Tufts Medicine
Tufts Medical Center

LEAPS

Learning Ecosystems Accelerator for
Patient-centered, Sustainable innovation

PART I

LEAPS Methods Innovation Team proposes new frameworks for RWE

Stakeholders across the healthcare community, from biomedical innovators to payers to clinicians, seek a unified goal: get the right treatments, to the right patients, at the right time.

Real-world data (RWD) are a critical element of the evidence needed to help reach this goal, potentially informing a new drug's development or providing insight for its use after it comes onto market. But current frameworks for analyzing RWD aren't sufficient to provide the detailed, patient-level information that's really needed for treatment decisions.

NEW FRAMEWORKS FOR RWE

Part I of a three-part series

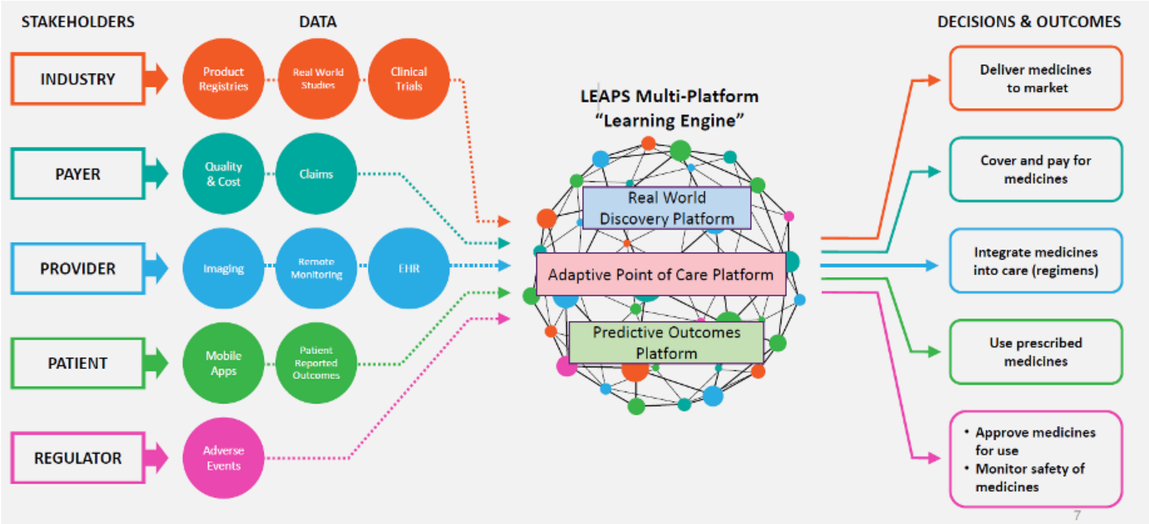
PART I
LEAPS Methods Innovation Team proposes new frameworks for RWE

PART II
Data Assessment & Risk Engineering

PART III
Federated Machine Learning to improve real-world evidence generation

The NEWDIGS LEAPS (Learning Ecosystems Accelerator for Patient-Centered, Sustainable Innovation) Project has sought new methods of planning, producing, and using real-world evidence (RWE) for the last five years by convening experts from different parts of the healthcare community and researching potential options. The project analyzed and designed solutions that would generate evidence to improve patient outcomes, while also benefiting all key stakeholders (pharmaceutical companies, payers, providers, regulators, and patients), in order to ensure sustainability.

LEAPS envisions a learning health system where all stakeholders contribute to and benefit from the knowledge produced.



Within LEAPS, the Methods Innovation Team has specifically focused on finding and evaluating new analysis tools that may be used to generate clinically meaningful hypotheses from RWE. The team's research and discussion in 2022 led to two major proposals for other stakeholders in this field:

1. **Data Assessment and Risk Engineering (DARE):** Healthcare data researchers should focus more heavily on evaluating their underlying data sources for potential bias. After this assessment, they should seek to mitigate bias through statistical methods and by pulling together results from disparate datasets. Resulting conclusions will be more reflective of real-world populations impacted by the research.
2. **Federated Machine Learning (FML):** This machine learning strategy, which connects disparate datasets for analysis, is a promising option for generating hypotheses from real-world healthcare data. The team argues that FML has the potential to simultaneously address two major challenges: data silos within the decentralized U.S. healthcare system, and biased datasets that overrepresent patients who receive easier access to care at the expense of minority groups.

In this package, we describe the NEWDIGS LEAPS Methods Innovation Team's findings on DARE and FML. We'll outline the potential value that these proposed strategies for healthcare data analysis may bring to other researchers and will provide new frameworks and research directions for data scientists interested in pursuing these topics further.

INTRODUCING THE TEAM

In early 2022, the LEAPS Project formed a new working group called the Methods Innovation Team. The team aimed to explore how distributed data networks and machine learning tools could provide hypotheses to inform drug therapy regimens; hypothesis generation was the primary goal, rather than clinical decision-making.

Like other workstreams and teams under the NEWDIGS umbrella, the Methods Innovation Team brought together health experts representing from various specialties and backgrounds, including data scientists, epidemiologists, and clinicians from academia, biopharmaceutical and health technology companies, and others. Overall, the team represented different stakeholders interested in RWE.

In early meetings, some team members expressed particular interest in compiling modeling results from different datasets across a distributed network, seeking to mimic the process of meta-analysis. After further research and discussion, the team arrived at FML as a tool that would serve this purpose; using this machine learning strategy, disparate datasets—which can't be combined due to security and business concerns—may participate in the same study. The team also investigated the problem of biased data in RWE research, which led to developing the DARE framework.

RESEARCH PROCESS

The Methods Innovation Team collectively considered FML during a series of meetings in 2022, including team-specific meetings and broader LEAPS Project Design Labs. NEWDIGS staff conducted an extensive literature review to better understand potential approaches and use cases for FML, which informed the meeting discussions.

In July 2022, the team split into three subtask teams to examine FML's potential in more detail:

- **Data and technical skills:** This team developed a framework to evaluate potential data providers for FML research, focusing on data sources that could be used in a possible pilot study that NEWDIGS considered running. In doing so, the team identified potential metrics, strengths, and limitations for datasets that might be used.
- **Machine learning and statistical models:** This team assessed different machine learning and statistical models that could potentially be utilized in a FML project. They considered a variety of characteristics for each potential model, including its level of transparency, ease of interpreting results, flexibility for additional research, and potential bias that could be introduced by the model.
- **Federated learning types and implementation:** This team explored several FML techniques (i.e., different distributed analysis methods) and platforms that could be used to run an FML project. They also identified some challenges that will arise for researchers who use FML for analyzing healthcare data, which may need to be addressed through further study.

The subtask teams initially planned for their work to inform a NEWDIGS-led test of FML for healthcare data. They planned to study which patients with advanced non-small-cell lung carcinoma (NSCLC), a common type of lung cancer, would be most likely to benefit from a chemotherapy treatment targeting immune system checkpoints. This study would have utilized a diverse group of datasets (potentially including electronic health records, insurance claims, patient-reported outcomes, etc.). While this pilot project did not materialize due to challenges with obtaining buy-in from data providers, the team members found their research helpful in developing frameworks for future projects.

In addition to research on FML, the team studied the problem of biased data in healthcare, which we propose may be addressed through this machine-learning strategy. NEWDIGS staff conducted an additional literature review on this topic, examining how healthcare data scientists assess their source datasets for bias and potential techniques (in addition to FML) which may be used to mitigate this issue.

MULTI-STAKEHOLDER APPROACH

As with other NEWDIGS projects, the FML research included a variety of stakeholders both on the Methods Innovation Team and outside of it. In this approach, it is particularly important to solicit feedback from clinicians, patient representatives, and others who represent the end user of a particular research product, as the product's success ultimately depends on those groups.

The team primarily sought outside feedback at Design Labs, NEWDIGS events that convene key stakeholders and cross-disciplinary researchers. The November 2022 Design Lab included presentations by Methods Innovation Team members about their FML research, as well as brainstorming sessions that invited participants to share their ideas on this novel topic.

For the team members, Design Labs and other feedback sessions were valuable opportunities to entertain new questions about their work. Outside experts expressed excitement about FML's potential while introducing potential challenges for further research.

While NEWDIGS's FML test did not materialize as initially planned, the team's work has informed a test project at Merrimack College, which will be described in further detail in Part III of this research brief series.

ABOUT LEAPS

LEAPS, a major project of the MIT NEWDIGS initiative that advances the knowledge and practice of Precision Medicine by modernizing how we plan, produce, and use real-world evidence (RWE). We take a systems approach to enhancing the efficiency and scalability of real-world learning to ensure that the right drug therapies are delivered to the right patient at the right time. Our participatory design approach involves stakeholders in the system who hold the data, use the evidence, and—only together—have the power to ensure that healthcare is both patient-centered and economically sustainable.

ABOUT NEWDIGS

The NEW Drug Development ParadIGmS (NEWDIGS) Initiative at Tufts Medical Center is an international "think and do tank" dedicated to delivering more value faster to patients, in ways that work for all stakeholders. NEWDIGS designs, evaluates, and initiates advancements that are too complex and cross-cutting to be addressed by a single organization or market sector. Its members include global leaders from patient advocacy, payer organizations, biopharmaceutical companies, regulatory agencies, clinical care, academic research, and investment firms.

For more information, visit newdigs.tuftsmedicalcenter.org.

LEAPS

Learning Ecosystems Accelerator for Patient-centered, Sustainable innovation

PART II

Data Assessment & Risk Engineering

Healthcare data are inherently biased. Most scientists who work with these data are aware of the problem, yet their papers often fail to clearly acknowledge it. As a result, their work may risk perpetuating the issue further.

To address this issue, the LEAPS Project's Methods Innovation Team recommends that healthcare researchers assess their data sources and clearly present those sources' bias in scientific papers. Our Data Assessment framework for this field suggests that researchers compare their dataset's demographics to the real-world population impacted by a condition under study, in order to identify discrepancies and to assess their potential impact on the findings.

NEW FRAMEWORKS FOR RWE

Part II of a three-part series

PART I
LEAPS Methods Innovation Team proposes new frameworks for RWE

PART II
Data Assessment & Risk Engineering

PART III
Federated Machine Learning to improve real-world evidence generation

After that assessment, researchers can take the work further, using techniques that make their data more representative of the real-world patient population under study. NEWDIGS has evaluated several potential methods for this purpose, collected under a framework called Risk Engineering. These methods primarily fall into two categories: modifying data prior to conducting the analysis and combining multiple datasets to fill gaps.

THE PROBLEM

Healthcare datasets, like the system that produces them, tend to focus on certain populations at the expense of others. This happens through "exclusion cycles," as described in a 2022 paper in *Science*: medical datasets disproportionately include dominant groups at the expense of underrepresented ones, leading to biased analyses, next leading to results that don't apply to the underrepresented group, then leading patients from that group to distrust the medical system, and further biasing future datasets.[1] The data exclusion cycle mirrors a similar cycle in clinical care, as doctors' biases lead patients to withdraw.

Key takeaways

Healthcare datasets, like the system that produces them, tend to focus on certain populations at the expense of others.

The LEAPS Methods Innovation Team proposes a framework called Data Assessment and Risk Engineering (DARE).

In Data Assessment, healthcare data researchers should focus more heavily on evaluating their underlying data sources for potential bias.

Researchers should seek to mitigate bias through available Risk Engineering techniques.

Conclusions resulting from work informed by this framework will be more reflective of real-world populations impacted by the research.

Patient race is one common driver of exclusion, but the same process happens for other demographic factors, according to the team's research. For example, the team examined an analysis of lung cancer risk published in 2019, which utilized electronic health records. The study's authors did not include race in their analysis or discuss this demographic factor in their paper; in fact, their underlying dataset included a very small non-white patient population, compared to white patients.[2] As a result, the study's findings represent a predictive model that may not represent the true population impacted by lung cancer.

Health risk prediction models like this paper should disclose the differences between their dataset population and the actual population impacted by the condition under study. But articles frequently fail to clearly explain this limitation of their analysis, leading the biased results to potentially be taken back to the clinical setting—and utilized as a basis for further study.

To examine this problem, the Methods Innovation Team undertook a literature review of papers describing healthcare data analysis. A student researcher at NEWDIGS reviewed 20 papers from top journals published in the last five years, focusing on papers describing analysis with machine learning. In general, these papers had little mention of potential bias or resulting limitations, the review found. A paper from this sample might include a table showing the demographics of its underlying dataset but would fail to explain the implications of those demographics.

While the team reviewed a relatively small subset of papers, the common pattern suggests cause for concern. NEWDIGS additionally examined several articles that propose data analysis models for specific subfields of healthcare research, such as Emergency Medical Services (EMS) data. While these models may be helpful within their niches, it's difficult to generalize their approaches to other fields.

Healthcare data scientists may not be able to address bias issues with an underlying dataset under study, because they are limited by the organizations providing data. For example, a researcher using a hospital's data may not have access to health records beyond those the hospital has deidentified.

As a result, it is up to researchers to analyze their data's bias and clearly disclose the results. By supporting this work—and improving upon past disregard for this issue—medical journals could encourage real-world evidence results that more closely match the real-world population.

DATA ASSESSMENT

In the Methods Innovation Team's proposed Data Assessment framework, a researcher would compare their source dataset to the population impacted by their analysis, utilizing a comprehensive dataset such as those offered by the CDC and other government agencies. In this comparison, researchers may examine demographic and geographic factors, such as race, ethnicity, age, gender, urban vs. rural, and so on. For each factor, researchers should determine the gap between their dataset and the true population.

Following from this assessment, healthcare data science papers should map out the demographic factors of their source datasets, overlaid with actual incidence rates for the condition under study. This mapping could take the form of a typical "Table 1," but should include details about the overall population in addition to the sample under study. As a result, the paper's audience should gain a clearer understanding of the data's advantages a nd limitations.

There are some health conditions for which detailed prevalence data in the real-world setting may not be available, such as diseases that impact very small numbers of patients or those that emerged recently as research topics. For research on such conditions, the team recommends still describing the patient dataset under study as completely as possible and looking out for potential gaps, such as a geographic region that may be missing.

Additionally, even government datasets may have underlying biases or may underrepresent key populations. By continuing to focus on improving their data's representation, researchers from academic and private sectors can work with government data scientists to improve these sources over time.

The team recommends that researchers begin incorporating this type of Data Assessment into their studies and reporting the results in their papers. Researchers could include these assessments as part of the methods section, as well as discussing their results in conclusions (along with other potential limitations of their work). Journal editors may even consider including Data Assessment as a requirement for healthcare data analysis submissions.

If this assessment becomes a priority for data science teams, it may inspire further research to address the biases and gaps illuminated. For example, if one paper's Data Assessment finds that its source dataset fails to include a certain underrepresented population that is disproportionately impacted by the health condition under study, for their next paper, the researchers could seek out a new dataset that specifically represents that population. Researchers could also consider analyzing multiple datasets from different sources simultaneously, with techniques such as federated machine learning.

RISK ENGINEERING

After assessing their data, researchers can take the work further, using techniques that make their data more representative of the real-world patient population under study. The Methods Innovation Team refers to these techniques under a framework called Risk Engineering.

TECHNICAL DATA ADJUSTMENT

Using technical or statistical methods can mitigate a dataset's bias before starting analysis.[3] This could include suppressing a particular demographic attribute to prevent a model from incorporating it, changing data labels for some objects based on statistical ranking, or weighing different patient groups in the analysis to better reflect real-world demographics, according to a paper published in Knowledge and Information Systems in 2011.[4]

Options include:

- Data collection and preparation methods: If researchers have control over their data from initial collection steps, they can compile datasets that are as diverse and large as possible, in order to represent all potential patient groups. During this process, data engineers should also carefully monitor any potential errors in software that may be used to classify different patient groups.
- Data preprocessing methods: When starting with a biased dataset, researchers may use processing solutions to remove elements of discrimination before an algorithm is strained upon the data. These solutions may involve fully removing specific demographic attributes from the data, changing labels of some objects in the dataset to reduce potential discrimination, assigning weights to different groups in the dataset so that the data population more closely matches the real-world population, and taking a sample from a dataset that represents the real-world population (potentially duplicating some values and taking out others).[5]
- Analyzing records in clusters: In the analysis phase of a project, researchers may use techniques that separate health records into specific groups based on their demographic attributes, either manually or with a machine-learning technique called clustering. When records are analyzed in groups, the results will be more specifically applicable to those groups rather than providing an overall pattern that might not describe less-represented minorities in the population.

These data adjustment methods may be helpful for a single analysis when only biased data are available. But they are only partial solutions that may not fully disrupt cycles of exclusion. Even the biggest datasets don't necessarily offer useful results because they can simply be biased on larger scales, according to the team's research. In such cases, data scientists must seek more comprehensive, inclusive sources.

COMPLEMENTARY DATASET EXPANSION

The Methods Innovation Team recommends that researchers seek to fill gaps in their original datasets by appending other datasets that include more diverse populations. This could mean building a master dataset that incorporates records from different healthcare systems; it could also mean using a machine learning approach to analyze several different datasets through the same algorithm, while keeping the source data siloed. For example, if the population in a researcher's source dataset is 90% white, they may address this bias by pulling in additional datasets that better represent non-white patient groups.

Federated machine learning (FML) offers one strategy for researchers seeking to reduce bias in their results while simultaneously addressing the challenge of data fractured in different source locations that can't easily be combined.[6] In the FML technique, researchers analyze several different datasets with diverse populations; they can even analyze different kinds of data.

The FML approach allows healthcare researchers to send algorithms to different datasets, analyze them separately, and then bring the results together at the end. In addition to the opportunity for reducing bias, FML is advantageous due to its decentralized nature: researchers can evaluate multiple, diverse datasets without combining or standardizing them. This process circumvents privacy concerns about sharing data, as people working on different datasets can contribute to the same analysis without sending records back and forth. It also circumvents the challenges that often occur when researchers try to bring together datasets that aren't interoperable.

For additional details about FML, refer to Part III in this research brief series.

Put together, the Data Assessment and Risk Engineering (DARE) framework provides data scientists with a mechanism to interrogate their sources and push for more inclusive datasets across the healthcare field. Data analysis papers, particularly papers focusing on analytical methods, offer an opportunity for researchers to create more awareness about bias in healthcare datasets and to seek new strategies for mitigating that bias.

ACKNOWLEDGEMENT

NEWDIGS wishes to thank and acknowledge the leaders and members of the LEAPS Methods Innovation team and subteams for their contribution to this research brief and to the LEAPS Project overall.

REFERENCES

1. Bracic A, Callier SL, Price WN. Exclusion cycles: Reinforcing disparities in medicine. Science. 2022;377(6611):1158-1160. doi:doi:10.1126/science.abo2788
2. Data Study Group team. (2020). Data Study Group Final Report: Roche. Zenodo. Accessed May 1, 2023. https://doi.org/10.5281/zenodo.3876989
3. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. Communications Medicine. 2021/08/23 2021;1(1):25. doi:10.1038/s43856-021-00028-w
4. Kamiran F, Calders T. Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems. 2012/10/01 2012;33(1):1-33. doi:10.1007/s10115-011-0463-8
5. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. 2022/06/01/ 2022;3(1):91-99. https://doi.org/10.1016/j.gltp.2022.04.020
6. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y). Oct 8 2021;2(10):100347. doi:10.1016/j.patter.2021.100347

LEAPS

Learning Ecosystems Accelerator for Patient-centered, Sustainable innovation

PART III

Federated Machine Learning to improve real-world evidence generation

Federated Machine Learning (FML) is a distributed approach for asking questions of data. Its potential to pool results—but not records—from different datasets offers opportunities to address long-standing issues in real-world evidence generation and to produce clinically meaningful hypotheses from different types of data.

In this approach, rather than compiling a large number of records in one place for a unified analysis, researchers may ask the same question of multiple smaller datasets and then pool the results. Only the analytical models and their results are shared in a central location, retaining privacy and security for all participating data providers. The decentralized approach also allows researchers to pull in a diverse group of datasets, including those that serve less-resourced populations and may not meet standards for more unified analysis.

NEW FRAMEWORKS FOR RWE

Part III of a three-part series

PART I

LEAPS Methods Innovation Team proposes new frameworks for RWE

PART II

Data Assessment & Risk Engineering

PART III

Federated Machine Learning to improve real-world evidence generation

INTRODUCTION TO FML

Following months of research and discussion about FML, the LEAPS Project's Methods Innovation Team argues that this technique has the potential to simultaneously address two major challenges in the healthcare data space:

- **Data silos:** Within the decentralized U.S. healthcare system, every organization tracks health information independently. Two different hospitals, two different insurance companies, or two different public health agencies may not be able to share the most basic data, such as immunization records, in an interoperable manner. And even if they could share data, many organizations are hesitant to do so due to the business potential of proprietary health information. As a result, it's difficult for data scientists to compile

Key takeaways

Federated Machine Learning (FML), which connects disparate datasets for analysis, is a promising option for generating hypotheses from real-world healthcare data.

The LEAPS Methods Innovation Team argues that FML has the potential to simultaneously address two major challenges:

- Data silos within the decentralized U.S. healthcare system
- Biased datasets that overrepresent patients who receive easier access to care at the expense of minority groups.

While some pilot projects have shown how FML may be used in healthcare research, more research will be needed to fully understand its utility for real-world evidence generation.

the large, diverse datasets that are often necessary to study complex medical questions. Barriers to data sharing also may benefit larger and more resourced health organizations, which have the capacity to analyze their own records for improving care, while smaller organizations (such as health centers serving marginalized populations) are less likely to do so.

- **Biased data:** Across the healthcare system, records disproportionately include patients who receive easier access to care—such as white people, insured people, and people in higher income brackets—at the expense of minority groups who are less prioritized. This bias in health records contributes to biased analyses, which in turn lead to clinical results and medical guidance that are more attuned to well-resourced patients than to minority patients. Those patients may find doctors are less capable of caring for them—leading them to withdraw from the medical system, and further biasing future health records datasets. Such "exclusion cycles" persist across the system, connecting gaps in clinical practice to broader health disparities. To inform better care for these minority patients, the healthcare data field requires more diverse data and novel analysis frameworks that address bias rather than perpetuating it. Refer to Part II in this research brief series for more details on this issue.

Machine learning scientists at Google first developed FML in order to analyze data from smartphones while preserving the privacy of every individual phone's owner. In their framework, described in a 2016 paper, the scientists deliver an algorithm to every smartphone which computes a specific user research problem, such as how to improve the phone's ability to predict a user's next word while they're typing. On every phone, the algorithm tests its problem and sends the results—not the underlying data, just the results—back to a central server. The central server then compiles results from many smartphones and identifies overarching conclusions, informing further improvements to the overall network.[1]

While FML has primarily been a research interest for computer scientists, healthcare data experts have devoted more attention to the technique in recent years. There are a variety of potential applications, as discussed in a review paper published in March 2023.[2] Academic researchers have begun testing the technique, as have startups such as the AI biotech company Owkin.[3] The team views the production of novel hypotheses from real-world evidence and combining disparate datasets to address bias as two particularly promising areas for FML.

TEAM FINDINGS

The Methods Innovation Team produced four frameworks based on its research into FML and bias issues. The first framework (described in Part II of this research brief series) describes overall strategies for healthcare data scientists to evaluate and address biases in their data sources. The other three frameworks, compiled by the three subteams examining different aspects of FML, focus on considerations for using FML techniques in this field.

The team argues that FML has the capacity to address two analysis challenges at once. This technique addresses the data silos challenge by enabling organizations to collaborate on analyses without sharing their proprietary data with each other. It addresses the bias challenge by enabling researchers to pull together data sources that represent diverse groups of patients, rather than relying on a single source that may be biased.

With FML, researchers can utilize datasets that represent a broader diversity of patients, deliberately including groups that may otherwise not be included in research such as those in rural or lower-income communities. They can also utilize diverse data types, including electronic health records, administrative claims, wearables, patient-reported outcomes, social media posts, and more.

In essence, FML may enable healthcare data scientists to scale up a research question from a single dataset to several datasets—without going through the logistical steps needed to build a central data lake. This tool may be the next iteration of meta-analysis, a long-standing technique used to compile results from different studies into broadly-applicable findings.

If further studies demonstrate success with the FML approach, this tool could be valuable for a variety of medical research applications, such as drug discovery, clinical trial design, developing diagnostic tools, and informing healthcare policy. However, relatively few research projects have truly tested FML's potential in the healthcare data field so far. Additional study is needed, and the researchers taking on these projects will face many challenges.

CHALLENGES AND LIMITATIONS FOR FML

The Methods Innovation Team has identified several key challenges that future research should address:

- Motivating data providers to participate: Healthcare organizations with proprietary datasets, particularly larger companies that utilize their data as a revenue stream, may be less likely to participate in FML projects. Additional research is needed to motivate data providers to join these efforts.

- Smaller test projects, such as pilot studies at Owkin and an upcoming study at Merrimack College, may serve as proof-of-concept projects to demonstrate FML's potential value to these organizations.[4] For example, Owkin's Substra software has supported MELLODDY, a collaboration between 10 major pharmaceutical companies on drug discovery research.[5]
- Another potential strategy for motivating data providers to join FML projects may be prioritizing transparency in all communications, according to the team's research. Team members recommend that scientists soliciting data for FML projects should be clear about both the potential gains and potential risks for different data providers who may participate.

- Data harmonization: One great advantage of FML is that it may incorporate different datasets that cannot be standardized. Without standardizing, however, researchers may run into the question of how to interpret results from different datasets in a unified manner; they may be comparing apples to oranges, as the saying goes.
  - In one pilot project at Owkin, the team addressed this challenge through collaboration: scientists at different hospitals coded tumor images in similar ways and data engineers worked together on analysis methods.[6]
  - Another, less resource-intensive strategy for harmonizing a FML project may be to require data dictionaries or similar documentation from each data provider. Researchers could use the documentation to select similar metrics from each dataset for analysis and set up a unified framework for interpreting results.
  - Highly structured datasets, such as those adhering to Common Data Models, could still participate in FML. But in order to broaden the horizon of evidence used in hypothesis generation, the team proposes that it's also important to use datasets that may come from less-resourced organizations that are unable to adhere to Common Data Models.
- Resource and operational needs: FML projects may require extensive computer programming resources, as well as staff scientists with expertise in a variety of healthcare data-related fields to conduct analysis and interpret results (i.e., data engineers to program machine learning models are needed, and so are clinicians and patient representatives who can help to interpret results' clinical significance).
  - These operational needs could represent barriers for smaller healthcare organizations that are interested in participating in FML. Researchers running FML projects should consider how to assist such organizations that may have fewer resources, yet still have valuable data to offer a project.
- Privacy and security concerns: Privacy concerns around source datasets should be minimal in an FML project because data are not shared outside of their host institution. However, additional steps may be needed to ensure the security of the algorithms used in the analysis and of the results from each dataset, which will be compiled in a centralized location.

NEXT STEPS FOR RESEARCH

The Methods Innovation Team has concluded its investigation into FML and related healthcare data analysis questions with this package, which NEWDIGS hopes can be a valuable resource for other organizations seeking to take FML from the theoretical research phase into real-world test projects.

One such project is now in its early stages at Merrimack College, led by Fotios Kokkotos, former Director of Data Science at NEWDIGS and member of the Methods Innovation Team. The FML research described above has informed this project, which will serve as a pilot study based on these findings.

The project at Merrimack aims to simulate FML with test data available on Owkin's Substra, an open-source software designed for FML projects analyzing healthcare data.[4] Kokkotos and his colleagues chose to use Substra after reviewing other available software (such as Microsoft's FLUTE)[7] and finding it the most user-friendly option.

Working with public data on Substra, the Merrimack team seeks to identify the probability that a patient may develop heart disease. The team will simulate FML by splitting a single data source into multiple datasets, running the same analytical algorithm on these smaller datasets in isolation, then combining the results. Upon combining the results, the team will generate a singular predictive algorithm informed by all participating datasets.

As the public datasets used in this study are specifically designed for FML tests, the researchers will not face logistical issues that may occur with FML analysis of real-world data, such as missing values and a lack of harmonization across participating datasets. This test aims to examine the overall principles and process for FML; logistical issues with data analysis will be a topic for future projects.

Consortium members interested in connecting with Kokkotos and his colleagues about this test project may reach out to NEWDIGS staff at tuftsmcnewdigs@tuftsmedicine.org.

In addition to the project at Merrimack, our team intends to follow further work from Owkin, the AI startup Rhino Health, and other companies advancing FML research.[8] These projects will advance FML as a tool for healthcare data analysis by demonstrating this technique's potential while revealing further challenges. Like other machine learning techniques, FML will push this field to the next phase of real-world evidence generation, taking full advantage of the most diverse and inclusive datasets available.

ACKNOWLEDGEMENT

NEWDIGS wishes to thank and acknowledge the leaders and members of the LEAPS Methods Innovation team and subteams for their contribution to this research brief and to the LEAPS Project overall.

REFERENCES

1. McMahan HB, Moore E, Ramage D, Aguera y Arcas B. Federated Learning of Deep Networks using Model Averaging. arXiv: Cornell University; 2016. p. 11.
2. Dasaradharami Reddy K, Gadekallu TR. A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics. Comput Intell Neurosci. 2023;2023:8393990. doi:10.1155/2023/8393990
3. Owkin AI: The AI biotech company website. Accessed May 4, 2023. https://owkin.com
4. Substra: The most proven federated learning software designed for healthcare research. Accessed May 5, 2023. https://owkin.com/substra
5. Largest-ever pharma industry artificial intelligence collaboration utilizing a federated model holds potential benefits for drug discovery without compromising IP, landmark project finds. 2022. Accessed May 1, 2023. https://www.melloddy.eu/y3announcement
6. Ogier du Terrail J, Leopold A, Joly C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer.
7. Dimitriadis D, Garcia MH, Madrigal Diaz DE, Manoel A, Sim R. FLUTE: A scalable federated learning simulation platform. Microsoft Research Blog blog. May 2, 2022. https://www.microsoft.com/en-us/research/blog/flute-a-scalable-federated-learning-simulation-platform/
8. Rhino Health: Healthcare AI with Federated Learning and Edge Computing. Accessed May 6, 2023. https://www.rhinohealth.com/