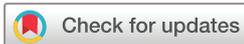


Healthcare provider evaluation of machine learning-directed care: reactions to deployment on a randomised controlled study

Julian C Hong,^{1,2,3} Pranalee Patel,⁴ Neville C W Eclow,⁴ Sarah J Stephens,⁴ Yvonne M Mowery,^{4,5} Jessica D Tenenbaum,⁶ Manisha Palta⁴

To cite: Hong JC, Patel P, Eclow NCW, *et al.* Healthcare provider evaluation of machine learning-directed care: reactions to deployment on a randomised controlled study. *BMJ Health Care Inform* 2023;**30**:e100674. doi:10.1136/bmjhci-2022-100674

Received 01 September 2022
Accepted 28 January 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Radiation Oncology, University of California San Francisco, San Francisco, California, USA

²Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, California, USA

³Joint Program in Computational Precision Health, UCSF-UC Berkeley, San Francisco, California, USA

⁴Department of Radiation Oncology, Duke University, Durham, North Carolina, USA

⁵Department of Head and Neck Surgery & Communication Sciences, Duke University, Durham, North Carolina, USA

⁶Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA

Correspondence to

Dr Julian C Hong;
julian.hong@ucsf.edu

ABSTRACT

Objectives Clinical artificial intelligence and machine learning (ML) face barriers related to implementation and trust. There have been few prospective opportunities to evaluate these concerns. System for High Intensity Evaluation During Radiotherapy (NCT03775265) was a randomised controlled study demonstrating that ML accurately directed clinical evaluations to reduce acute care during cancer radiotherapy. We characterised subsequent perceptions and barriers to implementation.

Methods An anonymous 7-question Likert-type scale survey with optional free text was administered to multidisciplinary staff focused on workflow, agreement with ML and patient experience.

Results 59/71 (83%) responded. 81% disagreed/strongly disagreed their workflow was disrupted. 67% agreed/strongly agreed patients undergoing intervention were high risk. 75% agreed/strongly agreed they would implement the ML approach routinely if the study was positive. Free-text feedback focused on patient education and ML predictions.

Conclusions Randomised data and firsthand experience support positive reception of clinical ML. Providers highlighted future priorities, including patient counselling and workflow optimisation.

INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) has the potential to transform medical practice. Despite many retrospective studies, randomised controlled trials (RCTs), particularly interventional trials, remain limited.^{1–3} Thus, there have been limited opportunities to formally characterise barriers to the implementation of healthcare AI and ML and identify solutions.^{4,5} There are minimal reports describing provider opinions following a prospective randomised interventional study of healthcare ML.

One application of healthcare ML is in the prediction and reduction of acute care (emergency visits and hospitalisations) during outpatient cancer therapy,^{1 6–9} prioritized by the Centers for Medicare and

Medicaid Services.¹⁰ The System for High Intensity Evaluation During Radiotherapy study (SHIELD-RT; NCT03775265) was a randomised controlled quality improvement study of an ML model predicting acute care visits (emergency department visits and/or hospitalisation) during radiotherapy (RT) or chemoradiotherapy (CRT).^{1 6} ML identified high-risk patients for supplemental clinical evaluations, which reduced acute care rates from 22.3% to 12.3%, with low-risk patients experiencing a 2.7% rate. Radiation oncology care uniquely requires a diverse clinical staff, including attending and resident physicians, advanced practice provider (APPs), nurses and radiation therapists (RTTs), each with different viewpoints on how ML can optimally play a role in delivering care. Following the completion but prior to final analysis of SHIELD-RT, we administered a survey to understand the perspectives of healthcare providers with regard to the acceptability and feasibility of ML-directed strategies, addressing key components of the implementation outcomes framework.¹¹ The objective was to evaluate specific barriers to planned long-term implementation.

METHODS

We conducted a single institution survey of perceptions of SHIELD-RT, during which all outpatient adult courses of RT and CRT initiated from 7 January 2019 to 30 June 2019 were evaluated during the first week of treatment by ML to identify high-risk patients with >10% risk of an acute care visit during RT.^{1 6} Patients were randomised to standard of care (mandatory weekly on-treatment and clinically indicated ad hoc visits) versus mandatory twice-weekly visits. Interventional second weekly visits were facilitated through an alert

Table 1 Responses to survey questions

Paraphrased survey question	Response, median (25–75th percentiles)*					
	Overall	Attending physicians	Resident physicians	Advanced practice providers (APP)	Nurses	Radiotherapy technologists (RTT)
The study disrupted my clinical workflow†	4 (4–5)	4.5 (3.25–5)	4 (4–4)	3 (3–3.5)	4 (3.25–4)	4 (4–4.5)
Patients on the intervention were at high risk for acute care visits‡	2 (2–3) 13 not aware	2 (2–3)	2 (2–2)	2 (2–2)	2 (2–2)	2.5 (2–3)
I was aware of my patients who were undergoing intervention	2 (2–3)	3.5 (2.25–5)	2 (2–3)	2 (1.5–2.5)	2.5 (2–3)	2 (2–3)
The study altered my clinical management‡	4 (3–4) 5 not aware	4 (3.5–4)	3 (3–3.5)	4 (4–4)	4 (4–4.75)	4 (3–4.5)
I would implement the machine learning system routinely if the study is positive	2 (1–3)	2 (2–2)	2 (1–3)	3 (2–3)	2 (1–2)	2 (2–3)
My opinion of machine learning to assist with clinical care is now...	3 (2–3)	3 (2.25–3)	2 (2–3)	3 (2–3)	3 (3–3)	2 (2–3)
Patients understood the study after their first mandatory visit	3 (2–3)	3 (3–3)	3 (2–3)	3 (2.5–3)	3 (2–3)	3 (3–3.5)

*Responses on a 5-point Likert-type scale ranging from 1 (strongly agree; much better) to 5 (strongly disagree; much worse).
†APPs, nurses and RTTs were most frequently directly involved in the supplemental visits.
‡Also included the option for respondents to indicate that they were not aware which patients were identified as high risk by the algorithm.

that notified RTTs to bring patients to an appropriate clinic room to then be seen by an APP, nurse clinician, resident physician or attending physician. The primary endpoint was rate of acute care visits during RT. Additional details of SHIELD-RT and its primary analysis and implementation workflow were previously reported.^{1 12}

Involved attending and resident physicians, APPs, nurses and RTTs were invited to participate in an anonymous survey to characterise workflow satisfaction and evaluation of potential barriers to future adoption. This included eight questions on a Likert-type scale characterising respondents' attitudes with an optional free-text comment field.

RESULTS

A total of 59/71 (83%) of invited staff completed the survey, including 14/16 attending physicians (MD), 9/9 resident physicians, 3/5 APPs, 10/11 nurses, 23/30 RTTs (table 1). Eighty-one per cent of staff disagreed or strongly disagreed that the study disrupted their workflow. Only 51% of respondents agreed or strongly agreed that they were aware of their patients undergoing the intervention; 3% agreed that their clinical management beyond the study intervention was altered. Of those aware of patients seen twice weekly, 67% agreed or strongly agreed that patients undergoing intervention were high risk. Most staff (64%) neither agreed nor disagreed that patients understood the study. Willingness for future adoption was favourable, as 75% of respondents agreed

or strongly agreed that they would implement the intervention routinely if the study was positive; 41% agreed or strongly agreed and none disagreed that their opinion of clinical ML improved following the study.

There were 8 (16%) free-text comments. Three (two RTT and one nurse) indicated confusion among staff and patients with the need and logistics of the supplemental visit. One nurse noted that they felt ML overestimated the risk of their patients (specifically in brain tumours). Two MD responses indicated that they had minimal contact with patients on study. Two (one MD and one RTT) responses expressed anticipation for the results of the study.

DISCUSSION

Our study highlights an overall positive reception towards ML implementation in an academic radiation oncology clinic. Our survey supports that RCT results drive willingness to routinely adopt clinical ML. ML-guidance and supplemental visits were integrated successfully into our clinical workflow with minimal perceived disruption.

This analysis shows how some concerns regarding ML may be overcome. In addition to randomised evidence, direction observation of ML operating in a controlled setting may have improved subjective opinions of clinical ML prior to the study. This is instrumental given recent data demonstrating the limitations of commercial prediction models,¹³ and ultimately, subsequent to this survey, the SHIELD-RT analysis demonstrated a reduction in

acute care events.¹ While ML will continue to require complementary input from healthcare professionals, these survey results are promising for adoption.¹⁴ Our clinic is currently incorporating this ML-directed clinical strategy into routine practice.

Overall, ML implementation had limited provider-perceived impact on clinical workflows, to the point of reducing MD awareness as indicated by survey responses. This was intentional in the design to minimise extra cognitive and functional effort to improve the likelihood of MD adoption.^{1 12} One relative exception to this was surveyed APPs, the majority of whom participated in the interventional second mandatory clinical evaluation. This suggests that ML-guided interventions may place greater burden on specific staff. This cost must be considered in model and interventional design.

Among limited free-text comments, staff reservations focused on patient education and ML risk predictions. Patients were not surveyed, although staff both anecdotally and in the survey highlighted logistical challenges surrounding location and timing of supplemental visits. While patients were educated when undergoing the supplemental evaluation, the neutral evaluation of patient understanding and anecdotal responses highlight the reported challenges of explaining the algorithm and its clinical implications to patients. This emphasises the need for transparent and explainable approaches, especially given increasingly opaque AI methods. Despite the single comment noting concern for overestimation, calibration analyses previously reported in the primary study results demonstrated good model performance in comparison to clinicians who were more inconsistent, with wide CIs, and assigned a 0% risk to a patient who had an acute care event.¹ It is possible that over time, both improved explainability and consistent observation of ML accuracy may demonstrate longitudinal improvements in clinician perception.

There are limitations to our study. We surveyed staff only following completion of the study, and direct comparisons pre-SHIELD-RT and post-SHIELD-RT were not possible. The results of this survey may be subject to bias, though we had a high rate of completion (83%) across a range of roles, with a high representation of non-academic staff (61% of respondents; APPs, nurses and RTTs).

The results of this study inform our future directions, primarily emphasising the importance of RCTs in demonstrated clinical ML benefit and highlighting the need for concerted efforts in patient and staff education. Other ongoing work focuses on optimising workflows, patient logistics, long-term ML surveillance and generalisability.

Twitter Julian C Hong @julian_hong

Contributors JCH and MP: concept and design. JCH, PP, NCWE, SJS, YMM, JDT, MP: acquisition, analysis or interpretation of data. JCH, PP, NCWE, SJS, YMM, JDT, MP: drafting of the manuscript. JCH, PP, NCWE, SJS, YMM, JDT, MP: critical revision of the manuscript for important intellectual content.

Funding This study was supported in part by the Duke Endowment, the Radiation Oncology Institute and the Conquer Cancer Foundation which had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The Duke Department of Radiation Oncology also provided funding. JCH is also supported by a career development grant from the American Society for Radiation Oncology and Prostate Cancer Foundation.

Competing interests JDT, MP and JH are coinventors on a pending patent, 'Systems and methods for predicting acute care visits during outpatient cancer therapy,' related to the current work.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Research data are stored in an institutional repository, and will be shared on reasonable request to the corresponding author.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Hong JC, Eclov NCW, Dalal NH, *et al*. System for high-intensity evaluation during radiation therapy (SHIELD-RT): a prospective randomized study of machine learning-directed clinical evaluations during radiation and chemoradiation. *J Clin Oncol* 2020;38:3652–61.
- Nimri R, Battelino T, Laffel LM, *et al*. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nat Med* 2020;26:1380–4.
- Wijnberge M, Geerts BF, Hol L, *et al*. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the hype randomized clinical trial. *JAMA* 2020;323:1052–60.
- Gaube S, Suresh H, Raue M, *et al*. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med* 2021;4:31.
- Henry KE, Kornfield R, Sridharan A, *et al*. Human-machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *NPJ Digit Med* 2022;5:97.
- Hong JC, Niedzwiecki D, Palta M, *et al*. Predicting emergency visits and hospital admissions during radiation and chemoradiation: an internally validated pretreatment machine learning algorithm. *JCO Clin Cancer Inform* 2018;2:1–11.
- Jairam V, Lee V, Park HS, *et al*. Treatment-Related complications of systemic therapy and radiotherapy. *JAMA Oncol* 2019;5:1028–35.
- Grant RC, Moineddin R, Yao Z, *et al*. Development and validation of a score to predict acute care use after initiation of systemic therapy for cancer. *JAMA Netw Open* 2019;2.
- Brooks GA, Uno H, Aiello Bowles EJ, *et al*. Hospitalization risk during chemotherapy for advanced cancer: development and validation of risk stratification models using real-world data. *JCO Clin Cancer Inform* 2019;3:1–10.
- Admissions and emergency department (ED) visits for patients receiving outpatient chemotherapy. Available: https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=2929 [Accessed 19 Dec 2019].
- Proctor E, Silmere H, Raghavan R, *et al*. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 2011;38:65–76.
- Hong JC, Eclov NCW, Stephens SJ, *et al*. Implementation of machine learning in the clinic: challenges and lessons in prospective deployment from the system for high intensity evaluation during radiation therapy (SHIELD-RT) randomized controlled study. *BMC Bioinformatics* 2022;23(Suppl 12):408.
- Wong A, Otles E, Donnelly JP, *et al*. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med* 2021;181:1065–70.
- Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: Humanism and artificial intelligence. *JAMA* 2018;319:19–20.