Check for updates

Approaches to identify scenarios for data science implementations within healthcare settings: recommendations based on experiences at multiple academic institutions

Lillian Sung[1]*, Michael Brudno[2], Michael C. W. Caesar[3], Amol A. Verma[4], Brad Buchsbaum[5], Ravi Retnakaran[6], Vasily Giannakeas[7], Azadeh Kushki[8], Gary D. Bader[9], Helen Lasthiotakis[10], Muhammad Mamdani[11] and Lisa Strug[10]

[1]Department of Paediatrics, The Hospital for Sick Children, Institute of Health Policy Management & Evaluation, University of Toronto, Toronto, ON, Canada, [2]Department of Computer Science, Vector Institute for Artificial Intelligence, University Health Network, University of Toronto, Toronto, ON, Canada, [3]Institute of Health Policy Management & Evaluation, University Health Network, University of Toronto, Toronto, ON, Canada, [4]Department of Medicine, Department of Laboratory Medicine and Pathobiology, and Institution of Health Policy Management & Evaluation; St. Michael's Hospital, Unity Health Toronto, University of Toronto, Toronto, ON, Canada, [5]Department of Psychology, Rotman Research Institute, Baycrest Centre, University of Toronto, Toronto, ON, Canada, [6]Division of Endocrinology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, [7]Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada, [8]Institute of Biomedical Engineering, Holland Bloorview Kids Rehabilitation Hospital, University of Toronto, Toronto, ON, Canada, [9]Department of Molecular Genetics, Temerty Faculty of Medicine, Toronto, ON, Canada, [10]Data Sciences Institute, University of Toronto, Toronto, ON, Canada, [11]Temerty Faculty of Medicine, Centre for Artificial Intelligence Education and Research in Medicine, Unity Health Toronto, University of Toronto, Toronto, ON, Canada

Objectives: To describe successful and unsuccessful approaches to identify scenarios for data science implementations within healthcare settings and to provide recommendations for future scenario identification procedures.

Materials and methods: Representatives from seven Toronto academic healthcare institutions participated in a one-day workshop. Each institution was asked to provide an introduction to their clinical data science program and to provide an example of a successful and unsuccessful approach to scenario identification at their institution. Using content analysis, common observations were summarized.

Results: Observations were coalesced to idea generation and value proposition, prioritization, approval and champions. Successful experiences included promoting a portfolio of ideas, articulating value proposition, ensuring alignment with organization priorities, ensuring approvers can adjudicate feasibility and identifying champions willing to take ownership over the projects.

Conclusion: Based on academic healthcare data science program experiences, we provided recommendations for approaches to identify scenarios for data science implementations within healthcare settings.

KEYWORDS

machine learning, data sciences, healthcare, use case, scenario

Background

Healthcare institutions rely upon data to influence decision making at the organizational level and to improve patient outcomes (1). Examples range from simple reports to rule- or machine learning- (ML) based algorithms. Constraints to using data to drive decision making include, but are not limited to, challenges with data access, security and privacy restrictions, need to ensure patient safety and address ethical aspects, fragmented data ecosystems and funding/resource limitations. These remain open challenges in implementing data science approaches. It is important to identify solutions to these challenges to accelerate data science-directed advancements in healthcare.

Scenario identification (or "use case" identification) is a specific challenge for data sciences in the clinical context because of the need to curate solutions and optimize use of scarce resources. Scenario refers to both the operational or clinical problem in addition to the specific setting in which the data will be applied. For example, sepsis is an example of a clinical problem whereas sepsis prediction during admission to a general ward is an example of a scenario.

There are several papers that have described frameworks for healthcare institutions to develop and implement data science solutions, including ML (2–8). However, most propose solutions that are generated from a theoretical or a single institution perspective. In contrast, we hypothesized that building upon the experiences of several academic health sciences institutions implementing data science solutions would reveal more general patterns associated with successful implementations. We define success as scenarios leading to institutional adoption, favorable clinical or operational impact and continued implementation. Successful scenarios include early warning systems for clinical deterioration (9, 10), mortality (11) prediction, and sepsis (12) prediction as examples. Our objective was to describe successful and unsuccessful approaches to identify scenarios for data science implementations within healthcare settings and to provide recommendations for future scenario identification procedures.

Materials and methods

The Data Sciences Institute (DSI) is a multi-divisional, tri-campus and multidisciplinary hub for data science activity at the University of Toronto. The DSI mission is to accelerate the impact of data sciences across disciplines to address pressing societal questions and drive positive social change. They define data sciences as the science of collecting, manipulating, storing, visualizing, learning from and extracting useful information from data in a reproducible, fair and ethical way. DSI partners with seven healthcare institutions in Toronto and collaborates with the Temerty Faculty of Medicine Centre for Artificial Intelligence Education and Research in Medicine (T-CAIREM). T-CAIREM supports a community of over 1,500 members focused on research, education and infrastructure for artificial intelligence in health care.

The DSI and T-CAIREM co-hosted a one-day in-person workshop with representation from the following seven DSI-affiliated healthcare institutions: Baycrest Centre, Holland Bloorview Kids Rehabilitation Hospital, Mount Sinai Hospital, The Hospital for Sick Children, Unity Health Toronto, University Health Network and Women's College Hospital. The workshop was held at the DSI offices on April 5, 2024. Each institution was asked to provide a brief introduction to their data science program and to provide an example of a successful and unsuccessful approach to scenario identification at their institution. Each presentation was followed by a question and answer period.

Once all presentations were completed, attendees first outlined some general observations about the potential utility and limitations of experience reflections from this group. Next, we outlined general considerations required to develop and implement data science solutions in healthcare, where the focus of the present meeting was to discuss scenario identification approaches. Finally, the discussion centered on common observations across experiences with successful and unsuccessful approaches to identify scenarios for data science implementations.

Notes were taken during the meeting by one author (HL). Using the presentations themselves and the meeting notes, content analysis (13) was performed by a single reviewer (LS). These results were evaluated and revised by a second reviewer (HL). If discrepancies were identified, the two reviewers met to come to consensus. Summarized results were then reviewed and confirmed by all authors.

Results

We observed that sharing of experiences was useful for learning and collaborating, with the potential to accelerate each hospital's program by leveraging insights made at other institutions. We also observed that a synthesis of approaches that are consistently successful and those that are consistently unsuccessful is likely to be generalizable across many settings. High-level approaches are more likely to be generalizable rather than the specific plans to implement, which will typically be driven by local circumstances.

Table 1 describes the general considerations required to develop and implement data science solutions in healthcare, where scenario identification was the focus of the workshop. Table 2 describes common observations with successful and unsuccessful approaches to identify scenarios. Examples were

TABLE 1 General considerations required to implement data science solutions in healthcare.

Considerations to implement data science solutions
Scenario identification: idea generation and value proposition, prioritization and approval[a]
Data acquisition: access and governance
Data infrastructure including ongoing support and quality assurance
Solution design and/or procurement, including model development
Pre- and post- deployment planning and monitoring including education, change management including implementation science considerations and maintenance
Evaluation of impact on patient or operational outcomes and return on investment

[a]Focus for this manuscript.

TABLE 2 Common observations with successful and unsuccessful approaches to identify scenarios for data science implementations.

Theme	Success examples	Unsuccessful examples	Recommendations
Idea generation and value proposition	Promote a portfolio of ideas, allowing many to propose different scenarios and solutions Clear value proposition Using data-driven approaches to problem identification	Multiple groups proposing same scenario and solution working in silos Operational mindset that does not permit innovation or change	Require approval from target users who can identify and harmonize duplicate efforts Endorsement by both clinical and operational leadership Allow some risk tolerance
Prioritization	Aligned with organizational priorities Successfully solves a real operational or clinical problem	Focus on research without articulation of clinical or operational benefit	Articulate strategic priority Align research energy and culture with problems prioritized by clinicians and leaders
Approval	Single or small group of individuals with knowledge and commitment to evaluate proposal	Large executive committee that may lack knowledge to identify promising and feasible solutions or may not invest energy to understand proposal	Small group Can judge problem importance, feasibility, return on investment Need prioritization framework
Champions	Individuals at senior level at decision-making table as well as implementation champions Willing to take ownership Knowledgeable	No one willing to "own" solution or be accountable for poor outcomes	Individual or group who takes responsibility for project success or failure Requires representation from target users

shown by idea generation and value proposition, prioritization, approval and champions. Table 2 also provides recommendations related to each of these areas.

Idea generation and value proposition

In the context of scenario identification, we recommend that there is a mechanism, within an organization, that allows many ideas to be proposed for downstream prioritization. Ideas can be generated and shaped from a wide variety of sources spanning operational and clinical needs to an individual clinician's or researcher's ideas. To address unmet needs, an operational perspective that prioritizes aligned innovation or improvement is required. We also recognized the importance of balancing risk, and that some risk tolerance should be promoted as long as it is monitored and it is balanced against potential benefit. A critical component of scenario identification is clear articulation of the value proposition and anticipated return on investment for scenario development and deployment. In identifying scenarios, it is important to use local data to ensure problems or outcomes are sufficiently common to justify institutional investment.

An unsuccessful example was when the same scenario was being proposed by multiple groups for the same target user, without adequate engagement of the actual target user. These different groups might represent different operational, clinical or research perspectives. Identifying duplicative siloed efforts is important; target user may be in the best position to recognize and reconcile them.

Prioritization

It is important to distinguish between approval to explore in the research phase vs. approval to implement. We suggest that if there are no or limited resource implications, researchers should be encouraged to innovate and identify solutions as freely as possible, while adhering to regulatory and privacy frameworks. Prioritization should occur prior to the point of clinical implementation, or when

there are implications for institutional resources or impact. Ultimately, the prioritization of initiatives will be driven by a balance between the anticipated benefits and the resources required to develop, deploy and maintain the solution.

Prioritization is a critical process that should consider limited resources, organizational strategy, prioritized improvement efforts and overall cost-benefit. While it is important that prioritization aligns with operational and clinical needs, how to actually prioritize specific projects and who performs the prioritization will likely differ between institutions. Institutions should promote alignment of research energy and culture with problems prioritized by clinicians and leaders where feasible. The specific roles of relevant clinical and operational leaders will vary across institutions.

Approval

The mechanism to approve scenarios was particularly heterogeneous. The group agreed that the approval process must take into account what operational or clinical scenarios are important to the organization as well as determine the feasibility of a data science approach. The approval process must also be able to assess likely return on investment. Successful examples included single or a small number of knowledgeable individuals to approve scenario identification. An unsuccessful example was when large executive groups made the decision of which data science solutions were to proceed as they may not have the requisite knowledge to be able to judge clinical impact or data science feasibility. With large groups, there may be few members who engage in the discussion or make the effort to fully understand the proposal, its feasibility and likelihood for success.

Champions

It is important that there is an individual or group of individuals who are willing to take ownership over the project implementation and to be accountable for the measured success

and failure. This ownership must occur throughout the various stages of the project. The specific professional characteristics of champions will vary depending on the specific scenario and across institutions. These individuals are likely to be at both the clinical and executive levels. In contrast, projects were not successful when no one was willing to take responsibility for them.

Discussion

We summarize successful and unsuccessful approaches to identify scenarios for data science implementations in healthcare. Observations were coalesced to idea generation and value proposition, prioritization, approval and champions. Successful experiences included promoting a portfolio of ideas, articulating value proposition, ensuring alignment with organization priorities, ensuring approvers can adjudicate feasibility and identifying champions willing to take ownership over the projects.

While these observations arose out of real-world experiences, we recognize that there are many appropriate ways to generate and approve scenarios and our observations are unlikely to be applicable to all settings. Nonetheless, despite different institutions and patient populations (e.g., spanning neonatal to geriatric), observed commonalities support some generalizability. We hope that sharing our successes and challenges can contribute to the broader discourse around data-driven decision making in healthcare.

Scenario identification, prioritization and approval is only one step in data science implementations in healthcare. Future workshops could focus on identifying shared experiences related to other aspects such as data access and governance as an example. Another important aspect that warrants further exploration is approaches that promote successful implementations and sustainability post deployment.

The strength of this report lies in the number and heterogeneity of contributing organizations. However, this report is limited as all institutions belonged to the same umbrella networks, namely DSI and T-CAIREM. Factors that influence decisions at academic centers in Toronto may be different than in other places. In particular, these findings may not be generalizable to very different contexts or countries. Further, despite the number and heterogeneity of organization involved, they do not cover all possible types of healthcare contexts and thus continued analysis of diverse settings will be beneficial.

In summary, based on academic healthcare data science program experiences, we provided recommendations for approaches to identify scenarios for data science implementations within healthcare settings. Future efforts should focus on other requirements to successfully deploy data sciences solutions in healthcare.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

LS: Writing – original draft, Writing – review & editing. MB: Writing – review & editing. MC: Writing – review & editing. AV: Writing – review & editing. BB: Writing – review & editing. RR: Writing – review & editing. VG: Writing – review & editing. AK: Writing – review & editing. GB: Writing – review & editing. HL: Writing – original draft, Writing – review & editing. MM: Writing – review & editing. LS: Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Gagalova KK, Leon Elizalde MA, Portales-Casamar E, Görges M. What you need to know before implementing a clinical research data warehouse: comparative review of integrated data repositories in health care institutions. *JMIR Form Res*. (2020) 4(8): e17687. doi: 10.2196/17687

2. Harris S, Bonnici T, Keen T, Lilaonitkul W, White MJ, Swanepoel N. Clinical deployment environments: five pillars of translational machine learning for health. *Front Digit Health*. (2022) 4:939292. doi: 10.3389/fdgth.2022.939292

3. Jeong H, Kamaleswaran R. Pivotal challenges in artificial intelligence and machine learning applications for neonatal care. *Semin Fetal Neonatal Med*. (2022) 27(5):101393. doi: 10.1016/j.siny.2022.101393

4. Liao F, Adelaine S, Afshar M, Patterson BW. Governance of clinical AI applications to facilitate safe and equitable deployment in a large health system: key elements and early successes. *Front Digit Health*. (2022) 4:931439. doi: 10.3389/fdgth.2022.931439

5. Novak LL, Russell RG, Garvey K, Patel M, Thomas Craig KJ, Snowdon J, et al. Clinical use of artificial intelligence requires AI-capable organizations. *JAMIA Open*. (2023) 6(2):ooad028. doi: 10.1093/jamiaopen/ooad028

6. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. (2022) 28(1):31–8. doi: 10.1038/s41591-021-01614-0

7. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. (2019) 25(9):1337–40. doi: 10.1038/s41591-019-0548-6

8. Verma AA, Trbovich P, Mamdani M, Shojania KG. Grand rounds in methodology: key considerations for implementing machine learning solutions in quality improvement initiatives. *BMJ Qual Saf*. (2024) 33(2):121–31. doi: 10.1136/bmjqs-2022-015713

9. Verma AA, Stukel TA, Colacci M, Bell S, Ailon J, Friedrich JO, et al. Clinical evaluation of a machine learning-based early warning system for patient deterioration. *CMAJ*. (2024) 196(30):E1027–37. doi: 10.1503/cmaj.240132

10. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med*. (2020) 383(20):1951–60. doi: 10.1056/NEJMsa2001090

11. Lu J, Sattler A, Wang S, Khaki AR, Callahan A, Fleming S, et al. Considerations in the reliability and fairness audits of predictive models for advance care planning. *Front Digit Health*. (2022) 4. doi: 10.3389/fdgth.2022.943768

12. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med Inform*. (2020) 8(7):e15182. doi: 10.2196/15182

13. Columbia University Mailman School of Public Health. Content Analysis. Available online at: https://www.publichealth.columbia.edu/research/population-health-methods/content-analysis (accessed May 18, 2024)