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

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

How to cite

SYROWATKA, Ania et al. Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review. In: The Lancet. Digital health, 2022, vol. 4, n° 2, p. e137–e148. doi: 10.1016/S2589-7500(21)00229-6

This publication URL: https://archive-ouverte.unige.ch/unige:166052

Publication DOI: 10.1016/S2589-7500(21)00229-6

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.

Lancet Digit Health 2022; 4: e137–48

Published Online November 23, 2021 https://doi.org/10.1016/S2589-7500(21)00229-6

Division of General Internal Medicine (A Syrowatka PhD, W Song PhD, M G Amato PharmD, D Foer MD, H Edrees PharmD, Z Co BS, S Dulgarian BS, D L Seger RPh, A Simona MD, Prof D W Bates MD)**, and Division of Allergy and Clinical Immunology** (D Foer)**, Brigham and Women's Hospital, Boston, MA, USA; Department of Medicine** (A Syrowatka, W Song, D Foer, A Simona, Prof D W Bates) **and Countway Library of Medicine** (P A Bain PhD)**, Harvard Medical School, Boston, MA, USA; Massachusetts College of Pharmacy and Health Sciences, Boston, MA, USA** (M G Amato, H Edrees)**; Harvard Business School, Boston, MA, USA** (M Kuznetsova MPH)**; IBM Watson Health, Cambridge, MA, USA** (G Purcell Jackson MD, K Rhee MD)**; Department of Pediatric Surgery, Vanderbilt University Medical Center, Nashville, TN, USA** (G Purcell Jackson)**; CVS Health, Wellesley Hills, MA, USA** (K Rhee)**; Harvard T H Chan School of Public Health, Boston, MA, USA** (Prof D W Bates)

Correspondence to: Dr Ania Syrowatka, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02120, USA **asyrowatka@bwh.harvard.edu**

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".[1] 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.[2] 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.[3] 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.[4] 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.[5]

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.[6] About one in three ADEs are considered preventable.[7]

Algorithms and tools based on artificial intelligence (AI)—ie, computer applications that can perform tasks that normally require human intelligence[8]—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[8]—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).[9]

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*.[1] The main concepts of AI, prediction, early detection, and ADEs were mapped to the most relevant controlled

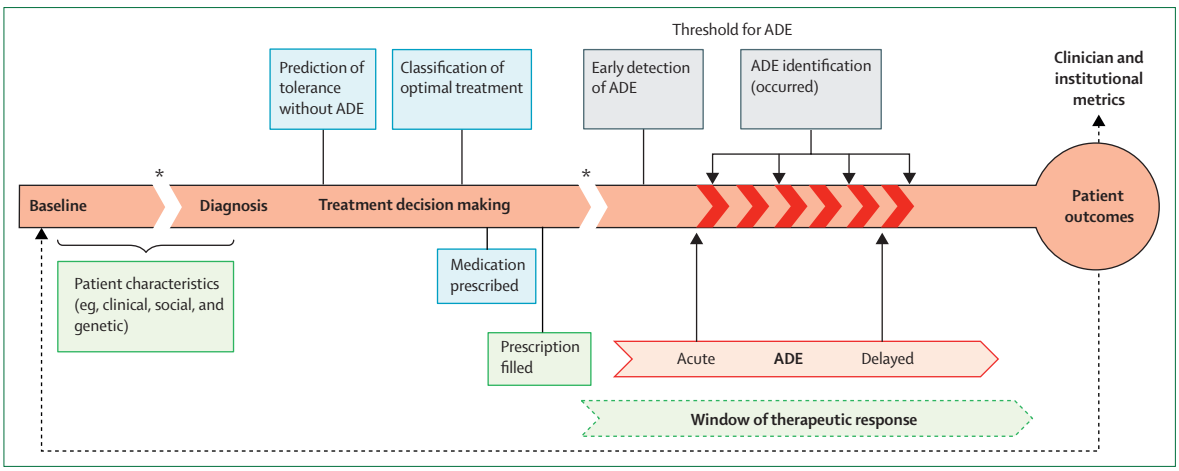

Figure: **Timeline for prediction and early detection of ADEs**
The green boxes represent patient-centred variables, the blue boxes represent provider-centred variables, and the black–grey boxes represent the threshold for ADEs. Asterisks indicate a disontinuity in time. ADE=adverse drug event.

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"[1] 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).

See Online for appendix

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.[9,10] 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

For more on the **Covidence systematic review management program** see https://www.covidence.org/

	Studies (n=78)
Domain	
Prediction	67 (86%)
Early detection	11 (14%)
Publication year	
1998–2005	3 (4%)
2006–10	7 (9%)
2011–15	10 (13%)
2016–20	58 (74%)
Regions and countries	
North America	33 (42%)
USA	33 (42%)
Europe	16 (21%)
France	3 (4%)
Italy	3 (4%)
Netherlands	2 (3%)
Other	8 (10%)
Asia	21 (27%)
Taiwan	10 (13%)
Japan	4 (5%)
China	3 (4%)
South Korea	2 (3%)
Other	2 (3%)
Middle East	2 (3%)
Israel	1 (1%)
Turkey	1 (1%)
Oceania	1 (1%)
Australia	1 (1%)
Africa	1 (1%)
Nigeria	1 (1%)
Multiple countries	4 (5%)
Sample size	
200–500	22 (28%)
501–1000	13 (17%)
1001–10 000	19 (24%)
10 001–100 000	14 (18%)
100 001–1 000 000	7 (9%)
>1 000 000*	3 (4%)

(Table 1 continues in next column)

(Continued from previous column)

	Studies (n=78)
Outcome†	
Adverse drug event	45 (58%)
Cardiovascular	7 (9%)
Renal	7 (9%)
Opioid overdose or opioid use disorder	6 (8%)
Metabolic or endocrine	3 (4%)
Abnormal blood count	3 (4%)
Allergic reaction	2 (3%)
Weight gain	2 (3%)
Nervous system	1 (1%)
Fall with injury	1 (1%)
Genitourinary	1 (1%)
Haemorrhage	1 (1%)
Multiple adverse drug events	3 (4%)
Not specified	8 (10%)
Treatment response	18 (23%)
Optimal dose	14 (18%)
Most appropriate treatment	3 (4%)
AI model†	
Neural network	36 (46%)
Random forest	31 (40%)
Support vector machine	24 (31%)
Decision tree	19 (24%)
Bayesian	19 (24%)
K-nearest neighbours	12 (15%)
Gradient boosting machine	11 (14%)
Other‡	19 (24%)
Medication classification†	
Analgesics	13 (17%)
Narcotic	9 (12%)
Non-narcotic	4 (5%)
Antineoplastics	12 (15%)
Antibiotics	9 (12%)

(Table 1 continues in next column)

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

	Studies (n=78)
(Continued from previous column)	
Anticoagulants	7 (9%)
Antivirals	6 (8%)
Cardiovascular	4 (5%)
Antipsychotics	4 (5%)
Anaesthetics	4 (5%)
Glucocorticoids	3 (4%)
Immunomodulators	3 (4%)
Anticonvulsants	2 (3%)
Immunosuppressants	2 (3%)
Antihyperglycaemics	2 (3%)
Multiple medication classes	6 (8%)
Not specified	5 (6%)
Other§	4 (5%)
Data source†	
Health records	46 (59%)
Electronic	23 (29%)
Unclear whether electronic	23 (29%)
Secondary use of research data	17 (22%)
Genetic data	13 (17%)
Publicly available dataset	12 (15%)
Administrative health data	10 (13%)
Biosensor data	1 (1%)
Quality assessment (scale: levels 1 to 5)[11]	
Level 2: well designed controlled trial without randomisation or prospective comparative cohort trial	9 (12%)
Level 3: case-control study or retrospective cohort study	68 (87%)
Level 4: case series or cross-sectional study	1 (1%)
Equivalent clinical study phase (scale: phases 1 to 5)[12]	
Phase 0: discovery and invention	40 (51%)
Phase 1: technical performance and safety	33 (42%)
Phase 2: efficacy and side-effects	5 (6%)

Data are n (%), with all 78 studies included in the denominator in every case. AI=artificial intelligence. *Includes sample sizes of 1 247 722, 1 540 732, and 1 807 159 patients. †Percentages add to more than 100; some studies cover multiple categories. ‡Includes principal component analysis (n=3), k-means clustering (n=3), adaptive boosting (n=3), ensemble techniques (n=2), natural language processing (n=2), reinforcement learning (n=2), extremely randomised trees (n=1), simulated treatment learning (n=1), kernel machine learning (n=1), and bagging (n=1). §Includes urinary antispasmodic smooth muscle relaxants (n=1), contrast agents (n=1), thyroid hormones (n=1), and colony stimulating factors (n=1).

Table 1: **Characteristics of studies included in the scoping review**

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,[13–79] 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.[14,15]

15 (44%) of 34 studies assessed the performance of a single AI model.[13,14,16,17,20,26,30–32,34,36,37,41,42,44] 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.[27]

	Medication class (name)*	Studies (n=45)
Cardiovascular		
Venous thromboembolism	Antineoplastics (platinum compounds, fluoropyrimidines, anthracyclines, taxanes, paclitaxel, bevacizumab, gemcitabine, irinotecan, pemetrexed, herceptin, anti-tyrosine kinase, aromatase inhibitors)	2 (4%)
Hypotension	Anaesthetics and analgesics (sevoflurane, propofol, fentanyl, remifentanil, midazolam, etomidate, bupivacaine)	3 (7%)
Cardiovascular diseases	Analgesics: non-narcotic (osteoarthritis medications including NSAIDs)	1 (2%)
Arrythmia or dysrhythmia, heart failure	Antineoplastics (methotrexate, cyclophosphamide, carboplatin, others not specified)	1 (2%)
Renal		
Acute kidney injury	Antibiotics (vancomycin; n=4), antineoplastics (cisplatin; n=1), and multiple classes or not specified (n=1)	6 (13%)
Acute tubular necrosis	Contrast agents (medications not specified)	1 (2%)
Metabolic or endocrine		
Hepatotoxicity (jaundice, elevated liver function test)	Analgesics: non-narcotic (acetaminophen [paracetamol])	1 (2%)
Hypoglycaemia	Antihyperglycaemics (acarbose, acetohexamide, alogliptin, canagliflozin, chlorpropamide, colesevelam, dapagliflozin, glibenclamide, glimepiride, glipizide, linagliptin, miglitol, metformin, nateglinide, pioglitazone, repaglinide, rosiglitazone, saxagliptin, sitagliptin, tolazamide, voglibose)	1 (2%)
Metabolic syndrome	Antipsychotics (risperidone, olanzapine, clozapine)	1 (2%)
Opioid overdose or opioid use disorder		
Opioid abuse, overdose, poisoning, or dependence	Analgesics: narcotic (opioids)	6 (13%)
Abnormal blood count		
Low white blood count, neutropenia	Antineoplastics (platinum + taxanes, platinum + irinotecan; n=1) and antivirals (ganciclovir; n=1)	2 (4%)
Low platelets, heparin-induced thrombocytopenia	Anticoagulants (heparin)	1 (2%)
Allergic reaction		
Angioedema or anaphylaxis, urticaria, syncope, exanthema, unspecified	Antibiotics: penicillins, cephalosporins, other unspecified β-lactams (n=1), and benzylpenicillin, aminopenicillins, cephalosporins, other unspecified β-lactams (n=1)	2 (4%)
Nervous system		
Myopathy	Multiple classes or not specified	1 (2%)
Weight gain	Antipsychotics (amisulpride, aripiprazole, clozapine, medroxyprogesterone, flupentixol, fluphenazine decanoate, haloperidol, olanzapine, paliperidone, prochlorperazine, promazine, quetiapine, risperidone, trifluoperazine, sulpiride, zuclopenthixol; n=1) and medications not specified (n=1)	2 (4%)
Fall with injury		
Bone fracture	Glucocorticoids (inhaled corticosteroids)	1 (2%)
Genitourinary		
Post-operative urinary retention	Anaesthesia	1 (2%)
Haemorrhage		
Gastrointestinal haemorrhage	Anticoagulants and antiplatelets (warfarin, dabigatran, apixaban, rivaroxaban, clopidogrel, ticlopidine, aspirin, dipyridamole, cilostazol, ticagrelor)	1 (2%)
Multiple adverse drug events		
Adrenocortical insufficiency, aplastic anaemia, polyneuropathy, hypotension, generalised or localised skin eruption, osteoporosis, fetal damage, anaphylactic shock, allergy, vascular complications, infusion complications, cardiovascular related, other	Not specified (n=1), antineoplastics (platinum + gemcitabine, pemetrexed, paclitaxel, vinorelbine, or etoposide; n=1), and cardiovascular (benazepril, captopril, enalapril, lisinopril, labetalol, doxazosin, atenolol, metoprolol, pindolol, propranolol, clonidine, methyldopa, minoxidil, bumetanide, furosemide, reserpine, spironolactone, triamterene, hydrochlorothiazide, metolazone; n=1)	3 (7%)
Not specified	Multiple classes or not specified	8 (18%)

NSAID=non-steroidal anti-inflammatory drug. *The number of studies is presented in parentheses following the medication class, unless the medication class applies to all studies in that row.

Table 2: **Information about studies that focused on prediction or early detection of adverse drug events**

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[45–62] and focused on antineoplastics to treat patients with cancer (four [22%] of 18 studies), or antivirals with or without immunomodulators to treat patients with HIV or hepatitis C (five [28%]). Eight (44%) of 18 studies evaluated a single AI model.[47,49,51,56–58,60,61] 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[63–76] 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

	Medication classes (78 studies)	ADE classes (45 studies)
200–500 participants	Analgesics (n=1); anaesthetics and analgesics (n=1); antibiotics (n=3); anticoagulants (n=2); antineoplastics (n=6); antipsychotics (n=3); antivirals (n=1); cardiovascular (n=2); immunomodulators and antivirals (n=1); thyroid hormones (n=1); multiple classes or not specified (n=1)	Abnormal blood count (n=2); cardiovascular (n=1); metabolic or endocrine (n=1); renal (n=3); weight gain (n=2); not specified (n=1)
501–1000 participants	Analgesics (n=1); antibiotics (n=2); anticoagulants (n=1); antineoplastics (n=3); antivirals (n=1); glucocorticoids (n=1); immunomodulators and antivirals (n=2); immunosuppressants (n=1); urinary antispasmodic smooth muscle relaxants (n=1)	Cardiovascular (n=1); renal (n=1); multiple ADEs (n=1)
1001–10 000 participants	Analgesics (n=2); anaesthetics (n=1); antibiotics (n=4); anticoagulants (n=3); anticonvulsants (n=2); antineoplastics (n=2); antivirals (n=1); cardiovascular (n=1); colony stimulating factors (n=1); contrast agents (n=1); multiple classes or not specified (n=1)	Abnormal blood count (n=1); allergic reaction (n=2); cardiovascular (n=3); metabolic or endocrine (n=1); renal (n=2); multiple ADEs (n=1); not specified (n=1)
10 001–100 000 participants	Analgesics (n=2); anaesthetics and analgesics (n=2); anticoagulants (n=1); antihyperglycaemics (n=2); antineoplastics (n=1); cardiovascular (n=1); glucocorticoids (n=1); glucocorticoids and immunosuppressants (n=1); multiple classes or not specified (n=3)	Cardiovascular (n=2); fall with injury (n=1); genitourinary (n=1); haemorrhage (n=1); metabolic or endocrine (n=1); opioid overdose or opioid use disorder (n=2); renal (n=1); multiple ADEs (n=1); not specified (n=1)
100 001–1 000 000 participants	Analgesics (n=3); multiple classes or not specified (n=4)	Opioid overdose or opioid use disorder (n=3); not specified (n=4)
>1 000 000 participants	Analgesics and antipsychotics (n=1); multiple classes or not specified (n=2)	Nervous system (n=1); opioid overdose or opioid use disorder (n=1); not specified (n=1)

ADE=adverse drug event.

Table 3: **Medication and ADE classes by study sample size**

model, with six of these studies using Bayesian estimation.[63,66,67,71,73,75] 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[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.[18]

Genetic information emerged as a potentially valuable data source. 13 (19%) of 67 studies included genetic variables to develop prediction models.[41–45,47–50,74–77] 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.[41,76] In eight studies, both genetic and clinical factors were combined as model input variables.[42–45,47,74–76] 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.[42,43] In a study predicting nephrotoxicity, adding genetic variables increased the AUC-ROC from 0·64 to 0·73.[42] Although genetic features often improved model performance, non-significant results were also reported.[43]

Multiple studies used machine learning methods to identify genetic factors associated with ADEs, including single nucleotide polymorphisms (SNPs) and related genes.[41,47,50] 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.[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.[45]

One study developed a cancer (multiple myeloma) treatment learning system to predict patients' response to treatment using genetic similarities.[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)[80–89] or medication errors (four [36%] of 11);[87–90] 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.[85,90] Two studies tested models in patient care settings and measured prescriber response after receiving clinical decision support alerts identifying potential ADEs.[81,88] Most studies evaluated models that were designed to identify ADEs or errors in general, with

	Description	Studies, n	Common AI approaches*
Prediction		67†	
Predict which patients are likely to have an ADE	To prevent or manage ADEs	34	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	To limit ADEs caused by medications that will not provide therapeutic benefit to the patient	18	Neural network (n=10), random forest (n=9), support vector machine (n=9), Bayesian (n=5)
Predict optimal medication dose or adaptive dosing	To balance therapeutic benefit with ADE risk	14	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	To guide selection of safe and effective pharmacological therapies	3	No common models
Early detection		11†	
Detect ADEs	To reduce the duration or severity of harm	10	Neural network (n=6), random forest (n=3)
Detect medication prescribing errors	To prevent ADEs	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**

only three studies focused on identifying ADEs for specific medications.[81–83] Most studies tested a single AI model, with only four studies evaluating multiple models.[80,84,85,90]

Three studies used a software system (MedAware) that identified potential medication errors or ADEs in EHRs using several machine learning methods including neural networks.[87–89] This software was designed to alert prescribers about clinical or medication dosage outliers, time-dependent irregularities occurring between medication use and patients' data, and potential drug duplication. It was found to more accurately identify potential ADEs (85% of alerts were considered clinically relevant *vs* 16% with a rule-based legacy system) and reduce alert burden (alerted on 0·4% *vs* 37·1% of orders).[88] Another study used natural language processing to analyse admission history and physical notes, along with medication lists and laboratory results, to alert providers about potential ADEs.[81] Use of the algorithm also resulted in ordering of fewer potential adverse drug reaction-causing medications at any time during the hospital admission compared with a control group (47% *vs* 58%, p<0·001). Fewer adverse drug reaction-causing medications were also ordered within the first 24 h after admission in the intervention group (28% *vs* 39%, p<0·001). Other studies also reported successful detection of ADEs with the various models tested, but variability in study populations and performance of models between studies limited the ability to determine which type of AI model performed best for detecting ADEs.

Discussion

We performed this scoping review to summarise the published literature on the development, validation, and testing of AI-based algorithms and tools that provide patient-specific predictions to prevent ADEs and early detection to mitigate the harmful effects. 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.

For more on **MedAware** see https://www.medaware.com/

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.[91]

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[8]).[92] 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.[93] 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

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.[94] 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.[95,96]

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.[43] 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,[14,15] 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 allergen in drug hypersensitivity, including in fatal anaphylaxis,[97] 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.[98] 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.[99] 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 are 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.

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.

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,

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 DWB 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 DWB received salary support from a grant funded by IBM Watson Health. DWB 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 MDClone, 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. GPJ 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 Kohn 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, 2010 versus 2014. HCUP 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, Dhital 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 A, Schulam P, Wright A, Saria S. Reporting and implementing interventions involving machine learning and artificial intelligence. *Ann Intern Med* 2020; **172** (suppl)**:** 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.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 (accessed Nov 1, 2020).

12 Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. *JAMIA Open* 2020; **3:** 326–31.

13 Imai S, Yamada T, Kasashi K, Ishiguro N, Kobayashi M, Iseki K. Construction of a flow chart-like risk prediction model of ganciclovir-induced neutropaenia including severity grade: a data mining approach using decision tree. *J Clin Pharm Ther* 2019; **44:** 726–34.

14 Chiriac AM, Wang Y, Schrijvers R, et al. Designing predictive models for beta-lactam allergy using the drug allergy and hypersensitivity database. *J Allergy Clin Immunol Pract* 2018; **6:** 139–148.e2.

15 Moreno EM, Moreno V, Laffond E, et al. usefulness of an artificial neural network in the prediction of β-lactam allergy. *J Allergy Clin Immunol Pract* 2020; **8:** 2974–82.e1.

16 Ferroni P, Zanzotto FM, Scarpato N, Riondino S, Guadagni F, Roselli M. Validation of a machine learning approach for venous thromboembolism risk prediction in oncology. *Dis Markers* 2017; **2017:** 8781379.

17 Ferroni P, Zanzotto FM, Scarpato N, et al. Risk assessment for venous thromboembolism in chemotherapy-treated ambulatory cancer patients: a machine learning approach. *Med Decis Making* 2017; **37:** 234–42.

18 Kang AR, Lee J, Jung W, et al. Development of a prediction model for hypotension after induction of anesthesia using machine learning. *PLoS One* 2020; **15:** e0231172.

19 Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology* 2018; **129:** 675–88.

20 Lin CS, Chiu JS, Hsieh MH, Mok MS, Li YC, Chiu HW. Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks. *Comput Methods Programs Biomed* 2008; **92:** 193–97.

21 Liu L, Yu Y, Fei Z, et al. An interpretable boosting model to predict side effects of analgesics for osteoarthritis. *BMC Syst Biol* 2018; **12** (suppl 6)**:** 105.

22 Yang X, Gong Y, Waheed N, et al. Identifying cancer patients at risk for heart failure using machine learning methods. *AMIA Annu Symp Proc* 2020; **2019:** 933–41.

23 Chen YF, Lin CS, Wang KA, et al. Design of a clinical decision support system for fracture prediction using imbalanced dataset. *J Healthc Eng* 2018; **2018:** 9621640.

24 Hung CY, Lin CH, Chang CS, Li JL, Lee CC. Predicting gastrointestinal bleeding events from multimodal in-hospital electronic health records using deep fusion networks. *Annu Int Conf IEEE Eng Med Biol Soc* 2019; **2019:** 2447–50.

25 Speiser JL, Karvellas CJ, Wolf BJ, Chung D, Koch DG, Durkalski VL. Predicting daily outcomes in acetaminophen-induced acute liver failure patients with machine learning techniques. *Comput Methods Programs Biomed* 2019; **175:** 111–20.

26 Chiang WH, Shen L, Li L, Ning X. Drug-drug interaction prediction based on co-medication patterns and graph matching. *Int J Comput Biol Drug Des* 2020; **13:** 36–57.

27 Choudhury O, Park Y, Salonidis T, Gkoulalas-Divanis A, Sylla I, Das AK. Predicting adverse drug reactions on distributed health data using federated learning. *AMIA Annu Symp Proc* 2020; **2019:** 313–22.

28 Valeanu A, Damian C, Marineci CD, Negres S. The development of a scoring and ranking strategy for a patient-tailored adverse drug reaction prediction in polypharmacy. *Sci Rep* 2020; **10:** 9552.

29 Cheng P, Waitman LR, Hu Y, Liu M. Predicting inpatient acute kidney injury over different time horizons: how early and accurate? *AMIA Annu Symp Proc* 2018; **2017:** 565–74.

30 Imai S, Yamada T, Kasashi K, Kobayashi M, Iseki K. Usefulness of a decision tree model for the analysis of adverse drug reactions: evaluation of a risk prediction model of vancomycin-associated nephrotoxicity constructed using a data mining procedure. *J Eval Clin Pract* 2017; **23:** 1240–46.

31 Yin WJ, Yi YH, Guan XF, et al. Preprocedural prediction model for contrast-induced nephropathy patients. *J Am Heart Assoc* 2017; **6:** e004498.

32 Imai S, Yamada T, Kasashi K, Niinuma Y, Kobayashi M, Iseki K. Construction of a risk prediction model of vancomycin-associated nephrotoxicity to be used at the time of initial therapeutic drug monitoring: a data mining analysis using a decision tree model. *J Eval Clin Pract* 2019; **25:** 163–70.

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 Imai 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:** 1917–23.

38 Lo-Ciganic 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-Ciganic 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:** e0235981.

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:** 1885–91.

41 Huang K, Luo A, Li X, Li S, Wang S. Chemotherapy-induced neutropenia 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, Lauritsen 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 EW, 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. *Annu 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 Ubels J, Sonneveld P, van Beers EH, Broijl A, van Vliet MH, de Ridder J. Predicting treatment benefit in multiple myeloma through simulation of alternative treatment effects. *Nat Commun* 2018; **9:** 2943.

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* 2010; **5:** 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 S, 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:** 1255–66.

52 Sauer CM, Sasson 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 Murphree DH, Arabmakki E, Ngufor 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 Ekpenyong ME, Etebong PI, Jackson TC. Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy. *Heliyon* 2019; **5:** e02080.

57 Hatzakis GE, 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, Rajpurkar P, Laird D, Ng AY, 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 response phenotypes in severe asthma. *Am J Respir Crit Care Med* 2019; **199:** 1358–67.

60 Waljee AK, Lipson R, Wiitala WL, et al. Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning. *Inflamm Bowel Dis* 2017; **24:** 45–53.

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:** 2101–09.

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, Drusano 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 CL, Tai CT. Predicting warfarin dosage from clinical data: a supervised learning approach. *Artif Intell Med* 2012; **56:** 27–34.

65 Liu KE, Lo CL, 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, Jaehde U. Excel-based tool for pharmacokinetically guided dose adjustment of paclitaxel. *Ther Drug Monit* 2015; **37:** 725–32.

68 Hu YH, Tai CT, 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:** e0227324.

70 Escandell-Montero P, Chermisi M, Martínez-Martínez JM, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. *Artif Intell Med* 2014; **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, Doubleday AR, et al. Optimizing levothyroxine dose adjustment after thyroidectomy with a decision tree. *J Surg Res* 2019; **244:** 102–06.

73 Kraff S, Nieuweboer AJ, Mathijssen 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, Khalighi K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLoS One* 2018; **13:** e0205872.

75 Öztaner SM, Temizel TT, Erdem SR, Özer M. A Bayesian estimation framework for pharmacogenomics driven warfarin dosing: a comparative study. *IEEE J Biomed Health Inform* 2015; **19:** 1724–33.

76 Pavani A, Naushad SM, Kumar RM, Srinath M, Malempati AR, Kutala VK. Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization. *Pharmacogenomics* 2016; **17:** 121–31.

77 Blangero Y, Rabilloud M, Ecochard R, Subtil F. A Bayesian method to estimate the optimal threshold of a marker used to select patients' treatment. *Stat Methods Med Res* 2020; **29:** 29–43.

78 König R, Cao X, Oswald M, et al. Macrolide combination therapy for patients hospitalised with community-acquired pneumonia? An individualised approach supported by machine learning. *Eur Respir J* 2019; **54:** 54.

79 Desrochers J, Wojciechowski J, Klein-Schwartz W, Gobburu JVS, Gopalakrishnan M. Bayesian forecasting tool to predict the need for antidote in acute acetaminophen overdose. *Pharmacotherapy* 2017; **37:** 916–26.

80 Taft LM, Evans RS, Shyu CR, et al. Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery. *J Biomed Inform* 2009; **42:** 356–64.

81 Smith JC, Chen Q, Denny JC, Roden DM, Johnson KB, Miller RA. Evaluation of a novel system to enhance clinicians' recognition of preadmission adverse drug reactions. *Appl Clin Inform* 2018; **9:** 313–25.

82 Marchetti M, Barelli S, Zermatten MG, et al. Rapid and accurate Bayesian diagnosis of heparin-induced thrombocytopenia. *Blood* 2020; **135:** 1171–84.

83 Lin CC, Bai YM, Chen JY, et al. Easy and low-cost identification of metabolic syndrome in patients treated with second-generation antipsychotics: artificial neural network and logistic regression models. *J Clin Psychiatry* 2010; **71:** 225–34.

84 Jeong E, Park N, Choi Y, Park RW, Yoon D. Machine learning model combining features from algorithms with different analytical methodologies to detect laboratory-event-related adverse drug reaction signals. *PLoS One* 2018; **13:** e0207749.

85 Hoang T, Liu J, Roughead E, Pratt N, Li J. Supervised signal detection for adverse drug reactions in medication dispensing data. *Comput Methods Programs Biomed* 2018; **161:** 25–38.

86 Bagattini F, Karlsson I, Rebane J, Papapetrou P. A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records. *BMC Med Inform Decis Mak* 2019; **19:** 7.

87 Schiff GD, Volk LA, Volodarskaya M, et al. Screening for medication errors using an outlier detection system. *J Am Med Inform Assoc* 2017; **24:** 281–87.

88 Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. *J Am Med Inform Assoc* 2019; **26:** 1560–65.

89 Rozenblum R, Rodriguez-Monguio R, Volk LA, et al. Using a machine learning system to identify and prevent medication prescribing errors: a clinical and cost analysis evaluation. *Jt Comm J Qual Patient Saf* 2020; **46:** 3–10.

90 Hasan S, Duncan GT, Neill DB, Padman R. Automatic detection of omissions in medication lists. *J Am Med Inform Assoc* 2011; **18:** 449–58.

91 Coleman JJ, Pontefract SK. Adverse drug reactions. *Clin Med (Lond)* 2016; **16:** 481–85.

92 Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med Inform Decis Mak* 2019; **19:** 146.

93 Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. *arXiv* 2019; published online April 10. https://arxiv.org/abs/1904.05342 (preprint).

94 Pirmohamed M. Pharmacogenetics and pharmacogenomics. *Br J Clin Pharmacol* 2001; **52:** 345–47.

95 Dean L. Amitriptyline therapy and CYP2D6 and CYP2C19 genotype. In: Pratt VM, Scott SA, Pirmohamed M, et al, eds. Medical genetics summaries. Bethesda, MD: National Center for Biotechnology Information; March 23, 2017.

96 Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 2013; **138:** 103–41.

97 Pawankar R, Canonica GW, Holgate ST, Lockey RF, Blaiss M. World Allergy Organization (WAO) white book on allergy. Milwaukee, WI: World Allergy Organization, 2011.

98 Blumenthal KG, Shenoy ES, Varughese CA, Hurwitz S, Hooper DC, Banerji A. Impact of a clinical guideline for prescribing antibiotics to inpatients reporting penicillin or cephalosporin allergy. *Ann Allergy Asthma Immunol* 2015; **115:** 294–300.e2.

99 Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015; **68:** 25–34.