional symptom and behavior reporting, as they have helped more conveniently capture PRO histories among disease populations with complex symptoms and requiring complex care (e.g. juvenile idiopathic arthritis, chronic kidney disease, and substance abuse disorder; Lee et al., 2020; Goodday et al., 2020; Groveet al., 2019). For example, PROs collected using the Kansas City Cardiomyopathy Questionnaire from patients at high risk of heart failure correlate with the New York Heart Association's classification system and are used to predict hospitalization and death (Hawwa et al., 2017). These validated PROMs and correlations can therefore become powerful tools and drivers of quality and foresight in patient-centered care.

Popular paper-based PROM surveys include the Harris Hip Score (HHS), the 36-Item Short-Form Health Survey (SF-36) Physical Component Summary (PCS), the SF-36 Mental Component Summary (MCS), and the EuroQol five-dimension three-level (EQ-5D) index. Their conversion from paper to remote electronic format warranted further study to determine if patient responses might differ based on this delivery format (White et al., 2018; Meirte et al., 2020). PROMs, collected in remote electronic fashion or otherwise, however, are validated and used with the intent to drive clinical decision making in patient-centered care and understand overall patient outcomes based on a collection of factors from multiple data sources.

For instance, the Center for Outcomes Research and Evaluation (CORE), led by Yale University and the Mayo Clinic, are undergoing initiatives to explore mechanisms to capture longitudinal PROMs using mobile health technology and ways to integrate and combine PROMs with data from multiple sources, like health insurance claims databases, wearable/mobile device data, pharmacy data, and electronic health records (EHRs) to support post-market medical device surveillance strategies (CORE, n.d.). Additionally, newly created or adapted PROMs (electronic and paper-based) have become a data strategy to assess the value and utility of relatively newer clinical services like genetic counseling, especially as genetic testing becomes more implemented in clinical practice for various purposes and disease groups and recommended in clinical guidelines (Pitini et al., 2019; Yuen et al., 2020).

1.2. Privacy Concerns

A systematic literature review of benefits and disadvantages to the electronic capture of PROs identified and described privacy protection, or lack thereof, as a key disadvantage to the systematic and routine collection of PROMs (Meirte et al., 2020). Two studies included in the review reported privacy concerns (Liu et al., 2016; Hartkopf et al., 2017). Liu et al. found that most of their surveyed patients (71.7 percent) thought their privacy should be adequately protected and Hartkopf et al. reported that 30 percent of their surveyed patients were concerned about privacy issues and that those concerns significantly influenced the patient's willingness to participate in the collection and reporting of electronic PROMs. As emerging evidence reveals potential benefits to the use and implementation of all types of machine learning in clinical practice, and as machine learning increasingly incorporates PROMs comprising both structured and unstructured electronic data, it is crucial to explore potential strategies that could help patients and other stakeholders (e.g. caregivers, regulatory agencies, etc.) overcome privacy concerns in the electronic, remote, and high-quality capture of PROMs.

2. Privacy-Preserving and Secure Machine Learning for Electronically-Reported PROs

PROs can be drawn from paper and pencil (PP) or digital records. Today, most providers, from pediatricians to clinical trialists, trust that patients rely on the wealth of computer and mobile phone based applications (apps) to track their behaviors (e.g., blood glucose testing), symptoms (e.g., painful menstruation), treatments (e.g., medication trackers), and other wellness related actions (e.g., diet and exercise; Fierce Biotech, 2011; Agency for Healthcare Research and Quality, 2019). Patients and caregivers place a high degree of trust in the developers of PROMs to secure their data, to share their data only for preferred purposes (e.g., medical research), and to use their data for legitimate user-interface/ user-experience (UI/UX) improvements, not only for advertising or promoting paid upgrades to the app (Eisner et al., 2019). Regardless of the purpose or design of a PROM, and as we've described, PROM data represents a wealth of information. For example, it is data for treatment, data for research, and data for UI/UX studies. Likewise, it is valuable data fodder for development of predictive tools, such as AI, that may nudge patients, build better more interactive and empathetic care robots, or fuel powerful correlational studies in health emergency situations. Management of privacy risk from uses of structured data, such as survey responses to numeric ratings of pain, are meaningfully different than is management of privacy and quality risk from uses of unstructured, user-entered, data. For instance, responses to structured queries, such as daily or otherwise routine questionnaires, can provide detailed information on a patient's lifestyle, behaviors, or even location, are measurable in a fixed and validated format, whereas unstructured data are more abstract and open to interpretation (Belarmino et al., 2019). Since PROM data shared via mobile apps can reveal fine grained information about a patient or caregiver, systems using this data should adopt a nuanced approach to creating and implementing privacy policies and machine readable patient consent to uphold patient trust (Benze et al., 2019). Additionally, systems should underwrite the risks and benefits to adopting federated machine learning to assess PROM data alongside external clinical records.

2.1. Privacy Policies and Machine-Readable Consent

A foremost and key component to preserving privacy of user's data is informing the users which types of data will be collected and how that will be used (O'Loughlin et al., 2019) In the case of apps which collect and use routine entry data, privacy policy language should reflect that the user's data is tracked for trends over time. Longitudinal or time series data reveal patterns of life that can be valuable information for practitioners but also valuable information for third parties, whether those third parties are neutral or positive (e.g., advertisers) or nefarious (e.g., hackers; Saleheen et al., 2016) Indicating how data is tracked within a system requires some nuance as data can be used in many ways. For example, if survey responses are time-aggregated for user inspection over a time period and or are time-aggregated prior to transmission, users should be informed that time is a key variable. Likewise, if the apps use location data to support users or augment services, the uses of that data should be made clear. Clear privacy policies will indicate to users which data will be used to generate or improve predictions or classifications of their individual responses or will be used in aggregate to improve UI/UX for all users.

Well written privacy policies may indicate how data is used, but research suggests that many users do not read, nor do they retain information from, privacy policies. "Clicking through" permissions and policy notifications is sufficiently common-place that new technologies are being built to help users manage their privacy preferences through aggregation of multiple apps' privacy policies. To ensure that such meta-apps are useful and operate well, machine readable privacy policies are essential. Likewise, machine readable consent terms are essential, especially when express consent is offered in a nuanced way to capture patient preferences across any data sharing and use spectrum (e.g. sharing PROM data with third-parties, PROM data use for product development purposes or generalizable knowledge research, incompatible secondary uses, etc.). Making consent machine-readable also allows for consent tools to become data themselves and searchable queries.

2.2. Federated Approaches to Privacy-Preserving Machine Learning

Securing private user data, whether text or images, structured or unstructured, throughout the multiple steps of use by PRO apps, presents an interesting challenge for app developers, systems architects, and machine learning programmers (Lin, Yang, and Wang, 2016). An essential consideration is whether models will be segmented and sent to users' devices for "on-device" training, or whether data will be encrypted and possibly aggregated and sent to a central model server. Whether a company chooses to disaggregate learning or to centralize learning depends on important expectations about user devices, performance of serialized models, and privacy demands. For example, to use on-device learning (see Figure 1), there must be an expectation that users' device capacity will allow the transmission of model components, the training of those components, and the transmission of revised model gradients and model components by the device without significant costs to users (e.g., data use or app lags).

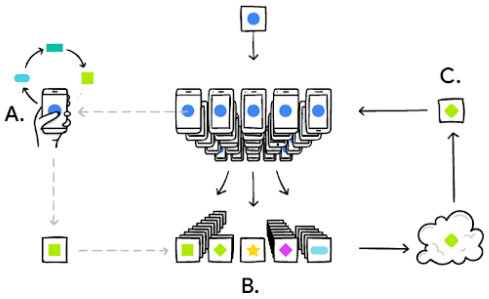

Figure 1. Google AI illustrates federated learning as "Your phone personalizes the model locally, based on your usage (A). Many users' updates are aggregated (B) to form a consensus change (C) to the shared model, after which the procedure is repeated" (Google AI, 2017).

Alternatively, to use a centralized system, there should be a reasonable expectation that data can be encrypted or perturbed (as in the case of homomorphic encryption) on the device prior to transmission to a central server, that devices can successfully cross-talk to mask individuals through creation of a federated average sent back, and that performance gains from centralizing the model are not outweighed by losses from breach of user data. Likewise, for both client-side and server-side learning, there must be a reasonable expectation that regular transmission patterns will not open the models for attacks, whether upon the model (e.g., model inversion attacks) or upon users' data (e.g., membership inversion attacks). These general problems for uses of federated learning approaches cover virtually any type of data, but uses of an on-device approach to federated learning using patient reported health data presents some special concerns.

When PROM tools are used with machine learning systems, whether in the context of massive-N multi-national post-market surveillance studies or small-N studies of unique patient populations, additional concerns arise. From the perspective of model architecture, a distributed server network, whether it transfers data, model components, or both may be subject to compromise due to local conditions. For example, localized rolling power failures can compromise systems as can changing requirements for local processing of citizens' data. From the perspective of model security, some authors have made the case that models trained on users' data become, in effect, a piece of personal data which attaches to that user. In this latter case, redefining model components as personal data changes the requirements for centralized or highly distributed processing of that model.

3. Conclusion

On-device PROM tools that use machine learning may employ the best privacy policies or consent mechanisms, but may ultimately leave key components, such as the security dimensions, of privacy up to the user. Keeping data in the hands of users opens the user up to unanticipated vectors of attack from adversaries striving to identify the valuable machine learning models or seeking to uncover data about a specific patient. User device security and security for transmission of either data (raw or processed) or model gradients are key concerns that must be addressed when using a federated learning system. For PROMs where similarity of user input over time or similarity between answers to structured and unstructured queries are salient, it is crucial that developers ensure that choices about models do not open users to attack or undue influence and thus do not open practitioners to liability for interpretation of false responses. Likewise, for the developers of PROM systems, which may be tied to clinical trials, post-market drug studies, or hospital performance scores, it is business critical to ensure that models are not compromised and valuable machine learning spending lost to competitors. Finally, federated machine learning architectures should be tested and validated to ensure quality of unstructured PROM data versus influencing or skewing PROs concerning safety, symptoms, and other important outcomes.

References

Agency for Healthcare Research and Quality. (2019, March 5). New App Designed To Help Patients Report Health Outcomes Wins Top Prize from AHRQ. https://www.ahrq.gov/news/newsroom/press-releases/stepup-app-phase2-winners.html

Belarmino, A., Walsh, R., Alshak, M., Patel, N., Wu, R., and Hu, J. C. (2019). Feasibility of a Mobile Health Application To Monitor Recovery and Patient-reported Outcomes after Robot-assisted Radical Prostatectomy. *European Urology Oncology, 2*(4), 425-428. doi:10.1016/j.euo.2018.08.016

Benze, G., Nauck, F., Alt-Epping, B., Gianni, G., Bauknecht, T., Ettl, J., . . . Gaertner, J. (2017). PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. *Management, 8*(2),104-111. doi:10.21037/apm.2017.07.05

CORE. (n.d.). Yale University-Mayo Clinic CERSI. https://medicine.yale.edu/core/current_projects/cersi/research/

Eisner, E., Drake, R. J., Berry, N., Barrowclough, C., Emsley, R., Machin, M., and Bucci, S. (2019). Development and long-term acceptability of ExPRESS, a mobile phone app to monitor basic symptoms and early signs of psychosis relapse. *JMIR mHealth and uHealth, 7*(3), e11568. doi:10.2196/11568

FDA. (n.d.). PRO Report Appendix: Patient-Reported Outcome Measure (PRO) Case Studies. https://www.fda.gov/media/125193/download

FDA. (2019). Patient-Reported Outcomes (PROs) in Medical Device Decision Making. https://www.fda.gov/about-fda/cdrh-patient-engagement/patient-reported-outcomes-pros-medical-device-decision-makingwhatare

FDA and NIH. (2016). BEST (Biomarkers, EndpointS, and other Tools). https://www.ncbi.nlm.nih.gov/books/NBK326791/

Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., and Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. *Journal of Biomedical Informatics, 46*(3), 541-562. doi:10.1016/j.jbi.2012.12.003

Fierce Biotech. (2011, February 7). Patient-reported Outcomes - Smartphone apps for clinical trials. https://www.fiercebiotech.com/special-report/patient-reported-outcomes-smartphone-apps-for-clinical-trials

Goodday, S. M., Atkinson, L., Goodwin, G., Saunders, K., South, M., Mackay, C., . . . Welch, J. (2020). The True Colours Remote Symptom Monitoring System: A Decade of Evolution. *Journal of Medical Internet Research, 22*(1), e15188. doi:10.2196/15188

Google AI. (2017, April 6). Federated Learning: Collaborative Machine Learning without Centralized Training Data. https://ai.googleblog.com/2017/04/federated-learning-collaborative.html?m=1

Grove, B. E., Ivarsen, P., de Thurah, A., Schougaard, L. M., Kyte, D., and Hjøllund, N. H. (2019). Remote follow-up using patient-reported outcome measures in patients with chronic kidney disease: the PROKID study - study protocol for a non-inferiority pragmatic randomised controlled trial. *BMC Health Services Research, 19*(1), 631. doi:10.1186/s12913-019-4461-y

Hartkopf, A. D., Graf, J., Simoes, E., Keilmann, L., Sickenberger, N., Gass, P., . . . Wallwiener, S. (2017). Electronic-based patient-reported outcomes: willingness, needs, and barriers in adjuvant and metastatic breast cancer patients. *JMIR Cancer, 3*(2), e11. doi:10.2196/cancer.6996

Hawwa, N., Vest, A. R., Kumar, R., Lahoud, R., Young, J. B., Wu, Y., . . . Cho, L. (2017). Comparison between the Kansas City cardiomyopathy questionnaire and New York heart association in assessing functional capacity and clinical outcomes. *Journal of Cardiac Failure, 23*(4), 280-285. doi:10.1016/j.cardfail.2016.12.002

Kent, J. (2020, March 20). *Privacy Protection Key for Using Patient Data to Develop AI Tools.* Analytics in Action News. https://healthitanalytics.com/news/privacy-protection-key-for-using-patient-data-to-develop-ai-tools

Lee, R. R., Shoop-Worrall, S., Rashid, A., Thomson, W., and Cordingley, L. (2020). "Asking Too Much?": Randomized N-of-1 Trial Exploring Patient Preferences and Measurement Reactivity to Frequent Use of Remote Multidimensional Pain Assessments in Children and Young People With Juvenile Idiopathic Arthritis. *Journal of Medical Internet Research, 22*(1), e14503. doi:10.2196/14503

Lin, W. Y., Yang, D. C., and Wang, J. T. (2016). Privacy preserving data anonymization of spontaneous ADE reporting system dataset. *BMC Medical Informatics and Decision Making, 16*(1), 58. doi:10.1186/s12911-016-0293-4

Liu, X., Wang, R., Zhou, D., and Hong, Z. (2016). Feasibility and acceptability of smartphone applications for seizure self-management in China: questionnaire study among people with epilepsy. *Epilepsy Behavior, 55*, 57-61. doi:10.1016/j.yebeh.2015.11.024

Meirte, J., Hellemans, N., Anthonissen, M., Denteneer, L., Maertens, K., Moortgat, P., and Van Daele, U. (2020). Benefits and Disadvantages of Electronic Patient-reported Outcome Measures: Systematic Review. *JMIR Perioperative Medicine, 3*(1), e15588. doi:10.2196/15588

O'Loughlin, K., Neary, M., Adkins, E. C., and Schueller, S. M. (2019). Reviewing the data security and privacy policies of mobile apps for depression. *Internet Interventions, 15*, 110-115. doi:10.1016/j.invent.2018.12.001

Pitini, E., D'Andrea, E., Rosso, A., Massimi, A., Unim, B., De Vito, C., . . . Villari, P. (2019). Genetic services for Hereditary Cancer: a systematic review of Patient Reported Outcomes studies: Erica Pitini. *European Journal of Public Health, 29*(Supplement$_4), ckz186 - 380.doi: 10.1093/eurpub/ckz186.380$

Saleheen, N., Chakraborty, S., Ali, N., Rahman, M. M., Hossain, S. M., Bari, R., . . . Kumar, S. (2016, September). mSieve: differential behavioral privacy in time series of mobile sensor data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 706-717). doi:10.1145/2971648.2971753

Squitieri, L., Bozic, K. J., and Pusic, A. L. (2017). The Role of Patient-Reported Outcome Measures in Value-Based Payment Reform. *Value in Health, 20*(6), 834–836. doi:10.1016/j.jval.2017.02.003

White, M. K., Maher, S. M., Rizio, A. A., and Bjorner, J. B. (2018). A meta-analytic review of measurement equivalence study findings of the SF-36® and SF-12® Health Surveys across electronic modes compared to paper administration. *Quality of Life Research, 27*(7), 1757-1767. doi:10.1007/s11136-018-1851-2

World Economic Forum. (2019). *Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data.* http://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf

Xu, J. and Wang, F. (2019). Federated Learning for Healthcare Informatics. *arXiv*, preprint arXiv:1911.06270.

Yuen, J., Lee, S. Y., Courtney, E., Lim, J., Soh, H., Li, S. T., . . . Ngeow, J. (2020). Evaluating empowerment in genetic counseling using patient-reported outcomes. *Clinical Genetics, 97*(2), 246-256. doi:10.1111/cge.13646