

REVIEW ARTICLE OPEN



The potential of artificial intelligence to improve patient safety: a scoping review

David W. Bates^{1,2,3}✉, David Levine^{1,2}, Ania Syrowatka^{1,2}, Masha Kuznetsova⁴, Kelly Jean Thomas Craig⁵, Angela Rui¹, Gretchen Purcell Jackson^{5,6} and Kyu Rhee⁵

Artificial intelligence (AI) represents a valuable tool that could be used to improve the safety of care. Major adverse events in healthcare include: healthcare-associated infections, adverse drug events, venous thromboembolism, surgical complications, pressure ulcers, falls, decompensation, and diagnostic errors. The objective of this scoping review was to summarize the relevant literature and evaluate the potential of AI to improve patient safety in these eight harm domains. A structured search was used to query MEDLINE for relevant articles. The scoping review identified studies that described the application of AI for prediction, prevention, or early detection of adverse events in each of the harm domains. The AI literature was narratively synthesized for each domain, and findings were considered in the context of incidence, cost, and preventability to make projections about the likelihood of AI improving safety. Three-hundred and ninety-two studies were included in the scoping review. The literature provided numerous examples of how AI has been applied within each of the eight harm domains using various techniques. The most common novel data were collected using different types of sensing technologies: vital sign monitoring, wearables, pressure sensors, and computer vision. There are significant opportunities to leverage AI and novel data sources to reduce the frequency of harm across all domains. We expect AI to have the greatest impact in areas where current strategies are not effective, and integration and complex analysis of novel, unstructured data are necessary to make accurate predictions; this applies specifically to adverse drug events, decompensation, and diagnostic errors.

npj Digital Medicine (2021)4:54; <https://doi.org/10.1038/s41746-021-00423-6>

INTRODUCTION

Adverse events related to unsafe care represent one of the top ten causes of death and disability worldwide, and a third to a half appear preventable¹. Investments in reducing harm can lead to substantial savings, and more importantly improve patient outcomes.

Twenty years after the Institute of Medicine's "To Err Is Human" report, problems with safety remain all too common² despite patient-centered strategies to create a culture of safety; for example, implementation of inpatient checklists, and computerization of prescribing and bar-coding^{3–6}. However, safety issues outside the hospital have received much less attention than hospital safety, yet care is increasingly being shifted outside the hospital.

The application of artificial intelligence (AI) has tremendous potential as a tool for improving safety, both inside and outside of the hospital, by providing solutions to predict harms, collect a variety of data including both new and already-available data, and as part of quality improvement initiatives. For instance, AI can provide decision support by identifying patients at high risk of hospital harm to guide prevention and early intervention strategies. Similarly, AI can be applied in outpatient, community, and home settings. When coupled with digital approaches, these technologies can improve communication between patients and healthcare providers to reduce the frequency of preventable harms. While existing data will be helpful, new data will be available through technologies like sensors which should improve predictions.

AI techniques, such as machine learning (ML), can be leveraged to provide clinical risk prediction to improve patient safety. Data-driven ML algorithms have advantages over rule-based approaches for risk prediction, as they allow simultaneous consideration of multiple data sources to identify predictors and outcomes. Healthcare organizations are increasingly implementing ML and other forms of AI to improve patient care and outcomes. However, substantial impacts to safety and reduction of associated costs related to safety issues will require further acceptance of these technologies across the larger ecosystem including regulatory agencies and the marketplace.

Evidence suggests that the majority of healthcare harms fall into the following domains: healthcare-associated infections (HAIs), adverse drug events (ADEs), venous thromboembolism (VTE), surgical complications, pressure ulcers, falls, insufficient decompensation detection, and diagnostic errors—including missed and delayed diagnoses^{7,8}. These domains are centered around hospital harm, and other issues undoubtedly play a role, but these adverse events account for the bulk of harm in hospitals. The goal of this paper was to conduct a scoping review to evaluate if AI has the potential to improve healthcare safety by reducing the frequency of adverse events within these eight major domains of harm.

METHODS

This scoping review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR)⁹.

¹Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA. ²Harvard Medical School, Boston, MA, USA. ³Harvard T. H. Chan School of Public Health, Boston, MA, USA. ⁴Harvard Business School, Harvard University, Boston, MA, USA. ⁵IBM Watson Health, Cambridge, MA, USA. ⁶Department of Pediatric Surgery, Vanderbilt University Medical Center, Nashville, TN, USA. ✉email: dbates@bwh.harvard.edu

Search strategy

A structured search was used to query MEDLINE (Ovid) for relevant articles published on or before October 25, 2019. Two main concepts of AI and patient safety, including the eight harm domains, were mapped to the most relevant controlled vocabulary using Medical Subject Headings (MeSH), and free-text terms were added where necessary. The full search strategy is provided in Supplementary Note 1.

Inclusion and exclusion criteria

The scoping review included studies that focused on the application of AI for prediction, prevention, and/or early detection of events in each of the harm domains in hospital, outpatient, community, and home settings. No comparisons were required, and all study designs were considered for inclusion. Articles were excluded if they were not published in the English language or reported on the use of AI to measure the frequency of harm events (e.g., post-marketing surveillance of drugs). Applications in robotics were also excluded. Detailed inclusion and exclusion criteria are provided in Supplementary Table 1.

Screening and data abstraction

Articles were screened in two stages using Covidence (Australia), a web-based review management tool. Titles and abstracts were screened for relevance, and eligible records were evaluated based on full-text articles by a single reviewer. Additional articles were identified through handsearching. For each article included in the scoping review, citation information was exported from Covidence into an Excel spreadsheet and harm domains were manually abstracted by a single reviewer.

Scoping review

The characteristics of studies that reported on the use of AI to improve patient safety were summarized. The literature was narratively synthesized for each harm domain highlighting key examples of how AI can be leveraged for prediction, prevention, and/or early detection of patient harms. Selected examples of traditional and novel data sources that could be used to develop AI algorithms to improve patient safety were summarized in tabular form.

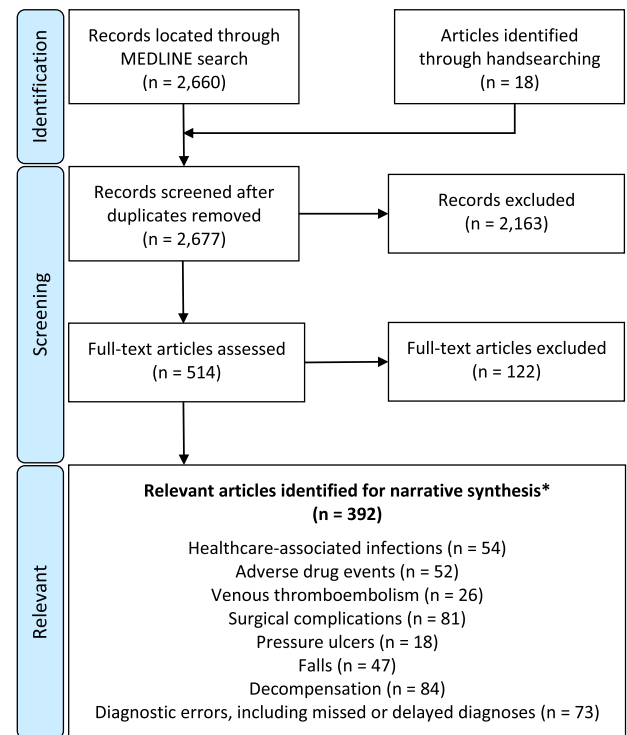
Evaluation of the potential for AI to improve patient safety

The findings of the scoping review were considered in the context of incidence, cost, and preventability of events to evaluate the potential of AI for improving safety. Current literature reporting on incidence, cost, and preventability was summarized for the eight harm domains in tabular form. Cost estimates were adjusted to United States dollars (USD, 2019) using the Producer Price Index to facilitate comparisons across the domains¹⁰. Projections around the likelihood of AI to improve safety in each of the harm domains were made and attractive early targets were identified as part of the Discussion.

RESULTS

Characteristics of included studies

From 2677 unique records, 392 articles met the inclusion criteria for the scoping review and are presented in Supplementary Table 2. A modified Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram is provided in Fig. 1. The majority of studies were pre-clinical and relied on retrospective analyses of data. Most algorithms were not externally validated or tested prospectively. The incidence, cost, and preventability of events for each harm domain are presented in



*Some articles addressed multiple harm domains.

Fig. 1 PRISMA flow diagram showing disposition of articles. The asterisk denotes that some studies addressed multiple harm domains.

Table 1. Traditional and novel data sources that can be used to develop AI algorithms are presented in Table 2.

Healthcare-associated infections

Approximately 3.2% of inpatients experienced HAIs in 2015 (ref. ¹¹). The estimated annual cost for five significant HAIs is 10.7 billion (USD, 2019)¹². Up to 70% of specific HAIs are considered preventable using existing evidence-based strategies¹³. The scoping review identified 54 articles (see Supplementary Table 2) describing the use of AI for prediction or early detection of HAIs.

ML and fuzzy logic (i.e., logical reasoning models based on incomplete or ambiguous data) have been applied for early detection of HAIs. Most algorithms were developed using claims-based data and information captured in electronic health records (EHRs) including laboratory test results and diagnostic imaging. With the integration of novel complex data, AI-based analytics could expedite detection and further improve diagnostic accuracy. For example, data from eNoses (i.e., chemical vapor sensors) have been analyzed using ML methods to rapidly detect ventilator-associated pneumonia (area under the curve (AUC) = 0.98), differentiate between six common wound pathogens (accuracy = 78%), and classify various strains of *Clostridium difficile* (sensitivities >80%; specificities >73%)^{14–16}.

AI can also contribute to infection control by providing real-time, accurate predictions of HAI risk to guide patient-specific interventions before an infection occurs. For example, a random forest classification algorithm can predict onset of central line-associated bloodstream infections with an AUC of 0.82 (ref. ¹⁷).

AI can also play a role in improving adherence to existing safety protocols; for instance, computer vision using a convolutional network classifier has been applied to monitor hand hygiene compliance in the hospital setting (accuracy = 75%). Similarly, an ML algorithm was developed to provide real-time hand hygiene

Table 1. Incidence, cost, and preventability of events in the eight harm domains from the peer-reviewed literature.

Patient safety domain	Population (years): incidence	Annual total cost adjusted to U.S. (2019) dollars ^a	Population: % preventable
Healthcare-associated infections	Inpatients (2015 data): 3.2% ¹¹	\$10.7 billion ¹² [five significant HAIs]	Inpatients: 65% to 70% ¹³ [CABSI or CAUTI]; 55% ¹³ [VAP or SSI]
Adverse drug events	Inpatients (2014 data): 2.1% ²² Present on admission (2014 data): 5.1% ²²	\$30.0 billion ²²	Inpatients: 28% ²³
Venous thromboembolism	Inpatients (2013 review): 3.3% ⁷	\$15.1 to \$30.4 billion ²⁸	Inpatient to 90 days after discharge: 70% ²⁹
Surgical complications	Inpatients (2014–2017 data): 16.0% within 30 days ³⁴ [invasive surgery]	\$7.5 billion ³⁵ [emergency general surgery]	Inpatients: 42.1% ³⁶ [emergency non- trauma surgery]
Pressure ulcers	Inpatients (2009–2010 data): 2.7% ⁴⁴	\$28.2 billion ⁴⁵	Inpatients: 97% ⁴⁶
Falls	Inpatients (2013 review): 1.1% ⁷	\$53.4 billion ⁵¹	Inpatients: 87.5% ⁴⁶
Decompensation	Inpatients (2013 data): 3.6% ⁵⁷ [septicemia] Inpatients (2005–2015 data): 13.2% ⁵⁸ [failure-to-rescue after complications of trauma surgery]	\$25.7 billion ⁵⁷ [septicemia]	Inpatients: 24.2% ⁵⁸ [failure-to-rescue after complications of trauma surgery]
Diagnostic errors	Outpatients (2014 review): at least 5.1% ⁷²	Exceeding \$100 billion ^b (ref. ⁷¹)	Unknown

CABSI catheter-associated bloodstream infection, CAUTI catheter-associated urinary tract infection, HAI healthcare-associated infection, SSI surgical site infection, U.S. United States, VAP ventilator-associated pneumonia.

^aEstimates adjusted to U.S. (2019) dollars using the Producer Price Index¹⁰.

^bEstimate not adjusted to U.S. (2019) dollars.

alerts in the outpatient setting based on data from multiple types of sensors, improving compliance from 54% to 100%^{18,19}. These technologies are increasingly being applied to complex problems and could be used to improve other aspects of infection control, including sanitation or adherence to condition-specific safety protocols^{20,21}.

Adverse drug events

In 2014, ADEs were associated with 1.6 million hospitalizations in the U.S., totaling an estimated 30.0 billion (USD, 2019), with ~½ million ADEs occurring during hospital stays (2.1% of inpatients) and ~1 million present on admission (5.1% of admissions)²². About one in four ADEs are considered preventable given what is known today²³. The review located 52 papers (see Supplementary Table 2) about leveraging AI to reduce the frequency of ADEs.

AI-based analytics can be applied to predict previously unreported ADEs based on drug similarities including chemical structure, mechanism of action, and polypharmacy side effects^{24,25}. Deep learning methods using neural fingerprints have been shown to not only predict adverse drug reactions with an AUC of ~0.85, but also identify the associated molecular substructures²⁶. These algorithms can inform the evidence-based development of safer medications. Similar techniques can be applied to predict drug–drug interactions for untested combinations of drugs²⁴.

At the point of care, ML can be applied to analyze multiple datasets, including traditional patient data documented in EHRs (e.g., medical history, laboratory test results) with novel data (e.g., bioactivity of single nucleotide polymorphisms (SNPs)), to provide personalized ADE risk estimates and treatment recommendations to support decision making. Using genomic sequencing data, an artificial neural network (ANN) algorithm was developed to guide safer and more effective dosing of warfarin, predicting therapeutic dose with an accuracy of 83% in patients with international normalized ratios (INRs) >3.5 (ref. ²⁷).

Venous thromboembolism

Approximately 3.3% of inpatients develop VTEs, including deep venous thromboses (DVT) and pulmonary emboli (PE), with an

estimated cost of 15.1–30.4 billion (USD, 2019) annually^{7,28}. Adherence to current evidence-based strategies could reduce up to 70% of healthcare-associated VTEs²⁹.

AI techniques can be used to identify patients at high risk for VTEs. The review located 26 articles (see Supplementary Table 2) about AI algorithms to prevent or safely rule out VTE. One study applied a super learner ensemble approach to identify inpatients at higher risk of future VTEs with an AUC of 0.69 (ref. ³⁰). Prediction can also be applied to manage at-risk populations in the outpatient setting; for example, a multiple kernel learning algorithm was developed to predict VTE risk among patients undergoing chemotherapy with a sensitivity of 89%, markedly outperforming the recommended Khorana score (sensitivity = 11%)³¹.

AI methods could also recommend optimal patient-specific treatments. As described above, ML leveraging genomic sequencing data was used to guide safer warfarin dosing resulting in a reduced time to achieving a therapeutic INR (OR = 6.7) compared with standard clinical dosing²⁷.

To date, AI has mostly contributed to VTE detection through the analysis of diagnostic imaging or radiologic reports. ML methods can also be applied to guide appropriate use of diagnostic imaging. For example, an ANN was applied to safely rule out DVT without ultrasonography in 38% of patients with a false-negative rate of only 0.2%³². Similarly, an ANN model was developed to guide computed tomography use for diagnosis of PE³³. The algorithm achieved an AUC of 0.90 using an internal validation sample and 0.71 using external data, reiterating the importance of external validation for all AI or ML models.

Surgical complications

Surgical complications are common; 16.0% of patients receiving invasive procedures experience a post-operative complication within 30 days³⁴. Annual U.S. costs associated with complications following emergency general surgery are 7.5 billion (USD, 2019)³⁵. It is estimated that 42.1% of complications following emergency non-trauma surgery are preventable³⁶.

ML use cases include predicting adverse events in both the operative and post-operative setting. Eighty-one papers that leveraged AI to reduce surgical complications were located through the scoping review (see Supplementary Table 2).

Table 2. Traditional and novel (italicized) data sources that can be used to develop artificial intelligence algorithms to improve patient safety: selected examples.

Patient safety domain	EHR	Claims	Risk scores	Genome sequencing	Sensors	Computer vision	Other
Healthcare-associated infections	X	X		<i>Of pathogens to monitor spread of nosocomial infections</i>	Continuous vitals, Chemical vapor sensors	<i>Of health care setting</i>	<i>Microbiology of random specimens, Smart sinks and dispensers</i>
Adverse drug events	X	X	BADRI model, Trivalle's risk score	<i>Of patients for Cytochrome P450 polymorphisms</i>	Continuous vitals, Continuous glucose monitoring		Patient report
Venous thromboembolism	X	X	Padua prediction score, Khorana score	<i>Of patients for Cytochrome P450 polymorphisms, Factor V Leiden, Prothrombin 20210 mutations</i>	Activity, Pressure, Location		Novel biochemical analytes
Surgical complications	X	X	Surgical risk scores		Continuous vitals	<i>Of operating room</i>	
Pressure ulcers	X	X	Braden score		Activity, Pressure, Location	<i>Of bed</i>	Novel biochemical analytes
Falls	X	X	Hendrich model		Activity, Pressure, Location	<i>Of common spaces</i>	
Decompensation	X	X	MEWS, CART score		Continuous vitals, Activity, Continuous glucose monitoring		Novel biochemical analytes
Diagnostic errors	X	X			Chemical vapor sensors		Clinician adjudication, Patient report

Novel data sources italicized. BADRI Brighton Adverse Drug Reactions Risk, CART Cardiac Arrest Risk Triage, EHR electronic health record, MEWS, Modified Early Warning Score.

Predicting blood loss, need for prolonged post-operative intubation, post-operative mortality, pain, nausea, and vomiting all represent areas with demonstrated improvements to current risk tools^{37–40}. For example, an ANN-based model achieved an accuracy of 92% at stratifying post-operative bleeding risk in patients undergoing cardiac pulmonary bypass³⁷. Another ANN algorithm was developed to predict the need for prolonged ventilation after coronary bypass grafting (AUC = 0.71–0.73)³⁸. Early intervention in these situations could translate into substantial improvements in patient safety.

An area of active research is the use of ML to recognize critical procedural steps in intra-operative videos. ANNs have been trained to identify the steps of laparoscopic sleeve gastrectomy procedures with an accuracy of 82%, and to determine whether the critical view of safety had been achieved in laparoscopic cholecystectomy videos, yielding an accuracy of 95%^{41,42}. ML algorithms that can identify key operative components might be used in the future during procedures to warn surgeons of deviations from an expected sequence of steps or omission of critical elements. Other ML approaches in surgery on the horizon include computer precision pre-operative evaluation, augmented reality in the operating room, technical skills augmentation such as suturing, and ultimately autonomous robotic surgery⁴³.

Pressure ulcers

Approximately 2.7% of hospitalized patients in the U.S. develop a pressure ulcer⁴⁴. The annual financial burden associated with treatment is estimated to be 28.2 billion (USD, 2019)⁴⁵. Up to 97% of hospital-acquired pressure ulcers are preventable⁴⁶.

The scoping review identified 18 articles (see Supplementary Table 2) that used AI for management of pressure ulcers. To date, most AI research in this area has focused on using sensor data for early detection; as such, using AI to predict future risk remains an area of opportunity. A recent study developed a random forest model, using EHR data to classify critical care patients based on their risk of developing pressure ulcers (AUC = 0.79 vs. 0.68 for the Braden Scale)⁴⁷. Earlier studies tested the feasibility of using smart beds and wheelchair cushions for pressure ulcer detection using fuzzy logic and ML models, respectively^{48,49}. Tracking data from embedded sensors, these algorithms detected a lack of movement and identified specific areas of skin that were at risk of developing an ulcer. Although the models were able to produce detection accuracy of up to 90% in experimental settings, their application and utility in notifying care providers and promoting early intervention remain uncertain.

Falls

In 2014, 7.0 million fall-related injuries occurred among adults aged 65 and older⁵⁰. These falls are estimated to account for 53.4 billion (USD, 2019)⁵¹. In the hospital, ~1.1% of inpatients experience a fall and 87.5% of these falls are considered preventable^{7,46}. Forty-seven articles (see Supplementary Table 2) identified through the scoping review described the use of AI for prediction or early detection of falls.

AI approaches could be used to predict fall risk at the point of care using existing data from EHRs. For example, a support vector machine model was able to predict inpatient falls based on data documented from the previous day⁵². However, the model showed a sensitivity of 65% and a specificity of 70%, which are comparable to existing clinical risk assessments.

Many studies have applied ML methods for the early detection of falls. Classification models using data from wearable sensors in a laboratory setting showed relatively high levels of accuracy (54–84%) at stratifying subjects based on their risk of falls^{53,54}. Using data from cameras, smart carpets, and wearable sensors intended for use in the home environment, support vector machine classifiers have been developed to detect falls, as well as

Table 3. Evaluation of the potential of artificial intelligence to improve patient safety in the eight harm domains.

Patient safety domain	Likelihood of impact
Healthcare-associated infections	AI may have a moderate impact on the reduction of HAIs given that current evidence-based practices are already effective when applied well.
Adverse drug events	AI can play a major role in ADE prevention. As more patients at risk of ADEs are accurately identified before a medication is administered or prescribed, a greater proportion of these events will become preventable. However, a key challenge lies in the lack of integrated high-quality datasets in which ADEs have been accurately captured. A variety of automated approaches have been effective at identifying patients likely to have experienced an ADE, but typically clinician adjudication is still required. ML could also help identify patients who may benefit from additional testing for specific single nucleotide polymorphisms to guide optimal drug therapy. These methods may also help to identify signals from the remainder of the genome beyond single nucleotide polymorphisms which may have prognostic impact.
Venous thromboembolism	We believe that AI will have a moderate effect on the reduction of VTE, as current evidence-based preventive strategies are already effective. AI solutions could provide further insights by identifying patients who could benefit from diagnostic testing for inherited thrombotic disorders to inform management of their condition.
Surgical complications	AI can be expected to have a moderate impact on the prediction and prevention of surgical complications both in the operating room and during recovery. Most complications felt to be preventable today are related to delayed diagnoses or intervention, technical issues, and infections. Given the overlap with other harm domains, focusing on advances in these other areas will likely also improve surgical safety.
Pressure ulcers	Pressure ulcers represent an attractive target with moderate to high potential for AI to prevent harm. Novel data sources such as motion and fluid sensors are now available, and large numbers of traditional clinical variables can be combined with the sensing data to predict who is at risk to guide evidence-based prevention.
Falls	AI is anticipated to have a moderate impact on fall prevention, given that this area has already received substantial attention and current risk mitigation strategies are effective. As with pressure ulcers, clinical data combined with sensing data can be used to predict when falls may occur, and which ones are likely to be associated with the most harm.
Decompensation	Leveraging novel data sources and AI has high potential to improve the prediction of decompensation to guide preventive strategies as well as early intervention to mitigate the impacts including premature death, given that current approaches are not effective. Given the serious nature of these events, preventing decompensation is a particularly attractive target. ML can deeply analyze data, beyond the standard values of heart rate or heart rate variability and will be critical to improving detection of decompensation and subsequent intervention.
Diagnostic errors	Diagnostic error is the most complex of the eight harm domains with vast opportunities for improvement using novel data sources and AI. ML could help to reduce the frequency of diagnostic errors by leveraging pattern recognition, bias minimization, and infinite capacity, areas where diagnosticians often falter. Although this area has garnered a lot of attention, many outstanding challenges remain, and current solutions only address a small fraction of what is possible. Most crucial to constructing valuable ML algorithms that help to reduce diagnostic error is the availability of large databases that accurately report errors.

ADE adverse drug event, AI artificial intelligence, HAI healthcare-associated infection, ML machine learning, VTE venous thromboembolism.

to identify deviating gait patterns as predictors of future falls^{55,56}. These models achieved accuracies of up to 100% in fall detection based on experimental and training datasets; however, their usability and applicability in real-world settings needs further testing.

Decompensation

Clinical deterioration in the hospital remains common. For example, 3.6% of inpatients develop sepsis, costing an estimated 25.7 billion (USD, 2019) annually⁵⁷. The failure-to-rescue rate following complications of trauma surgery, such as sepsis, is estimated at 13.2%, and one in four of these deaths are considered preventable⁵⁸. However, prediction and early detection of decompensation remain a challenge in all areas of medicine.

The review located 84 papers (see Supplementary Table 2) that used AI to predict or detect the early signs of decompensation. Most research has focused on sepsis detection, which has seen improvements compared to traditional methods although, as with most ML algorithms, its generalizability may be poor^{59–63}. It is likely that the detection of decompensation will improve by adding new categories of data, including biometric sensors such as continuous telemetry, motion activity sensors such as time spent in the bathroom or bedroom, novel biomarkers, and relevant patient-reported measures^{64–69}. For example, ML has been used for early detection of sepsis using novel gene

expression biomarkers with AUCs ranging from 0.86 to 0.92 (ref. ⁶⁸). An AI tool has also been developed using a random forest model to predict nocturnal hypoglycemia from midnight to 6 am with an AUC of 0.84 based on continuous glucose monitoring to provide real-time feedback to inform optimal diabetes management before going to sleep⁷⁰.

Diagnostic errors

Diagnostic errors—both missed and delayed diagnoses—are relatively common in both inpatient and outpatient settings and estimated to occur in at least 5.1% of the U.S. population each year, with associated costs exceeding 100 billion (USD, 2016) annually^{71,72}.

The scoping review identified 73 articles (see Supplementary Table 2) that leveraged AI to reduce diagnostic error. ML has widely demonstrated reduced errors in interpretation of imaging⁷³. It has also proven beneficial for early diagnosis of lung cancer by analyzing exhaled breath using an eNose sensor; the support vector machine was able to classify cancer patients vs. non-cancer controls with a sensitivity of 87% and a specificity of 71%⁷⁴. AI techniques are also being applied to reduce delays for critical diagnoses; for example, a clinical decision support system based on fuzzy logic was able to appropriately triage patients presenting to an emergency department with an accuracy of >99%—a 13% increase compared with traditional methods⁷⁵.

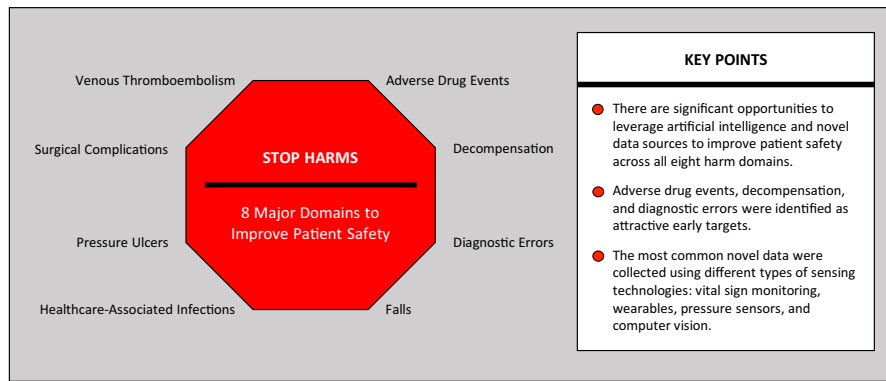


Fig. 2 Summary of major domains of harm and key points. The first panel highlights the eight major domains of harm. The second panel summarizes the key points from the article.

A recent issue of the journal *Diagnostics* was devoted to this area⁷⁶, and articles addressed diagnosis of a number of conditions. Another recent review summarized the main classes of problems that they believed AI systems are well suited to solve⁷⁷.

DISCUSSION

Based on epidemiologic evidence and our scoping review, we believe that there are major opportunities to improve safety using data and AI across the eight domains to reduce the frequency of harm (Table 3). We expect AI to have the greatest impact in areas where current strategies are not effective, and integration and complex analysis of novel, unstructured data are necessary to make accurate predictions, which applies specifically to ADEs, decompensation, and diagnostic errors.

However, the application of AI and ML to improve patient safety is an emerging field and most of these algorithms have not yet been externally validated or tested prospectively. Promising performance based on development or internal validation samples may not translate into improvements in real-world practice. Algorithms may be limited in generalizability, and performance may be affected by the clinical context where the solution is implemented. Although the level of evidence is modest for all domains, we are highlighting what we believe to be the most promising areas.

Future research must focus on careful evaluation of clinical decision support systems based on AI analytics prior to widespread implementation to ensure safety and accuracy. From a technical perspective, candidate algorithms and tools should be validated at other sites, account for differential performance in subgroups, and explicitly report the uncertainty around any estimates or recommendations⁷⁸. Furthermore, papers describing model development and performance assessments should adhere to reporting standards for transparency and provide important information about validity, biases, and generalizability to other settings⁷⁹. Once high-quality AI solutions are developed, additional factors beyond performance must be considered to increase the likelihood of successful implementation and adoption by individual providers. There is an active area of research focused on identifying key barriers and facilitators to implementation of AI-based tools in healthcare^{78,80,81}.

With data available today, especially laboratory information, imaging and continuous vital sign data, it should be possible to reduce the frequency of many types of harm. However, when the data are available, they are often unstructured, simply not in any documented form, or disputed. High-quality, large annotated databases will prove quite fruitful in minimizing patient harm in the future. New types of data, especially from the huge array of sensing technologies becoming available, but also including data from various other sources like information supplied directly by

patients, genomic sequencing, and social media, offer new opportunities to improve predictions as the first step toward development of preventive interventions to improve safety. These types of data are becoming available and more accessible over time for research and to drive innovation^{82–84}.

In addition, automated detection of safety issues of all types, but especially harm outside the hospital (e.g., post-marketing surveillance of drugs), will make routine measurement of the frequency of harm possible. While some of this will be rule-based, data-driven AI will also undoubtedly play a role.

This study has several limitations. The search query extracted evidence from a single database to identify published articles focused on the eight harm domains, and other literature may be available. Screening and data abstraction were completed by a single reviewer. The projections were informed by the incidence, cost, and preventability of harm as well as effectiveness of current strategies and promise of AI solutions.

CONCLUSIONS

Overall, AI has great potential to improve the safety of care (Fig. 2). In our view, harm domains including ADEs, decompensation, and diagnostic errors represent particularly attractive early targets. Transparent population-based datasets, which include diverse traditional (e.g., EHR, claims) and novel data (e.g., sensors, wearables, broader determinants of health), will be essential to build robust and equitable models. For AI to be effective, implementation of data-driven analytics will require organizations to develop, support, and iterate clinician, team, and system workflows for continued patient safety improvements.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article and its Supplementary Information.

Received: 27 April 2020; Accepted: 16 February 2021;
Published online: 19 March 2021

REFERENCES

1. Kohn, L., Corrigan, J. & Donaldson, M. *To Err Is Human* (National Academies Press, 2000).
2. Bates, D. W. & Singh, H. Two decades since to err is human: an assessment of progress and emerging priorities in patient safety. *Health Aff.* **37**, 1736–1743 (2018).
3. Pronovost, P. et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N. Engl. J. Med.* **355**, 2725–2732 (2006).
4. Haynes, A. B. et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N. Engl. J. Med.* **360**, 491–499 (2009).

5. Bates, D. W. et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* **280**, 1311 (1998).
6. Poon, E. G. et al. Effect of bar-code technology on the safety of medication administration. *N. Engl. J. Med.* **362**, 1698–1707 (2010).
7. Jha, A. K. et al. The global burden of unsafe medical care: analytic modelling of observational studies. *BMJ Qual. Saf.* **22**, 809–815 (2013).
8. Jha, A. K., Chan, D. C., Ridgway, A. B., Franz, C. & Bates, D. W. Improving safety and eliminating redundant tests: cutting costs in U.S. hospitals. *Health Aff.* **28**, 1475–1484 (2009).
9. Tricco, A. C. et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* **169**, 467–473 (2018).
10. U.S. Bureau of Labor Statistics. Producer price index by industry: selected health care industries (PCUASHCASHC). <https://fred.stlouisfed.org/series/PCUASHCASHC> (2020).
11. Magill, S. S. et al. Changes in prevalence of health care-associated infections in U.S. hospitals. *N. Engl. J. Med.* **379**, 1732–1744 (2018).
12. Zimlichman, E. et al. Health care-associated infections. *JAMA Intern. Med.* **173**, 2039 (2013).
13. Umscheid, C. A. et al. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infect. Control Hosp. Epidemiol.* **32**, 101–114 (2011).
14. Liao, Y.-H. et al. Machine learning methods applied to predict ventilator-associated pneumonia with *Pseudomonas aeruginosa* infection via sensor array of electronic nose in intensive care unit. *Sensors* **19**, 1866 (2019).
15. Saviak, T. et al. Electronic nose in the detection of wound infection bacteria from bacterial cultures: a proof-of-principle study. *Eur. Surg. Res.* **59**, 1–11 (2018).
16. Kuppasami, S., Clokie, M. R. J., Panayi, T., Ellis, A. M. & Monks, P. S. Metabolite profiling of *Clostridium difficile* ribotypes using small molecular weight volatile organic compounds. *Metabolomics* **11**, 251–260 (2015).
17. Beeler, C. et al. Assessing patient risk of central line-associated bacteremia via machine learning. *Am. J. Infect. Control* **46**, 986–991 (2018).
18. Haque, A. et al. Towards vision-based smart hospitals: a system for tracking and monitoring hand hygiene compliance. *Mach. Learn. Healthc. Conf.* (2017).
19. Geilleit, R. et al. Feasibility of a real-time hand hygiene notification machine learning system in outpatient clinics. *J. Hosp. Infect.* **100**, 183–189 (2018).
20. Mehra, R., Bianconi, G. M., Yeung, S. & Fei-Fei, L. Depth-based activity recognition in ICUs using convolutional and recurrent neural networks. *Mach. Learn. Healthc. Conf.* 1–9 (2017).
21. Suresh, H. et al. Clinical intervention prediction and understanding using deep networks. *Mach. Learn. Healthc. Conf.* **68**, 1–16 (2017).
22. Weiss, A., Freeman, W., Heslin, K. & Barrett, M. Adverse drug events in U.S. hospitals, 2010 versus 2014. <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb234-Adverse-Drug-Events.pdf> (2018).
23. Bates, D. W. et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE prevention study group. *JAMA* **274**, 29–34 (1995).
24. Zitnik, M., Agrawal, M. & Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **34**, i457–i466 (2018).
25. Ogallo, W. & Kanter, A. S. Towards a clinical decision support system for drug allergy management: are existing drug reference terminologies sufficient for identifying substitutes and cross-reactants? *Stud. Health Technol. Inform.* **216**, 1088 (2015).
26. Dey, S., Luo, H., Fokoue, A., Hu, J. & Zhang, P. Predicting adverse drug reactions through interpretable deep learning framework. *BMC Bioinformatics* **19**, 476 (2018).
27. Pavani, A. et al. Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization. *Pharmacogenomics* **17**, 121–131 (2016).
28. Mahan, C. E. et al. Venous thromboembolism: annualised United States models for total, hospital-acquired and preventable costs utilising long-term attack rates. *Thromb. Haemost.* **108**, 291–302 (2012).
29. Zeidan, A. M. et al. Impact of a venous thromboembolism prophylaxis “smart order set”: improved compliance, fewer events. *Am. J. Hematol.* **88**, 545–549 (2013).
30. Nafee, T. et al. Machine learning to predict venous thrombosis in acutely ill medical patients. *Res. Pract. Thromb. Haemost.* **4**, 230–237 (2020).
31. Ferroni, P. et al. Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients. *Med. Decis. Making* **37**, 234–242 (2017).
32. Willan, J., Katz, H. & Keeling, D. The use of artificial neural network analysis can improve the risk-stratification of patients presenting with suspected deep vein thrombosis. *Br. J. Haematol.* **185**, 289–296 (2019).
33. Banerjee, I. et al. Development and performance of the pulmonary embolism result forecast model (PERFORM) for computed tomography clinical decision support. *JAMA Netw. Open* **2**, e198719 (2019).
34. Corey, K. M. et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. *PLoS Med.* **15**, e1002701 (2018).
35. Scott, J. W. et al. Use of national burden to define operative emergency general surgery. *JAMA Surg.* **151**, e160480 (2016).
36. Linnebur, M. et al. Preventable complications and deaths after emergency non-trauma surgery. *Am. Surg.* **84**, 1422–1428 (2018).
37. Huang, R. S. P. et al. Post-operative bleeding risk stratification in cardiac pulmonary bypass patients using artificial neural network. *Ann. Clin. Lab. Sci.* **45**, 181–186 (2015).
38. Wise, E. S. et al. Prediction of prolonged ventilation after coronary artery bypass grafting: data from an artificial neural network. *Heart Surg. Forum* **20**, E007–E014 (2017).
39. Bertsimas, D., Dunn, J., Velmahos, G. C. & Kaafarani, H. M. A. Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. *Ann. Surg.* **268**, 574–583 (2018).
40. Wu, H.-Y. et al. Predicting postoperative vomiting among orthopedic patients receiving patient-controlled epidural analgesia using SVM and LR. *Sci. Rep.* **6**, 27041 (2016).
41. Hashimoto, D. A. et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve. *Ann. Surg.* **270**, 414–421 (2019).
42. Namazi, B., Sankaranarayanan, G., Devarajan, V. & Fleshman, J. A deep learning system for automatically identifying critical view of safety in laparoscopic cholecystectomy videos for assessment. In *SAGES 2017 Annual Meeting* (Sages, Houston, TX, 2017).
43. Hashimoto, D. A., Rosman, G., Rus, D. & Meireles, O. R. Artificial intelligence in surgery. *Ann. Surg.* **268**, 70–76 (2018).
44. Gardiner, J. C., Reed, P. L., Bonner, J. D., Haggerty, D. K. & Hale, D. G. L. Incidence of hospital-acquired pressure ulcers - a population-based cohort study. *Int. Wound J.* **13**, 809–820 (2016).
45. Padula, W. V. & Delarmente, B. A. The national cost of hospital-acquired pressure injuries in the United States. *Int. Wound J.* **16**, 634–640 (2019).
46. Landrigan, C. P. et al. Temporal trends in rates of patient harm resulting from medical care. *N. Engl. J. Med.* **363**, 2124–2134 (2010).
47. Alderden, J. et al. Predicting pressure injury in critical care patients: a machine-learning model. *Am. J. Crit. Care* **27**, 461–468 (2018).
48. Hsiao, R.-S. et al. Body posture recognition and turning recording system for the care of bed bound patients. *Technol. Health Care* **24**, S307–S312 (2015).
49. Luboz, V. et al. Personalized modeling for real-time pressure ulcer prevention in sitting posture. *J. Tissue Viability* **27**, 54–58 (2018).
50. Bergen, G., Stevens, M. R. & Burns, E. R. Falls and fall injuries among adults aged ≥65 years — United States, 2014. *MMWR Morb. Mortal. Wkly. Rep.* **65**, 993–998 (2016).
51. Florence, C. S. et al. Medical costs of fatal and nonfatal falls in older adults. *J. Am. Geriatr. Soc.* **66**, 693–698 (2018).
52. Yokota, S., Endo, M. & Ohe, K. Establishing a classification system for high fall-risk among inpatients using support vector machines. *CIN Comput. Inform. Nurs.* **35**, 408–416 (2017).
53. Howcroft, J., Kofman, J. & Lemaire, E. D. Prospective fall-risk prediction models for older adults based on wearable sensors. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1812–1820 (2017).
54. Howcroft, J., Lemaire, E. D. & Kofman, J. Wearable-sensor-based classification models of faller status in older adults. *PLoS ONE* **11**, e0153240 (2016).
55. Alazrai, R., Mowafi, Y. & Hamad, E. A fall prediction methodology for elderly based on a depth camera. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, 4990–4993 (2015).
56. Juang, L.-H. & Wu, M.-N. Fall down detection under smart home system. *J. Med. Syst.* **39**, 107 (2015).
57. Torio, C. M. & Moore, B. J. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. (Agency for Healthcare Research and Quality, Rockville, MD, 2016).
58. Kuo, L. E. et al. Failure-to-rescue after injury is associated with preventability: the results of mortality panel review of failure-to-rescue cases in trauma. *Surgery* **161**, 782–790 (2017).
59. Sanchez-Pinto, L. N., Venable, L. R., Fahrenbach, J. & Churpek, M. M. Comparison of variable selection methods for clinical predictive modeling. *Int. J. Med. Inform.* **116**, 10–17 (2018).
60. Ward, L., Paul, M. & Andreassen, S. Automatic learning of mortality in a CPN model of the systemic inflammatory response syndrome. *Math. Biosci.* **284**, 12–20 (2017).
61. Taylor, R. A. et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad. Emerg. Med.* **23**, 269–278 (2016).
62. Islam, M. M. et al. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput. Methods Prog. Biomed.* **170**, 1–9 (2019).

63. Wetzel, R. C., Aczon, M. & Ledbetter, D. R. Artificial intelligence: an inkling of caution. *Pediatr. Crit. Care Med.* **19**, 1004–1005 (2018).
64. Vandendriessche, B., Abas, M., Dick, T. E., Loparo, K. A. & Jacono, F. J. A framework for patient state tracking by classifying multiscalar physiologic waveform features. *IEEE Trans. Biomed. Eng.* **64**, 2890–2900 (2017).
65. Hackmann, G. et al. Toward a two-tier clinical warning system for hospitalized patients. *AMIA Annu. Symp. Proc.* **2011**, 511–519 (2011).
66. Brown, H., Terrence, J., Vasquez, P., Bates, D. W. & Zimlichman, E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. *Am. J. Med.* **127**, 226–232 (2014).
67. Sutherland, A. et al. Development and validation of a novel molecular biomarker diagnostic test for the early detection of sepsis. *Crit. Care* **15**, R149 (2011).
68. Taneja, I. et al. Combining biomarkers with EMR data to identify patients in different phases of sepsis. *Sci. Rep.* **7**, 10800 (2017).
69. Hassan, U., Zhu, R. & Bashir, R. Multivariate computational analysis of biosensor's data for improved CD64 quantification for sepsis diagnosis. *Lab Chip* **18**, 1231–1240 (2018).
70. Vu, L. et al. Predicting nocturnal hypoglycemia from continuous glucose monitoring data with extended prediction horizon. *AMIA Annu. Symp. Proc.* **2019**, 874–882 (2019).
71. Newman-Toker, D. The team sport of diagnosis: a culture shift can reduce missed diagnoses. *The Healthcare Blog* <https://thehealthcareblog.com/blog/2016/06/15/the-team-sport-of-diagnosis-a-culture-shift-can-reduce-missed-diagnoses/> (2016).
72. Singh, H., Meyer, A. N. D. & Thomas, E. J. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual. Saf.* **23**, 727–731 (2014).
73. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
74. Tirzite, M., Bukovskis, M., Strazda, G., Jurka, N. & Taivans, I. Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis. *J. Breath Res.* **11**, 036009 (2017).
75. Dehghani Soufi, M., Samad-Soltani, T., Shams Vahdati, S. & Rezaei-Hachesu, P. Decision support system for triage management: a hybrid approach using rule-based reasoning and fuzzy logic. *Int. J. Med. Inform.* **114**, 35–44 (2018).
76. Neri, E. & Pinker-Domenig, K. (eds) Special issue "Artificial Intelligence in Diagnostics". https://www.mdpi.com/journal/diagnostics/special_issues/AI_Diagnostics (2020).
77. Dias, R. & Torkamani, A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* **11**, 70 (2019).
78. Bates, D. W., Auerbach, A., Schulam, P., Wright, A. & Saria, S. Reporting and implementing interventions involving machine learning and artificial intelligence. *Ann. Intern. Med.* **172**, S137–S144 (2020).
79. Hernandez-Boussard, T., Bozkurt, S., Ioannidis, J. P. A. & Shah, N. H. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J. Am. Med. Assoc.* **317**, 2011–2015 (2020).
80. Watson, J. et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? *JAMIA Open* **3**, 167–172 (2020).
81. Shaw, J., Rudzicz, F., Jamieson, T. & Goldfarb, A. Artificial intelligence and the implementation challenge. *J. Med. Internet Res.* **21**, e13659 (2019).
82. Bates, D. W., Heitmueller, A., Kakad, M. & Saria, S. Why policymakers should care about "big data" in healthcare. *Health Policy Technol.* **7**, 211–216 (2018).
83. Open Data Science (ODSC). 15 Open datasets for healthcare. *Medium* <https://medium.com/@ODSC/15-open-datasets-for-healthcare-830b19980d9> (2019).
84. AltexSoft. Best public datasets for machine learning and data science: sources and advice on the choice. *AltexSoft* <https://www.altexsoft.com/blog/datascience/best-public-machine-learning-datasets/> (2019).

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Paul Bain for assistance in developing the search strategy for this scoping review. Dr. Syrowatka is supported by a Fellowship Award from the Canadian Institutes of Health Research. This work has been supported by IBM Watson Health (Cambridge, MA), which is not responsible for the content or recommendations made.

AUTHOR CONTRIBUTIONS

D.W.B., D.L., A.S., K.J.T.C., G.P.J., and K.R. were responsible for study conception and design; A.S., M.K., and A.R. reviewed the literature; D.W.B., D.L., A.S., and M.K. analyzed and interpreted the data; D.W.B., D.L., A.S., and M.K. drafted the manuscript, and all the remaining authors have made revisions to it. All the authors have approved the manuscript.

COMPETING INTERESTS

Dr. Bates consults for EarlySense, which makes patient safety monitoring systems. He receives cash compensation from CDI (Negev), Ltd, which is a not-for-profit incubator for health IT startups. He receives equity from ValeraHealth, which makes software to help patients with chronic diseases. He receives equity from Clew, which makes software to support clinical decision-making in intensive care. He receives equity from MDClone, which takes clinical data and produces deidentified versions of it. He receives equity from AESOP, which makes software to reduce medication error rates. Drs. Craig, Jackson, and Rhee are all employed by IBM Watson Health. The other coauthors have no disclosures.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00423-6>.

Correspondence and requests for materials should be addressed to D.W.B.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021