

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

Lillian Sung, Michael Brudno, Michael C. W Caesar, Amol Verma, Brad Buchsbaum, Ravi Retnakaran, Vasily Giannakeas, Azadeh Kushki, Gary Bader, Helen Lasthiotakis, Muhammad Mamdani, Lisa Strug

Submitted to: JMIR Medical Education  
on: June 19, 2024

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## ***Table of Contents***

---

<b>Original Manuscript.....</b>	<b>5</b>
---------------------------------	----------

Preprint  
JMIR Publications

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

Lillian Sung<sup>1, 2</sup> MD, PhD; Michael Brudno<sup>3, 4</sup> PhD; Michael C. W Caesar<sup>2, 5</sup> MBA; Amol Verma<sup>2, 6, 7</sup> MD, MPhil; Brad Buchsbaum<sup>8, 9</sup> PhD; Ravi Retnakaran<sup>10, 11</sup> MD; Vasily Giannakeas<sup>12</sup> PhD; Azadeh Kushki<sup>13, 14</sup> PhD; Gary Bader<sup>15</sup> PhD; Helen Lasthiotakis<sup>16</sup> PhD; Muhammad Mamdani<sup>17, 18</sup> PharmD, MA, MPH; Lisa Strug<sup>16</sup> PhD

<sup>1</sup>Department of Paediatrics The Hospital for Sick Children Toronto CA

<sup>2</sup>Institute of Health Policy Management & Evaluation University of Toronto Toronto CA

<sup>3</sup>Department of Computer Science University of Toronto Toronto CA

<sup>4</sup>Vector Institute for Artificial Intelligence University Health Network Toronto CA

<sup>5</sup>University Health Network Toronto CA

<sup>6</sup>Department of Medicine, Department of Laboratory Medicine and Pathobiology, and Institution of Health Policy Management & Evaluation University of Toronto Toronto CA

<sup>7</sup>St. Michael's Hospital Unity Health Toronto Toronto CA

<sup>8</sup>Department of Psychology University of Toronto Toronto CA

<sup>9</sup>Rotman Research Institute Baycrest Centre Toronto CA

<sup>10</sup>Division of Endocrinology Lunenfeld-Tanenbaum Research Institute Toronto CA

<sup>11</sup>Mount Sinai Hospital Toronto CA

<sup>12</sup>Women's College Research Institute Women's College Hospital Toronto CA

<sup>13</sup>Institute of Biomedical Engineering University of Toronto Toronto CA

<sup>14</sup>Holland Bloorview Kids Rehabilitation Hospital Toronto CA

<sup>15</sup>Department of Molecular Genetics Temerty Faculty of Medicine Toronto CA

<sup>16</sup>Data Sciences Institute University of Toronto Toronto CA

<sup>17</sup>University of Toronto Temerty Faculty of Medicine Centre for Artificial Intelligence Education and Research in Medicine University of Toronto Toronto CA

<sup>18</sup>Unity Health Toronto Toronto CA

## Corresponding Author:

Lillian Sung MD, PhD  
Department of Paediatrics  
The Hospital for Sick Children  
555 University Avenue  
Toronto  
CA

## Abstract

**Background:** Data sciences solutions such as artificial intelligence are increasing. A common challenge is identifying appropriate scenarios or “use cases” for data sciences implementation. Implementation frameworks are primarily generated from a theoretical or single institution based perspectives, highlighting the need for a multi-institutional experience to reveal patterns within successful implementations.

**Objective:** 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.

**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.

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

(JMIR Preprints 19/06/2024:63397)

DOI: <https://doi.org/10.2196/preprints.63397>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

## Original Manuscript

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

**Authors:** <sup>1</sup>Lillian Sung MD, PhD, <sup>2</sup>Michael Brudno PhD, <sup>3</sup>Michael C. W. Caesar MBA, <sup>4</sup>Amol A. Verma MD, MPhil, <sup>5</sup>Brad Buchsbaum PhD, <sup>6</sup>Ravi Retnakaran MD, <sup>7</sup>Vasily Giannakeas PhD, <sup>8</sup>Azadeh Kushki PhD, <sup>9</sup>Gary D. Bader PhD, <sup>10</sup>Helen Lasthiotakis PhD, <sup>11</sup>Muhammad Mamdani PharmD, MA, MPH, <sup>12</sup>Lisa Strug PhD

## Affiliations:

<sup>1</sup>Department of Paediatrics, The Hospital for Sick Children; Institute of Health Policy Management & Evaluation, University of Toronto, Toronto, Canada

<sup>2</sup>Department of Computer Science, University of Toronto; Vector Institute for Artificial Intelligence; University Health Network, Toronto, Canada

<sup>3</sup>Institute of Health Policy Management & Evaluation, University of Toronto; University Health Network; Toronto, Canada

<sup>4</sup>Department of Medicine, Department of Laboratory Medicine and Pathobiology, and Institution of Health Policy Management & Evaluation, University of Toronto; St. Michael's Hospital, Unity Health Toronto, Toronto, Canada

<sup>5</sup>Department of Psychology, University of Toronto; Rotman Research Institute, Baycrest Centre, Toronto, Canada

<sup>6</sup>Division of Endocrinology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada

<sup>7</sup>Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada

<sup>8</sup>Institute of Biomedical Engineering, University of Toronto; Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada

<sup>9</sup>Department of Molecular Genetics, Temerty Faculty of Medicine, Toronto, ON, Canada

<sup>10</sup>Data Sciences Institute, University of Toronto, Toronto, Canada

<sup>11</sup>University of Toronto Temerty Faculty of Medicine Centre for Artificial Intelligence Education and Research in Medicine, University of Toronto; Unity Health Toronto, Toronto, Canada

<sup>12</sup>Data Sciences Institute, University of Toronto, Toronto, Canada

## Address for correspondence:

Lillian Sung MD, PhD

Division of Haematology/Oncology

The Hospital for Sick Children,

555 University Avenue, Toronto, Ontario, M5G1X8, Canada

Telephone: 416-813-5287

Fax: 416-813-5979

Email: Lillian.sung@sickkids.ca

## Abstract

**Background:** Data sciences solutions such as artificial intelligence are increasing. A common challenge is identifying appropriate scenarios or “use cases” for data sciences implementation. Implementation frameworks are primarily generated from a theoretical or single institution based perspectives, highlighting the need for a multi-institutional experience to reveal patterns within successful implementations.

**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

## Introduction

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. 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.

## 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[9] 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: General considerations required to implement data sciences solutions in healthcare.

Scenario identification: idea generation and value proposition, prioritization and approval*
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 and maintenance
Evaluation of impact on patient or operational outcomes and return on investment

\* Focus for this manuscript

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

	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

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 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. 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 at the point of clinical implementation, or when there are implications for institutional resources or impact.

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.

**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. These individuals are likely to be at the executive level. 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. 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.

## Acknowledgments

LS is supported by the Canada Research Chair in Pediatric Oncology Supportive Care. We would like to acknowledge DSI and T-CAIREM for co-sponsoring the workshop and DSI for hosting.

## Conflicts of Interest

None declared.

## Abbreviations

DSI: The Data Sciences Institute

ML: Machine Learning

T-CAIREM: Temerty Faculty of Medicine Centre for Artificial Intelligence Education and Research in Medicine

## References

1. Gagalova KK, Leon Elizalde MA, Portales-Casamar E, et al. 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 Aug 27;2020;4(8):e17687. PMID:32852280

2. Harris S, Bonnici T, Keen T, et al. Clinical deployment environments: Five pillars of translational machine learning for health. Front Digit Health 2022;4:939292. PMID:36060542
3. Jeong H, Kamaleswaran R. Pivotal challenges in artificial intelligence and machine learning applications for neonatal care. Semin Fetal Neonatal Med Oct;2022;27(5):101393. PMID:36266181
4. Liao F, Adelaine S, Afshar M, et al. 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. PMID:36093386
5. Novak LL, Russell RG, Garvey K, et al. Clinical use of artificial intelligence requires AI-capable organizations. JAMIA Open Jul;2023;6(2):ooad028. PMID:37152469
6. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med Jan;2022;28(1):31-38. PMID:35058619
7. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med Sep;2019;25(9):1337-1340. PMID:31427808
8. Verma AA, Trbovich P, Mamdani M, et al. Grand rounds in methodology: key considerations for implementing machine learning solutions in quality improvement initiatives. BMJ Qual Saf Jan 19;2024;33(2):121-131. PMID:38050138
9. Columbia University Mailman School of Public Health. Content Analysis. Accessed 18 May, 2024. <https://www.publichealth.columbia.edu/research/population-health-methods/content-analysis>