Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review

Ania Syrowatka, Wenyu Song, Mary G Amato, Dinah Foer, Heba Edrees, Zoe Co, Masha Kuznetsova, Sevan Dulgarian, Diane L Seger, Aurélien Simona, Paul A Bain, Gretchen Purcell Jackson, Kyu Rhee, David W Bates

Adverse drug events (ADEs) represent one of the most prevalent types of health-care-related harm, and there is substantial room for improvement in the way that they are currently predicted and detected. We conducted a scoping review to identify key use cases in which artificial intelligence (AI) could be leveraged to reduce the frequency of ADEs. We focused on modern machine learning techniques and natural language processing. 78 articles were included in the scoping review. Studies were heterogeneous and applied various AI techniques covering a wide range of medications and ADEs. We identified several key use cases in which AI could contribute to reducing the frequency and consequences of ADEs, through prediction to prevent ADEs and early detection to mitigate the effects. Most studies (73 [94%] of 78) assessed technical algorithm performance, and few studies evaluated the use of AI in clinical settings. Most articles (58 [74%] of 78) were published within the past 5 years, highlighting an emerging area of study. Availability of new types of data, such as genetic information, and access to unstructured clinical notes might further advance the field.

Introduction

The US National Academy of Medicine has defined an adverse drug event (ADE) as “an injury resulting from medical intervention related to a drug”. These events include non-preventable ADEs (also called adverse drug reactions), and adverse events resulting from medication errors. ADEs represent one of the most important types of health-care-related harm, both inside and outside the hospital, and there is substantial room for improvement in how we predict and detect these events.

The true incidence of ADEs is unknown; these events are often not identified and they are systematically under-reported. An analysis of 28 US state inpatient databases showed that ADEs occurred during 2-1% of all inpatient stays and were present on admission in 5-1% of stays, and management of these ADEs has been estimated to cost US$28 billion annually. However, this analysis was based on documented diagnostic codes, and undoubtedly underestimated true rates. A systematic review of potentially preventable ADEs showed that rates varied widely across inpatient populations, ranging from less than 0-1% to 13-3%, and depended on the approach for event detection, with more cases identified using prospective reporting methods than retrospective or voluntary reporting methods. The occurrence of ADEs in primary care is estimated to be higher than in inpatients. A scoping review showed that rates of ADEs in primary care varied widely according to the study population, setting, medications, and ADEs under study; estimates ranged from 6% in community-dwelling patients prescribed medications for dementia to 81% in patients treated for drug-resistant tuberculosis.

These events are costly and morbid; patients with ADEs have longer hospital stays with greater associated costs and a higher likelihood of mortality than those who do not. About one in three ADEs are considered preventable.

Algorithms and tools based on artificial intelligence (AI)—ie, computer applications that can perform tasks that normally require human intelligence—have the potential to inform clinical decision making in real time to reduce the frequency, duration, and severity of ADEs. Advances in computing power, availability of large-scale patient databases, and machine learning algorithms—ie, algorithms and models that machines can use to learn without explicit instructions—provide the capacity and capability to integrate various data sources and analyse complex inter-relationships between risk factors and outcomes at the point of care. For example, AI could provide timely and accurate predictions of which patients are likely to have ADEs before medications are prescribed. Identification of patients at risk could allow intervention to prevent ADEs; AI could also expand knowledge about which ADEs are preventable and identify new types of ADEs.

The objective of this scoping review was to identify the most promising areas (or key use cases) in which AI could be used to reduce the frequency of harm by providing patient-specific (ie, personalised) predictions to help prevent ADEs or by leading to early detection, mitigating the effects of ADEs. This scoping review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR).

Methods

Search strategy and selection criteria

Two databases (PubMed [National Center for Biotechnology Information] and Embase [Elsevier]) were searched to identify relevant literature published between Jan 1, 1998, and Sept 9, 2020, to correspond with the release of the draft National Academy of Medicine report To Err is Human: Building a Safer Health System. The main concepts of AI, prediction, early detection, and ADEs were mapped to the most relevant controlled...
vocabulary using Medical Subject Headings (MeSH) and free-text terms. An ADE was defined as “an injury resulting from medical intervention related to a drug” and included adverse drug reactions and adverse events related to medication errors. Although the search strategy identified published literature on all ADEs, additional MeSH terms and keywords were added to capture important ADEs: medication errors, allergic reactions, and adverse effects on the blood (eg, thrombocytopenia), as well as drug-induced cardiotoxicity, neurotoxicity, hepatotoxicity, and nephrotoxicity.

The scoping review focused on modern AI techniques, such as neural networks, tree-based algorithms, support vector machines, and natural language processing. The full list of AI models is provided in the appendix (p 1). The review did not include traditional AI approaches, such as logistic or linear regression, due to the vast amount of literature and shift towards using more complex modelling approaches that can integrate large amounts of data from disparate sources to provide more accurate estimations or predictions. Full search strategies are provided in the appendix (pp 2–4).

The scoping review included studies that leveraged AI techniques to develop, validate, or evaluate prediction models to help prevent or manage ADEs, or early detection models to help identify existing ADEs and mitigate the effects, such as severity and duration. The figure depicts the timeline and delineates the difference between prediction and early detection models. Baseline patient characteristics can be integrated into both prediction and early detection models. Use cases for AI to reduce the frequency of ADEs are related to both patient and provider variables and span the full timeline. At the time of treatment decision making, clinical decision support can be provided to clinicians, such as predicting whether the patient is likely to have an ADE and whether there is a preferred treatment given the patient’s risk profile. This type of information can help prevent ADEs. Detection begins once the medication is prescribed, when medication errors can be identified even before medications are taken by the patient. The figure then shows the window of therapeutic response, which also corresponds with the window for occurrence of ADEs. These events can happen immediately, or patients can have delayed reactions. Early detection can help to reduce the severity or duration of ADEs. Both prevention and early detection of ADEs are expected to improve patient outcomes.

Articles were excluded if they were not published in the English language or did not report on original research or a structured review of the literature reported in accordance with PRISMA guidelines. Original research was excluded if the sample size was less than 200 patients, or if the article did not report standard model performance metrics (eg, accuracy or area under the receiver operating curve [AUC-ROC]) or comparisons with a control group for at least one AI model. Studies focused on ADE monitoring or post-marketing surveillance, child or youth populations (ie, aged <18 years), vaccines, dietary supplements, experimental medications not yet approved for use, or recreational use of medications were also excluded. Recreational use was defined as the use of medications that require a prescription but were not prescribed for the individual. Detailed inclusion and exclusion criteria are provided in the appendix (p 1).

**Screening and data abstraction**
Screening was completed in two stages using the Covidence systematic review management program. Articles were screened for relevance on the basis of information provided in the title and abstract and then evaluated for inclusion on the basis of the full text. Two of five reviewers (ASy, HE, ZC, MK, SD) independently screened articles at each stage. Disagreements were
resolved by discussion and consensus between at least two reviewers or by a third independent reviewer.

The following information was abstracted for studies included in the review: citation information; use case domain (ie, prediction, early detection); population under study (ie, region, population description); sample size; AI models and number of variables; medication class; model outcome; type of ADE; performance metrics for the best performing AI model; model validation approach (eg, sample splitting, cross-validation); quality rating; clinical research phase equivalency; and data sources. Data were abstracted by one reviewer and then validated by a second reviewer (ASy, WS, MGA, DF, HE, ZC, MK, SD, DLS, ASi); sections were assigned on the basis of areas of expertise.

The accuracy or AUC-ROC was abstracted for the best performing AI model, since these metrics are most often reported in the medical literature. When this information was not available, other performance metrics were abstracted including sensitivity, specificity, mean error, and area under the precision-recall curve. If modern AI techniques were outperformed by traditional approaches, such as logistic regression, the information was also abstracted for comparison.

The quality of the studies was evaluated on the basis of the study design using a simplified version of the Oxford Centre for Evidence-Based Medicine levels of evidence.11 The clinical research phase equivalency for each study was assessed using Park and colleagues’ 2020 framework to describe the relatively early stages of this type of work, ranging from phase 0 (discovery and invention) to phase 5 (safety and effectiveness).12 Common data sources were
identified and abstracted: health record data, genetic information, administrative health data, publicly available data, and secondary use of research data.

**Data synthesis**

Characteristics of the included studies were summarised. Key use cases for AI to reduce the frequency of ADEs were narratively synthesised. Use cases were identified according to the purpose of the model, either prediction or early detection, and the outcome being modelled. Commonly used AI techniques, main insights including best performing models, and future directions were summarised.

**Results**

**Study characteristics**

From 7218 unique records, 78 articles met the target criteria and were included in the scoping review. A PRISMA flow diagram is presented in the appendix (p 5), and data abstracted from the articles are also provided in the appendix (pp 6–20). The scoping review did not identify any structured reviews focused on the use of AI to reduce the frequency of ADEs. A list of studies excluded at full-text screening is provided in the appendix (pp 21–34).

Characteristics of included studies are summarised in tables 1 and 2. Most studies were published within the past 5 years of our search (2016–20) and were conducted using patient data from North America, Asia, and Europe (table 1).

Studies developed and evaluated a wide range of AI algorithms. Half the studies focused on four medication classes: analgesics, antineoplastics, antibiotics, and anticoagulants. Table 3 shows the breakdown of medication and ADE classes by study sample size.

Most studies (73 [94%] of 78) reported on the development and validation of AI algorithms (equivalent to phases 0 or 1 of clinical trials), and five studies assessed the efficacy and unintended effects of AI solutions (equivalent to phase 2 of clinical trials).12 68 (87%) of these studies used case-control or retrospective cohort study designs corresponding with level 3 quality of evidence (table 1).

**Prediction use cases**

Most studies (67 [86%] of 78) developed, validated, or tested AI-based prediction models to help reduce the frequency of ADEs,\(^3\,7\,13\) and mapped onto four interrelated use cases (table 4).

The first use case was prediction of which patients were likely to have a future ADE, which should help to prevent or effectively manage ADEs.13–46 Most focused on prediction of specific types of ADEs associated with a class of medications; the most common types of ADEs studied were renal (seven [21%] of 34 studies) or cardiovascular (seven [21%]) adverse events, and opioid overdose or opioid use disorder (six [18%]). Notably, only two articles addressed prediction of allergic reactions and focused on prediction of β-lactam hypersensitivity.\(^3\,13\)

15 (44%) of 34 studies assessed the performance of a single AI model.\(^17\,1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\) The other studies compared the performance of multiple models with neural networks and tree-based algorithms demonstrating the best performance based on accuracy and AUC-ROC, or other metrics reported in the studies. One study showed similar performance between federated learning (ie, training algorithms using multiple decentralised databases) and centralised approaches for development of AI-based ADE prediction models.\(^5\)
The second use case was prediction of therapeutic response to medications. The scoping review identified many studies that developed or validated models to predict therapeutic response for which prevention of ADEs in patients not expected to benefit from treatment was stated as a motivation for model development. 18 (27%) of 67 studies addressed this use case and focused on antineoplastics to treat patients with cancer (four [22%] of 18 studies), or antivirals with or without immuno-modulators to treat patients with HIV or hepatitis C (five [28%]). Eight (44%) of 18 studies evaluated a single AI model. The remaining studies compared multiple models; use of support vector machines or tree-based algorithms generally resulted in the most favourable performance.

The third use case was prediction of optimal medication dosing to balance therapeutic benefit with ADE risk related to a specific medication. Several studies (14 [21%] of 67) covered this use case with a focus on anticoagulants (five [36%] of 14 studies), cardiovascular medications (two [14%]), and antineoplastics (two [14%]). Eight studies assessed the performance of a single AI model.
model, with six of these studies using Bayesian estimation.\textsuperscript{41,46,50,73,77} Six studies tested multiple AI-based models, with random forests and ensemble models demonstrating the best performance.

The fourth prediction use case was prediction of the most appropriate treatment option to help guide selection of safe and effective pharmacological therapies. Only a small number of studies reported on this use case\textsuperscript{77–79} (three [4\%] of 67 studies).

Overall, about a quarter of the prediction studies (16 [24\%] of 67) developed algorithms using routinely collected, structured electronic health record (EHR) data including laboratory results, and an additional 21 studies (31\%) developed algorithms using medical records, although it was not clear if the data were electronic. Eight studies (12\%) used administrative health data and 17 studies (25\%) relied on secondary use of research data; only one study integrated information from biosensors.\textsuperscript{9}

Genetic information emerged as a potentially valuable data source. 13 (19\%) of 67 studies included genetic variables to develop prediction models.\textsuperscript{41,46,50,70,74,75} Most studies including genetic information (nine [69\%] of 13) were conducted using secondary research data and compared the performance of multiple AI algorithms. Genetic information was extracted from genetic variant genotyping data and expression profiles.

Integration of genetic and clinical data was one of the most important topics, in which one or several machine learning algorithms were applied to combine clinical and genetic risk factors to improve the performance of ADE prediction models.\textsuperscript{47,76} In eight studies, both genetic and clinical factors were combined as model input variables.\textsuperscript{42,43,47,74,76} In most cases, genetic factors were found to be associated with outcomes and contributed to the model performance. In two studies, the value of adding genetic variables was evaluated.\textsuperscript{42,43} In a study predicting nephrotoxicity, adding genetic variables increased the AUC-ROC from 0·64 to 0·73.\textsuperscript{42} Although genetic features often improved model performance, non-significant results were also reported.\textsuperscript{41}

Multiple studies used machine learning methods to identify genetic factors associated with ADEs, including single nucleotide polymorphisms (SNPs) and related genes.\textsuperscript{41,50,76} Among them, a 12-gene signature was developed using mRNA expression data and a hybrid model to predict tumour recurrence in patients with lung adenocarcinoma.\textsuperscript{49} In another lung cancer study, the association between both SNPs and genes with drug response and toxicity was estimated by logistic regression and neural network models.\textsuperscript{41}

One study developed a cancer (multiple myeloma) treatment learning system to predict patients’ response to treatment using genetic similarities.\textsuperscript{48} Better outcomes (ie, progression-free survival) for bortezomib and lenalidomide were shown, based on patients’ gene expression signatures.

**Early detection use cases**

11 (14\%) of 78 studies used AI-based models to identify ADEs (ten [91\%] of 11)\textsuperscript{48–43} or medication errors (four [36\%] of 11);\textsuperscript{85–90} three studies detected both. Data were obtained from health records for nine studies (noted as EHR data in seven studies) and pharmacy dispensing data for two studies.\textsuperscript{85,90} Two studies tested models in patient care settings and measured prescriber response after receiving clinical decision support alerts identifying potential ADEs.\textsuperscript{48,50} Most studies evaluated models that were designed to identify ADEs or errors in general, with
The included studies were heterogeneous with regard to the types of AI models used and medications and ADEs studied. Most articles only evaluated the technical performance of AI-based algorithms; there were only a few examples of studies describing the clinical evaluation of algorithms or tools. We identified key use cases to guide work in two areas: prediction and early detection.

**Prediction use cases**

The prediction literature identified various mechanisms to reduce the frequency of harm beyond simply predicting which patients are likely to have an ADE; there were four inter-related use cases (table 4). AI-based predictions could play an important role in reducing the overall frequency of ADEs. Identifying patients at higher risk of ADEs could inform dosing adjustments or additional interventions necessary to manage ADEs.

Tree-based methods were widely used and performed well across the prediction use cases. These approaches are particularly well suited for clinical settings, given the relative interpretability compared with deep learning algorithms (a subset of machine learning that generally uses neural networks). Most studies developed predictive models based on structured data routinely documented in health records or administrative health databases. Therefore, most of the necessary information, if not all, would be accessible through EHRs if these algorithms were implemented at the point of care. However, these findings also suggest that important information captured in clinical notes would be missed.

Natural language processing and newer deep learning approaches including transformer neural networks could be leveraged to access data contained in unstructured fields to improve the performance of predictive models. Although considerably more complex than current methods, these types of extraction technologies are being developed by industry and academia, and these solutions are likely to become available to advance research and development of clinical decision support over the next few years. In addition, integration of additional

| Description | Studies, n Common AI approaches* |
|-------------|----------------------------------|
| Prediction  |                                   |
| Predict which patients are likely to have an ADE | 67† Neural network (n=17), random forest (n=17), decision tree (n=11), support vector machine (n=11), gradient boosting machine (n=7), k-nearest neighbours (n=6), Bayesian (n=5) |
| Predict therapeutic response (or non-response) to medications | 34 Neural network (n=10), random forest (n=9), support vector machine (n=9), Bayesian (n=5) |
| Predict optimal medication dose or adaptive dosing | 18 Bayesian (n=6), decision tree (n=5), k-nearest neighbours (n=5), neural network (n=5), random forest (n=4), support vector machine (n=4) |
| Predict most appropriate treatment options | 3 No common models |
| Early detection |                                   |
| Detect ADEs | 11† Neural network (n=6), random forest (n=3) |
| Detect medication prescribing errors | 4 Neural network (n=3) |

ADE=adverse drug event. AI=artificial intelligence. *Applied in three or more studies. †Some studies addressed multiple use cases.

Table 4: Summary of prediction and early detection use cases with commonly used AI approaches

For more on MedAware see https://www.medaware.com/
information (eg, clinical guidelines and comprehensive drug knowledge bases) as inputs into AI algorithms to complement data captured within the EHR could further serve to provide more appropriate recommendations at the point of care.

Genetics can play a pivotal role in patients’ adverse and therapeutic responses to medications. Many genome-wide association studies have been conducted to identify loci and genes associated with ADEs. For example, the P450 genes, a gene family with important implications for drug metabolism, have been assessed in multiple ADE studies.

In our scoping review, multiple studies showed the value of genetic data for ADE prediction. When machine learning-based algorithms were used, they were able to incorporate high-dimensional features to reflect complex patterns in large-scale datasets. Despite the promising potential of integrating genetic information with clinical information, substantial limitations were also noted. Few studies were purely based on genetic variables, suggesting a relatively small degree of variation explained by genetic factors compared with clinical factors. Also, due to the shortage of patient-level genotyping data sources, only a small percentage of ADE studies included genetics.

Literature in the allergy domain was remarkably scarce and is an attractive target for future work. Our scoping review identified just two allergy studies, and both were published in the past 3 years, indicating that the allergy domain might be a developing area of focus for AI application. Antibiotics are the most common allergens in drug hypersensitivity, including in fatal anaphylaxis, and the focus on β-lactam allergy aligns with a high volume of publications in the allergy literature focused on risk stratification of β-lactam allergy. Machine learning applications in allergy remain an unmet need with the potential for substantial impact on patient outcomes. Drug allergy is an area with high patient morbidity motivating further work; it can be conceptualised as the severity of the drug allergy reaction itself—eg, death, anaphylaxis, or severe cutaneous ADEs. Alternatively, it could be viewed through the lens of far more common mild allergic reactions (eg, rashes, hives) for which avoidance of the associated medication (or drug class) can lead to selection of suboptimal or second-line drug choices, resulting in a high risk of morbidity—eg, the increased incidence of *Clostridium difficile* colitis with use of overly broad antibiotics when avoiding β-lactams.

**Early detection use cases**

Several types of AI models have been successful in detecting ADEs when tested using data from health records or pharmacy dispensing data. However, only a few of these systems have been tested in patient care settings. The models show promise in identifying ADEs that providers might otherwise miss and in reducing alert burden seen with rule-based clinical decision support systems that provide alerts for all patients fitting a rule, but that might not be clinically useful in all patients (eg, a drug–drug interaction alert in all patients taking a combination of two medications, even for those that are often recommended to be used together, are being tolerated well by the patient, or are being monitored appropriately). Further testing of these models should be conducted to determine their performance in affecting providers’ decision making when caring for patients.

There is a shortage of studies focused on identifying the best patient contact method, time and frequency, treatment setting, or patient population to screen for potential ADEs. It would be important to know, for example, whether using AI to identify patients to contact via patient portal, telephone call, or text when they are at home identifies ADEs sooner than limiting screening to visits or admissions in health-care settings. AI could also be used to identify the best time intervals to screen for ADEs after starting new medications or when continuing medications, or whether screening at some patient care locations (eg, inpatient, outpatient, or long-term care) is more successful for identification of ADEs than at other locations. More studies are needed to evaluate the added value of using natural language processing to access and integrate data from unstructured clinical notes.

**Challenges of using AI to reduce the frequency of ADEs**

Leveraging AI to reduce the frequency of ADEs is an emerging area of study, and further work is required to ensure that accurate, equitable, and meaningful tools are available at the point of care to inform clinical decision making. Most challenges relating to the implementation of AI solutions in health care apply here as well. High-performing models developed and validated at one or a small number of health-care sites might not translate well to different contexts and will require re-calibration to ensure efficacy and safety. Similarly, models need to be tested and calibrated for population subgroups to ensure that models perform well across all patient groups. To develop these equitable models, developers need access to data with high coverage of the underlying patient population. Prediction and detection of rare ADEs (eg, aplastic anaemia) poses additional challenges requiring careful consideration around the metrics used to evaluate model performance, applying sampling strategies to balance the data, and generating sufficiently large datasets to ensure enough cases for model development.

**Overarching themes and recommendations**

We present main insights and future directions for each domain (panel) and have identified several overarching themes. First, the application of AI to reduce the frequency of ADEs is an emerging area of study with most of the literature published within the past 5 years. Second, despite many AI-based algorithms showing promising performance, most studies were in early
phases of development with few evaluations beyond technical performance. As most AI algorithms that support or advise clinicians are not regulated, evaluations are often rudimentary, which is an ongoing limitation in other areas of health care. It is crucial for high-performing algorithms to be systematically and comprehensively evaluated in prospective trials in clinical settings to show real-world impact and generate the evidence necessary for transparent, safe, and effective implementation. This research requires close collaboration between clinicians and informaticians to make the evaluation process efficient and successful. Third, genetic information was identified as a key data source that has the potential to substantially improve the performance of AI algorithms. With genotyping becoming more commonplace, this type of data should become more accessible over time for both model development and use at the point of care.

Limitations
This study had limitations. The search focused on the concepts of AI, prediction, early detection, and ADEs to identify key use cases for AI to reduce the frequency of ADEs. As such, the review summarised the literature in which ADEs were the focus of the study or in which reduction of ADEs was identified as a key motivation for model development. Related literature not included in this review might be available for some of the use cases.

Due to the large heterogeneity between studies with regard to the types of AI models used as well as the medications and ADEs studied, we were not able to do a formal assessment of the predictive validity of the different AI models. However, our work could inform systematic reviews aimed to answer more focused research questions about the use of AI for specific medications or ADEs and further delineate which AI techniques are most appropriate given different contexts and care settings.

In our study, AI model performance was generally evaluated using accuracy or AUC-ROC, which are most often reported in the medical literature. These metrics might not be the most informative, as they can overestimate performance of algorithms predicting rare ADEs and could account for the similar performances of AI models in studies that evaluated multiple techniques. Other metrics from the computer science literature are better able to account for unbalanced datasets than accuracy or AUC-ROC, such as the area under the precision-recall curve (reported in a few studies included in this scoping review), and we expect that these metrics will become more widely reported in the medical literature over time.

Conclusion
We performed a scoping review, summarised the main insights, and identified several use cases in which AI could contribute to reducing the frequency and consequences of ADEs. Most studies only evaluated technical algorithm performance, and very few studies evaluated the use of AI in clinical settings. Research on predicting allergic reactions was scarce and only a small number of studies incorporated genetic data. Most studies were published in the past 5 years, highlighting an emerging area of study, and we expect many more studies in the next few years. Availability of new types of data and access to unstructured EHR notes might further advance the field.

Contributors
ASy, WS, MGA, DF, DLS, GPJ, KR, and DWB were responsible for study conception and design. ASy and PAB developed the literature searches. ASy, HE, ZC, MK, and SD reviewed the literature. ASy, WS, MGA, DF, and ASy, HE, ZC, MK, and SD were responsible for study implementation. This research requires close collaboration between clinicians and informaticians to make the evaluation process efficient and successful. Third, genetic information was identified as a key data source that has the potential to substantially improve the performance of AI algorithms. With genotyping becoming more commonplace, this type of data should become more accessible over time for both model development and use at the point of care.

Panel: Main insights and future directions for AI to reduce the frequency of ADEs

Prediction
Insights
Most studies developed predictive algorithms and applied a wide range of AI approaches, and focused on specific medication or ADE classes. Tree-based methods performed well across the prediction use cases. A limitation was that algorithms were generally developed using structured data. In many of the studies, genetic factors were applied to improve ADE clinical prediction models; both positive and negative results were reported. One of the remaining challenges is how to identify and select the most relevant genetic variables among large amounts of genetic profile information. Machine learning approaches showed promising progress to address this problem, but are still in the early stages of development.

Opportunities
• Integration of unstructured clinical notes and data from external sources as additional variables to complement structured EHR data to improve the performance of AI-based algorithms
• Inclusion of genetic components in ADE clinical predictions, given the rapid development of causal genetic variant models, such as fine mapping; large-scale clinical biobanks and machine learning algorithms will facilitate this process and make patient genetic information more applicable for clinical decisions to prevent ADEs than is currently possible
• More studies are needed in the fields of EHR-derived ADE genetic association models and machine learning-based clinical and genetic integration systems

Early detection
Insights
Several different AI approaches using data (mostly structured, some unstructured) from EHRs and pharmacy dispensing data showed success in identifying ADEs and prescribing errors. These models have the potential to detect ADEs earlier and more accurately than current methods, and could reduce alert fatigue resulting from too many alerts presented to prescribers using current rule-based alert systems.

Opportunities
• Additional studies of specific models to identify the best AI methods to detect ADEs and errors
• Prospective studies using AI models in patient care settings
• Attempts to identify the best patient population, treatment setting, patient contact method, and time and frequency to screen for ADEs
• Comparisons of benefits of safety outcomes between systems focused on detecting ADEs for high-risk medications, specific ADEs, or general ADE detection

ADE=adverse drug event. AI=artificial intelligence. EHR=electronic health record.
HE, ZC, SD, DLS, and ASi abstracted and verified the data. ASy, WS, MGA, and DF analysed and interpreted the data. All authors had full access to all the data in the study. ASy, WS, MGA, DF, and DSB drafted the original manuscript, and all co-authors reviewed the draft and provided critical feedback. All authors contributed to and approved the final manuscript.

Declaration of interests
ASy, WS, MGA, DF, HE, ZC, SD, and DSB received salary support from a grant funded by IBM Watson Health. DSB has received research support and consults for EarlySense, which makes patient safety monitoring systems. He receives cash compensation from CDI (Negev), which is a not-for-profit incubator for health IT startups. He receives equity from Valera Health, which makes software to help patients with chronic diseases. Clew, which makes software to support clinical decision making in intensive care, and MDClose, which takes clinical data and produces deidentified versions of it. He consults for and receives equity from AESOP, which makes software to reduce medication error rates, and FeelBetter. He has received research support from MedAware. GP is employed by IBM Watson Health, and her compensation includes salary and equity. KR was employed by IBM Watson Health, and is employed by CVS Health; his compensation from both IBM and CVS Health includes salary and equity. All other authors declare no competing interests.

Acknowledgments
ASy is supported by a Fellowship Award from the Canadian Institutes of Health Research. ASi is supported by a Fellowship Award from the Geneva University Hospitals. This work has been supported by IBM Watson Health (Cambridge, MA), which is not responsible for the content or recommendations made.

Editorial note: the Lancet Group takes a neutral position with respect to territorial claims in published tables.

References
1. Kohr LT, Corrigan JM, Donaldson MS. To err is human: Building a safer health system. Washington, DC: Institute of Medicine of the National Academy of Sciences, 1999.
2. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the US Food and Drug Administration ADVERSE EVENT REPORTING SYSTEM (FAERS). Expert Opin Drug Saf 2017; 16: 761–67.
3. Weiss AJ, Freeman WJ, Heslin KC, Barrett ML. Adverse drug events in US hospitals, 2000 versus 2014. HUCP statistical brief #234. Rockville, MD: Agency for Healthcare Research and Quality, 2018.
4. Wolfe D, Yazdi F, Kanji S, et al. Incidence, causes, and consequences of preventable adverse drug reactions occurring in inpatients: a systematic review of systematic reviews. PLoS One 2018; 13: e0205426.
5. Khalil H, Huang C. Adverse drug reactions in primary care: a scoping review. BMC Health Serv Res 2020; 20: 5.
6. Poudel DR, Acharya P, Ghimire S, Dhitai R, Bharati R. Burden of hospitalizations related to adverse drug events in the USA: a retrospective analysis from large inpatient database. Pharmacoepidemiol Drug Saf 2017; 26: 635–41.
7. Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. JAMA 1995; 274: 29–34.
8. Bates DW, Auerbach P, Schulam P, Wright A, Saria S. Reporting and implementing interventions involving machine learning and artificial intelligence. Ann Intern Med 2020; 172 (suppl 1): S137–44.
9. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018; 169: 467–73.
10. Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 6: e1000097.
11. Centre for Evidence-Based Medicine. Oxford Centre for Evidence-Based Medicine: levels of evidence (March 2009). 2009. https://www.cebm.net/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 (accessed Nov 1, 2020).
33 Zasowski EJ, Murray KP, Trinh TD, et al. Identification of vancomycin exposure-toxicity thresholds in hospitalized patients receiving intravenous vancomycin. Antimicrob Agents Chemother 2017; 62: 62.

34 Imari S, Takekuma Y, Kashiwagi H, et al. Validation of the usefulness of artificial neural networks for risk prediction of adverse drug reactions used for individual patients in clinical practice. PLoS One 2020; 15: e0236789.

35 Dong X, Rashidian S, Wang Y, et al. Machine learning based opioid overdose prediction using electronic health records. AMIA Annu Symp Proc 2020; 2019: 389–98.

36 Ellis RJ, Wang Z, Genes N, Ma’ayan A. Predicting opioid dependence from electronic health records with machine learning. BioData Min 2019; 12: 3.

37 Hastings JS, Howison M, Inman SE. Predicting high-risk opioid prescriptions before they are given. Proc Natl Acad Sci USA 2020; 117: 1885–91.

38 Lo-Cigaran WH, Huang JL, Zhang HH, et al. Evaluation of machine-learning algorithms for predicting opioid overdose risk among Medicare beneficiaries with opioid prescriptions. JAMA Netw Open 2019; 2: e190968.

39 Lo-Cigaran WH, Huang JL, Zhang HH, et al. Using machine learning to predict risk of incident opioid use disorder among fee-for-service Medicare beneficiaries: a prognostic study. PLoS One 2020; 15: e0215981.

40 Li X, Yu S, Zhang Z, et al. Predictive modeling of hypoglycemia for clinical decision support in evaluating outpatients with diabetes mellitus. Curr Med Res Opin 2019; 35: 1129–37.

41 Huang K, Luo A, Li X, Li S, Wang S. Chemotherapy-induced neuropathy during adjuvant treatment for cervical cancer patients: development and validation of a prediction model. Int J Clin Exp Med 2015; 8: 10835–44.

42 Garcia SL, Laurinen J, Zhang Z, et al. Prediction of nephrotoxicity associated with cisplatin-based chemotherapy in testicular cancer patients. JNCI Cancer Spectr 2020; 4: pkaa032.

43 Harrison RNS, Gaughran F, Murray RM, et al. Development of multivariable models to predict change in body mass index within a clinical trial population of psychotic individuals. Sci Rep 2017; 7: 14738.

44 Lan TH, Loh FW, Wu MS, et al. Performance of a neuro-fuzzy model in predicting weight changes of chronic schizophrenic patients exposed to antipsychotics. Mol Psychiatry 2008; 13: 1129–37.

45 Yin JY, Li X, Li XP, et al. Prediction models for platinum-based chemotherapy response and toxicity in advanced NSCLC patients. Cancer Lett 2016; 377: 65–73.

46 Nickerson P, Tighe P, Shickel B, Rashidi P. Deep neural network architectures for forecasting analgesic response. Ann Int Conf IEEE Eng Med Biol Soc 2016; 2016: 2966–69.

47 Ke WS, Hwang Y, Lin E. Pharmacogenomics of drug efficacy in the interferon treatment of chronic hepatitis C using classification algorithms. Adv Appl Bioinform Chem 2010; 3: 39–44.

48 Ubel J, Sonnefeld P, van Beers EH, Broijl A, van Vlet MH, de Riddler J. Predicting treatment benefit in multiple myeloma through simulation of alternative treatment effects. Nat Commun 2018; 9: 2941.

49 Wan YW, Sabbagh E, Raese R, et al. Hybrid models identified a 12-gene signature for lung cancer prognosis and chemoresponse prediction. PLoS One 2019; 14: e12222.

50 Lin E, Hwang Y. A support vector machine approach to assess drug efficacy of interferon-α and ribavirin combination therapy. Mol Diagn Ther 2008; 12: 219–23.

51 Gudin J, Mavroudi E, Korfiati A, Theofilatos K, Dietze D, Hurwitz P. Reducing opioid prescriptions by identifying responders on topical analgesic treatment using an individualized medicine and predictive analytics approach. J Pain Res 2020; 13: 1253–66.

52 Sauer CM, Susson D, Paik KE, et al. Feature selection and prediction of treatment failure in tuberculosis. PLoS One 2018; 13: e0207491.

53 Almas M, Parsons B, Whalen E. Prediction of therapeutic response to pregabalin in subjects with neuropathic pain. Curr Med Res Opin 2018; 34: 2041–52.

54 Murphy JE, Arbabzadeh K, Nigbor C, Storlie CB, McCoy RG. Stacked classifiers for individualized prediction of glycemic control following initiation of metformin therapy in type 2 diabetes. Comput Biol Med 2018; 103: 109–15.

55 Ng T, Chew L, Yap CW. A clinical decision support tool to predict survival in cancer patients beyond 120 days after palliative chemotherapy. J Palliat Med 2012; 15: 863–69.

56 Ekenpong MF, Etebolong PI, Jackson TC. Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy. Helviva 2019; 5: e02080.

57 Hatzigeke G, Mathur M, Gilbert L, et al. Neural network-longitudinal assessment of the Electronic Anti-Retroviral Therapy (EARTH) cohort to follow response to HIV-treatment. AMIA Annu Symp Proc 2005; 2005: 301–05.

58 Duan T, Raiipurkar P, Laird D, Ng AE, Basu S. Clinical value of predicting individual treatment effects for intensive blood pressure therapy. Circ Cardiovasc Qual Outcomes 2019; 12: e005010.

59 Wu W, Bang S, Bleecker ER, et al. Multiview cluster analysis identifies variable corticosteroid responsiveness associated with severe asthma. Am J Respir Crit Care Med 2019; 199: 1358–66.

60 Waljee AK, Lipsan R, Wiitala WL, et al. Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning. Inflamm Bowel Dis 2017; 24: 45–51.

61 Maiellaro PA, Cozzolongo R, Marino P. Artificial neural networks for the prediction of response to interferon plus ribavirin treatment in patients with chronic hepatitis C. Curr Pharm Des 2004; 10: 2301–981.

62 Sheyn D, Ju M, Zhang S, et al. Development and validation of a machine learning algorithm for predicting response to anticholinergic medications for overactive bladder syndrome. Obstet Gynecol 2019; 134: 946–57.

63 Preston SL, Druzano GL, Berman AL, et al. Levofloxacin population pharmacokinetics and creation of a demographic model for prediction of individual drug clearance in patients with serious community-acquired infection. Antimicrob Agents Chemother 1998; 42: 1098–104.

64 Hu YH, Wu F, Lo CI, Tai CT. Predicting warfarin dosage from clinical data: a supervised learning approach. Artif Intell Med 2012; 56: 27–34.

65 Liu KE, Lo CI, Hu YH. Improvement of adequate use of warfarin for the elderly using decision tree-based approaches. Methods Inf Med 2014; 53: 47–53.

66 Bondareva IB, Jelliffe RW, Andreeva OV, Bondareva KI. Predictability of individualized dosage regimens of carbamazepine and valproate mono- and combination therapy. J Clin Pharm Ther 2011; 36: 625–36.

67 Kraff S, Lindauer A, Joerger M, Salamone SJ, Jaeaude U. Excel-based tool for pharmacokinetically guided dose adjustment of paclitaxel. Ther Drug Monit 2015; 37: 725–32.

68 Hu YH, Tai CI, Tsai CF, Huang MW. Improvement of adequate digoxin dosage: An application of machine learning approach. J Healthc Eng 2018; 2018: 3948245.

69 Levy AE, Biswas M, Weber R, et al. Applications of machine learning in decision analysis for dose management for dofetilide. PLoS One 2019; 14: e0227124.

70 Escandell-Montero P, Chermisi M, Martinez-Martinez JM, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. Artif Intell Med 2019; 62: 47–60.

71 Marquet P, Bedu A, Monchaud C, et al. Pharmacokinetic therapeutic drug monitoring of Advagraf in more than 500 adult renal transplant patients, using an expert system online. Ther Drug Monit 2018; 40: 285–91.

72 Chen SS, Zaborek NA, Doublady AR, et al. Optimizing levodopa dose adjustment after thyroideectomy with a decision tree. J Surg Res 2019; 244: 102–06.

73 Kraff S, Newuweber AJ, Mathiessen RH, et al. Pharmacokinetically based dosing of weekly paclitaxel to reduce drug-related neurotoxicity based on a single sample strategy. Cancer Chemother Pharmacol 2015; 75: 975–83.

74 Ma Z, Wang P, Gao Z, Wang R, Kahligh K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. PLoS One 2018; 13: e0205872.

75 Ozatan SM, Temizel TT, Erdem SR, Ozzer M. A Bayesian estimation framework for pharmacogenomics driven warfarin dosing: a comparative study. IEEE J Biomed Health Inform 2015; 19: 1724–31.
Review