

Comparing risk prediction models aimed at predicting hospitalizations for adverse drug events in community dwelling older adults: a protocol paper

Michelle S. Keller (✉ michelle.keller@cshs.org)

Cedars-Sinai Medical Center <https://orcid.org/0000-0002-8157-7586>

Nabeel Qureshi

Cedars-Sinai Health System

Elaine Albertson

Cedars-Sinai Health System

Joshua Pevnick

Cedars-Sinai Health System

Nicole Brandt

University of Maryland School of Pharmacy

Alex Bui

David Geffen School of Medicine: University of California Los Angeles David Geffen School of Medicine

Catherine A. Sarkisian



David Geffen School of Medicine: University of California Los Angeles David Geffen School of Medicine

Research Article

Keywords: Drug adverse events, polypharmacy, older adults, medications, primary care

Posted Date: January 18th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2429369/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

The objective of this paper is to describe the creation, validation, and comparison of two risk prediction modeling approaches for community-dwelling older adults to identify individuals at highest risk for adverse drug event-related hospitalizations. One approach will use traditional statistical methods, the second will use a machine learning approach.

Methods

We will construct medication, clinical, health care utilization, and other variables known to be associated with adverse drug event-related hospitalizations. To create the cohort, we will include older adults (≥ 65 years of age) empaneled to a primary care physician within the Cedars-Sinai Health System primary care clinics with polypharmacy (≥ 5 medications) or at least 1 medication commonly implicated in ADEs (certain oral hypoglycemics, anti-coagulants, anti-platelets, and insulins). We will use a Fine-Gray Cox proportional hazards model for one risk modeling approach and DataRobot, a data science and analytics platform, to run and compare several widely used supervised machine learning algorithms, including Random Forest, Support Vector Machine, Extreme Gradient Boosting (XGBoost), Decision Tree, Naïve Bayes, and K-Nearest Neighbors. We will use a variety of metrics to compare model performance and to assess the risk of algorithmic bias.

Discussion

In conclusion, we hope to develop a pragmatic model that can be implemented in the primary care setting to risk stratify older adults to further optimize medication management.

Introduction

The use of unnecessary and/or potentially harmful medications is a serious problem in older adults. Four out of ten adults 65 years and older take medications which do not provide much benefit or have significant side effects.(1) *Polypharmacy* – the daily use of multiple prescription and non-prescription medications to manage several chronic diseases – is a related concern. Forty percent of older American adults take 5 or more medications daily,(2) increasing their risk for adverse drug reactions such as falls, bleeding, hospitalizations, and death.(3) Concerns about polypharmacy and its adverse outcomes has led to the recent development of quality measures such as the 2022 Healthcare Effectiveness Data and Information Set (HEDIS) benzodiazepine deprescribing measure.(4) HEDIS measures are used by more than 90% of U.S. health plans.(5) Inclusion of deprescribing measures in quality measure sets reflects an increasing recognition by health system and insurance leaders of the importance of carefully monitoring potentially inappropriate medication use and implementing interventions to manage this clinical issue.

Deprescribing, or stopping or tapering potentially inappropriate medications, has been hypothesized to improve quality of life, to reduce falls and fall-related hospitalizations, and reduce healthcare utilization in older adults.(6) However, the evidence on deprescribing interventions for community-based older adults has been mixed.(7) A systematic review and meta-analysis found comprehensive medication reviews may result in small reductions in mortality (OR: 0.74, 95% CI: 0.58–0.95), but found no effects from other types of deprescribing interventions,(8) and found no clear delineating of which patients would benefit the most. Additionally, the review found no significant effects of deprescribing interventions on hospitalizations or emergency department (ED) visits – outcomes important to health systems, older adults, and payors.(8)

As deprescribing medications can be time consuming and challenging, researchers have noted an important gap in identifying *which* patients would most benefit from deprescribing.(9, 10) With limited resources available to address polypharmacy among community-dwelling adults in the primary care setting, physicians, pharmacists, and health system managers have to prioritize which patients should receive various intensities of deprescribing interventions.(10) These could include, at the least intensive level, a comprehensive annual medication review by a pharmacist or clinical decision support advising discontinuation, and at the most intensive level, enrollment in a deprescribing clinic with multiple follow-up visits. While prior studies have used markers for identifying older adults at high risk of polypharmacy-related adverse events such as the number of medications taken every day (e.g., ≥ 5 or ≥ 10 medications), adding additional clinical data such as renal, liver, and cognitive function; comorbidities, age, and other risk factors could make for more precise patient selection for deprescribing interventions. Moreover, combining clinical data with social determinants of health, such as dual eligibility for Medicare-Medicaid, Limited English Proficiency (LEP), living status (e.g., living alone), could help to identify older adults with polypharmacy who have additional need for medication-related support.

Risk prediction models in medicine have been used for allocations to a variety of interventions, including models aimed at identifying individuals at higher risk of breast cancer for referral to genetic counseling,(11) screening patients for lung cancer,(12, 13) assessing risk of stroke for provision of lifestyle interventions,(14) targeting patients at high risk of suicide,(15) and identifying hospitalized patients at high risk of delirium for intervention.(16) Several risk prediction models have been developed for predicting adverse drug reactions and hospitalizations for adverse drug events.(17–21) A 2014 systematic review identified four models aimed at predicting adverse drug events (ADEs) or adverse drug reactions (ADRs), finding several limitations of the existing models, including predictor variables not easily quantifiable or defined, low statistical power available to detect events, and no evidence of impact or implementation of these models.(20) Moreover, while identifying ADEs/ADRs is important, these outcomes may be less important to patients, health systems, and payors compared to hospitalizations or ED visits. A review of the literature identified two models for predicting ADR-related hospitalizations among community-dwelling older adults.(17, 18) However, shortcomings of these models include the inclusion of predictors in the model not readily available in the electronic health record (such as functional status and health-related quality of life scales obtained from a clinical trial)(17) and identification of risk factors only among hospitalized community-dwelling patients and not a broader population.(18) Thus, there is a need for a risk prediction model that could be used in the primary care setting for allocation to different intensities of deprescribing interventions which uses outcomes important to patients and payors and employs easily available and updated data. Importantly, measures used in the predictor model should be pragmatic and easily found in administrative data (such as electronic health record data) for reproducibility in other health systems and settings. Moreover, ideally, to facilitate implementation, a practical risk prediction model should incorporate existing pharmacy and healthcare utilization measures that health systems may already be tracking or may use in the future, increasing the likelihood of implementation.

Most risk prediction models have employed traditional statistical methods such as logistic regression to predict the risk of a binary outcome. This approach uses a set of clinical, demographic, and other variables selected based on assumptions and prior literature to predict the outcome. As Goldstein *et al.* have noted, limitations of these approaches include the potential for a lack of linearity between the risk factors and the outcome, which may not be correctly modeled in a simple risk prediction model; heterogeneity of effects due to interactions; and a limit on the number of prediction variables.(22) However, newer approaches to risk prediction have used machine learning methods.(22) Advantages include the ability to include more predictors, to more easily incorporate non-linearity, and to mitigate missing data without imputation.(22) Limitations include the potential for incorporating and amplifying existing racial biases (see below), the lack of an easily clinically interpretable relationship between the predictors and outcome, issues incorporating temporal data, and challenges creating risk scores that can be meaningfully used in clinical applications when using some machine learning approaches.(22) Machine learning models may also not perform any better than models using traditional statistical methods, highlighting the importance of comparing both approaches.

An important consideration in the use of machine learning models is that they may amplify existing racial, ethnic, gender, and other biases.(23) There are known biases in healthcare access and delivery such as differential treatment of pain,(24) timeliness to radiography and surgery,(25, 26) use of physical restraints in the emergency setting,(27)among many other examples. Race corrections in clinical algorithms, such as the estimate glomerular filtration rate (eGFR), have recently been under scrutiny for exacerbating racial and ethnic disparities.(28) Incorporating clinical and healthcare utilization variables into any algorithm – machine learning or otherwise – carries a risk of integrating these existing biases into the algorithm. Thus, integrating a variety of “fairness metrics” can help identify whether different groups are treated equally in the algorithm.(29)

The objective of this paper is to describe the creation, validation, and comparison of two risk prediction modeling approaches for community-dwelling older adults to identify individuals at highest risk for adverse drug event-related hospitalizations. One approach will use traditional statistical methods, the second will use a machine learning approach. We used the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) checklist and the PROBAST (Prediction model Risk Of Bias Assessment Tool), a tool for assessing the risk of bias to guide our methods.(30, 31) If successful, the risk prediction model will be used to guide a deprescribing strategy at the health system level.

Methods

Data sources

For model development, we will include older adults (≥ 65 years of age) empaneled to a primary care physician within the Cedars-Sinai Health System primary care clinics ($\sim 20,000$ patients) with polypharmacy (≥ 5 medications) or at least 1 medication commonly implicated in ADEs (some oral hypoglycemics [sulfonylureas], anti-coagulants [warfarin and direct-acting oral anticoagulants], some anti-platelets [platelet P2Y12 receptor agonists] and insulins).(32–34) We selected to focus on empaneled patients given that the majority of their healthcare utilization would be captured within the health system’s electronic health records. While this may potentially limit generalizability to patients in non-managed care plans, for this study, we opted to prioritize internal validity. We will use data from the electronic health record warehouse, which includes demographics, diagnoses, lab values, and medications. In future studies we will test the external validity of the CSMC-derived model with datasets from other health systems (including those with more non-managed care patients).

Outcome

The primary outcome is a composite measure consisting of any adverse-drug related ED visit or hospital admission in the next year. We used a list of the most common ICD-10 codes used to identify adverse drug events in administrative data from a systematic review by Hohl et al.(35)

Cohort creation:

To create the cohort, we will include older adults (≥ 65 years of age) empaneled to a primary care physician within the Cedars-Sinai Health System primary care clinics with polypharmacy (≥ 5 medications) or at least 1 medication commonly implicated in ADEs (some oral hypoglycemics [sulfonylureas], anti-coagulants [warfarin and direct-acting oral anticoagulants], some anti-platelets [platelet P2Y12 receptor agonists] and insulins).(32–34) We will create several cohort years from 2016–2021. For each year, we will select the closest encounter date on or before January 1 for all patients. For example, for the year 2016, we will select the closest encounter date on or before January 1, 2016. We will obtain diagnoses and clinical values one year prior to this index encounter date from any encounter. To obtain the outcome, we will look one year forward from this index date. We will obtain electronic health records including all encounters at the health system, including outpatient and inpatient encounters.

Logistic regression risk prediction model

Predictive variables for ADR-related hospitalization. We will perform preliminary screening for collinearity using pairwise correlations between all predictors and will drop a predictor if the correlation is greater than 0.75.(36) Following best practices for risk prediction models, we will use multiple imputation for missing data.(36)

Medications and medication-related measures

- We will include medications identified as important based on the Age-Friendly Health Systems (AFHS) Framework and Pharmacy Quality Alliance.(37–39) These medications include those aimed at preventing, treating, and managing dementia, depression, and delirium across settings of care. Medications that should be reviewed as part of the AFHS framework include benzodiazepines, opioids, central nervous system (CNS) depressants, highly anticholinergic medications (e.g., diphenhydramine), prescription and over-the-counter sedatives, muscle relaxants, antipsychotics, and tricyclic antidepressants.(38) We will operationalize these medications in the following ways:
 - For benzodiazepines, we will convert the monthly dose to a diazepam daily equivalent and calculated the mean monthly dose. We will use diazepam dose equivalent conversion factors from Borrelli et al.(40)
 - For opioids, we will use measures from the Pharmacy Quality Alliance,(41) and create dummy variables if patients meet the following:
 - Use of opioids at high dosage in persons without cancer, defined as ≥ 120 morphine milligram equivalents for ≥ 90 consecutive days;
 - Multiple prescribers, defined as receiving opioid prescriptions from ≥ 4 providers and ≥ 4 pharmacies within 180 days (excluding patients with cancer, sickle cell disease, or receiving palliative care);
 - Concurrent use of opioids and benzodiazepines, defined as having ≥ 30 cumulative days of overlapping opioids and benzodiazepines among individuals having ≥ 2 opioid prescriptions and ≥ 2 benzodiazepine prescriptions.
 - For highly anticholinergic medications (e.g., antihistamines, antiparkinsonian agents), we will create a dummy variable defined as concurrent use for ≥ 30 days or ≥ 2 unique anticholinergic medications.(42)
 - For CNS depressants, we will create a dummy variable defined as concurrent use for ≥ 30 days or ≥ 3 unique CNS medications (including: antipsychotics, benzodiazepines and non-benzodiazepine sedative-hypnotics, opioids, selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, tricyclic antidepressants, and antiepileptics).(43)
 - For prescription and over-the-counter sedatives and sleep medications (recorded in the electronic health record), we will create a dummy variable if a patient had ≥ 30 days supply.
 - For muscle relaxants, we will create a dummy variable if a patient had any prescriptions within the last year.(38)
 - Tricyclic antidepressants, we will create a dummy variable if a patient had ≥ 30 days supply within the last year. (38)
 - For antipsychotics, we will create a dummy variable defined as at least one prescription and ≥ 30 days supply for any antipsychotic medication and no diagnosis of schizophrenia, bipolar disorder, Huntington’s disease, or Tourette’s syndrome.(39)
- A variable for the maximum count of all medications (over-the-counter and prescription) at any one encounter.

Demographics

We will include, age, sex (as a biological variable, as there may be sex-based differences in drug metabolism), race, ethnicity, interpreter needed (yes/no), primary language (English vs. not English), marital status (as a proxy for social support), and insurance status (Medicare Advantage, Medicare, or Medicare-Medicaid).

Diagnoses

In addition to including the Charlson Comorbidity Index (CCI) as a cumulative measure of multimorbidity burden,(44) we will include dummy variables for the following clinical diagnoses previously used in ADE-prediction models: depression, osteoarthritis, heart disease, hypertension, hyperlipidemia, cancer, chronic obstructive pulmonary disease, incontinence, congestive heart failure, substance use disorder.(18) We will use documented diagnoses in the electronic health record from the 1 year before the baseline date.

As dementia is often underdiagnosed and underreported, we will define dementia as either a diagnosis of dementia using the Bynum-standard algorithm and/or ≥ 2 prescriptions and ≥ 60 days' supply of a cholinesterase inhibitor or a N-methyl-D-aspartate (NMDA) receptor antagonist.(45) The Bynum-standard algorithm uses the following ICD-10 codes to identify dementia: F01.50-F01.51 (vascular dementia), F02.80-F02.81 (dementia), F03.90-F03.91 (unspecified dementia), F04 (amnesic disorder), G30.0/30.1/30.8/G30.9 (Alzheimer's disease), G31.01 (Pick's disease), G31.09 (frontotemporal dementia), G31.83 (dementia with Lewy bodies), G31.1 (senile degeneration), G31.2 (degeneration of nervous system), R41.81 (age-related cognitive decline).(46)

Clinical values

We will include clinical measures known to be associated with drug metabolism and potential adverse effects, including:

- Depression, as measured by the PHQ-2 and PHQ-9
- Glomerular filtration rate (eGFR),(47) calculated using serum creatinine, serum cystatin C, age, and sex
- Liver function, as calculated via the MELD 3.0 (Model for End-Stage Liver Disease) Score.(48) The MELD is used primarily to stratify individuals based on severity for transplant purposes, though it has been used for other purposes. The score ranges from 6 to 40, with higher numbers associated with higher severity. The score is derived from the following variables: Whether a patient is on dialysis, Creatinine, Bilirubin, Internal Normalized Ratio (INR), Serum Sodium(48)
- Hemoglobin concentration(49)
- Mean and minimum systolic and diastolic blood pressure(50)
- Mean and minimum pulse(50)
- Mean and maximum body temperature(51)

Healthcare utilization

We will also control for healthcare utilization throughout the previous year to account for informed presence bias. We will control for the number of outpatient and inpatient encounters in the last year.(52)

Social determinants of health screening

In January 1, 2021, Cedars-Sinai Medical Center began administering a social determinants of health screener to patients in the acute care setting and in some non-acute settings. We will use variables recorded in the electronic health record that measure social determinants of health, understanding that there may not be data available for all individuals in our sample, including:

- Food insecurity: ("Within the past 12 months, you worried that your food would run out before you got the money to buy more"; response choices: never true, sometimes true, often true, no response)

- Transportation needs (“In the past 12 months, has lack of transportation kept you from medication appointments, from getting medications, getting things needed for daily living?”; response choices: yes, no, no response)
- Financial resource strain (“How hard is it for you to pay for the very basics like food, housing, medical care, and heating?”; response choices: very hard, hard, somewhat hard, not very hard, not hard at all, no response)
- Access to care (“In the last 12 months, have you needed to see a doctor but could not because of cost or insurance issues?”; response choices: hardly ever, some of the time, often, no response)
- Health literacy (“Do you ever need help reading or understanding information about your medical condition?”; response choices: yes, no, no response)
- House stability (“What is your housing situation today?”; response choices: I have housing, I do not have housing, no response. Also, “Are you worried about losing your housing?”; response choices: yes, no, no response)
- Independent living (“Because of a physical, mental, or emotional condition, do you have difficulty doing activities or errands alone such as visiting a doctor’s office or shopping?”; response choices: hardly ever, some of the time, often, chose not to respond)
- Social connections (“How often do you lack companionship?”, “How often do you feel isolated from others?”, “How often do you feel left out?”; response choices: hardly ever, some of the time, often, chose not to respond)

As these variables are not available in every health system and the responses may not be standardized across different health systems, we will estimate models with and without this set of variables to ensure generalizability and potential dissemination of the model. We will also determine the level of missingness of these variables and consider whether multiple imputation is needed.

Statistical analysis. We will first use chi-square or t-tests to compare characteristics of individuals who did and did not experience an ADE-related hospitalization or ED visit within the year.

We will use a Fine-Gray Cox proportional hazards model⁽⁵³⁾ which accounts for competing risks to predict the time to an ADE-related hospitalization or ED visit. Competing risks will include mortality or a non-ADE-related hospitalization/ED visit. We will use the Stata function *stcrreg* function, which allows for multiple-record data (in this case, multiple observations per patient across the different cohort years). We will use robust standard errors.

For validation, we will use a bootstrap validation approach, where we will construct 200 bootstrap datasets by randomly sampling patients with replacement and will fit the model to each bootstrap dataset.⁽⁵⁴⁾ Estimated coefficients will be used to obtain predictions for patients in the original prediction model, which will be used to calculate the calibration slope for the fitted model.⁽⁵⁴⁾

Machine learning risk prediction algorithms

We will use DataRobot, a data science and analytics platform, to run and compare several widely used supervised machine learning algorithms, including Random Forest, Support Vector Machine, Extreme Gradient Boosting (XGBoost), Decision Tree, Naïve Bayes, and K-Nearest Neighbors.⁽⁵⁵⁾ DataRobot is an automated machine learning platform that can rapidly compare model types and tune hyperparameters to optimize prediction accuracy. Automated machine learning with DataRobot, and other platforms, has emerged as part of “state-of-the-art” healthcare analytics (Waring 2020). Prior studies have used DataRobot for healthcare risk prediction.^(56–58) The cohort selection will use the same process as above. We will use the same constructed measures as above, with the exception of the following:

- We will construct several summary-level values (means, minimum, maximum, standard deviation) for laboratory values.
- We will include all diagnoses in the year prior to the index date.

As described above, inadvertent algorithmic bias in ML is a recognized problem. We will assess data and algorithms through population stratification and quantitation of bias following three paths:(59)

- **Dataset representation.** It is important to determine to what extent the data used to train an algorithm comes from a population that is representative for its intended use (in this case, the overall empaneled population in the health system). We will assess statistical distributions, and also determine the levels of associated missingness. If a particular subgroup is underpowered and/or has a significant amount of missing data in each model feature, we will look to apply inverse-odds weighting/transportability as cited above to accordingly compensate.
- **Stratification and subgroup analyses.** We will calculate group benefit equality (GBE), an algorithmic fairness metric that quantifies the rate at which a particular event is predicted to occur within a subgroup compared to the rate at which it actually occurs; and apply the fairness principles which examine the equality of opportunities for various subgroups.(59) The use of the GBE can help identify whether a particular subgroup (e.g., a racial or ethnic subgroup) is being flagged for intervention at a lower rate than the prevalence of the event in the subgroup – in this case, because the predicted probability of ADE-related hospitalization is potentially incorrectly or biasedly under-predicted.
- **Ablation studies.** We will selectively remove and replace variables to ascertain any impact on guiding a predictive model's results, evaluating the relative impact.

From these results and application of the Prediction model Risk Of Bias Assessment Tool (PROBAST),(60) we will determine if our models are prone to potential bias and adapt the models accordingly to address these issues.

Model comparison

We will use a variety of performance metrics to compare the various models. These include:

- the Brier score, or squared-loss error, which takes the difference between the model's predictions and the actual outcomes;
- the misclassification rate;
- the c-statistic, which measures the model's ability to correctly classify those who have or do not experience an ADE-related hospitalization or ED visit based on predicted risk and is analogous to the area under the receiver operating characteristic curve, a plot of sensitivity versus (1–specificity) for a binary outcome at various thresholds;(36)
- Area under the precision-recall curve, which summarizes precision and recall;
- calibration plots where the mean predicted probability of the outcome is plotted against the observed proportion of outcomes for groups of the cohort;(36)
- the Hosmer-Lemeshow test;(36)
- the log loss or logarithmic score;
- the Gini coefficient, which measures a model's discriminatory power.

We will also create a confusion matrix to examine positive and negative predictive values and sensitivity and specificity for each model. Finally, we will conduct error analysis to descriptively examine the errors that each model makes (for example, comparing diagnostic/clinical/demographic characteristics for the patients correctly vs. incorrectly classified in each model).

For the machine learning algorithms, we will also compare variable importance for each final model. We will also examine any potentially bias by examining the proportionate number of people in different racial and ethnic categories identified as higher risk.

Discussion

Deprescribing interventions may require large amounts of time from physicians and pharmacists. Identifying which patients may benefit the most from deprescribing interventions has the potential to improve efficiencies and workflow. We have described a rigorous *a priori* methodology to create and validate a disseminatable tool that can use EHR data to identify older adults most likely to benefit from deprescribing interventions.

Limitations of this approach are similar to other models using administrative data such as electronic health records. For example, there may be miscoding, over-coding, or under-coding of diagnoses, which may introduce bias. For example, it is well known that dementia is under-diagnosed due to structural factors (e.g., lack of knowledge, resources, training, and time among primary care providers to diagnose dementia).(61) Moreover, an important consideration is that access to healthcare may be impeded by structural factors such as lack of transportation, insurance, language issues, discrimination and racism, etc. This issue may result in under-diagnosed conditions and lower quality of care experienced among patients from vulnerable and marginalized patients.(62) These populations may also have fewer encounters or fewer laboratory and test values, which could result in an underestimation of risk among these patients. As a result, it will be important to examine the demographic characteristics of individuals classified in the various models to explore bias in these models.

In conclusion, we hope to develop a pragmatic model that can be implemented in the primary care setting to optimize medication management in older adults.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Cedars-Sinai Medical Center Institutional Review Board.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received funding from the National Institute on Aging (R24AG064025, K01AG076865, K24AG047899). The funding body had no role in the design of the study.

Authors contributions

MSK wrote, edited, and revised the manuscript and obtained the primary funding for the study. NQ, EA, AB, JP, NB, CAS contributed to the study design, reviewed, and revised the manuscript.

Acknowledgments

Not applicable

References

1. Alhawassi TM, Alatawi W, Alwhaibi M. Prevalence of potentially inappropriate medications use among older adults and risk factors using the 2015 American Geriatrics Society Beers criteria. *BMC Geriatrics*. 2019;19(1):154.
2. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in Prescription Drug Use Among Adults in the United States From 1999-2012. *JAMA*. 2015;314(17):1818-30.
3. Wastesson JW, Morin L, Tan ECK, Johnell K. An update on the clinical consequences of polypharmacy in older adults: a narrative review. *Expert Opinion on Drug Safety*. 2018;17(12):1185-96.
4. NCQA. Proposed New Measure for HEDIS®1 MY 2022: Deprescribing of Benzodiazepines in Older Adults (DBO) Washington, DC2021 [Available from: <https://www.ncqa.org/wp-content/uploads/2021/02/05.-DBO.pdf>].
5. Office of Disease Prevention and Health Promotion OotASfH, Office of the Secretary, U.S. Department of Health and Human Services,. Healthcare Effectiveness Data and Information Set (HEDIS) [Available from: <https://health.gov/healthypeople/objectives-and-data/data-sources-and-methods/data-sources/healthcare-effectiveness-data-and-information-set-hedis#:~:text=The%20Healthcare%20Effectiveness%20Data%20and,report%20quality%20results%20using%20HEDIS>].
6. Patel T, Slonim K, Lee L. Use of potentially inappropriate medications among ambulatory home-dwelling elderly patients with dementia: a review of the literature. *Canadian Pharmacists Journal/Revue des Pharmaciens du Canada*. 2017;150(3):169-83.
7. Anderson LJ, Schnipper JL, Nuckols TK, Shane R, Sarkisian C, Le MM, et al. A systematic overview of systematic reviews evaluating interventions addressing polypharmacy. *American Journal of Health-System Pharmacy*. 2019;76(21):1777-87.
8. Bloomfield HE, Greer N, Linsky AM, Bolduc J, Naidl T, Vardeny O, et al. Deprescribing for Community-Dwelling Older Adults: a Systematic Review and Meta-analysis. *Journal of general internal medicine*. 2020:1-10.
9. Thompson W, Reeve E, Moriarty F, Maclure M, Turner J, Steinman MA, et al. Deprescribing: Future directions for research. *Research in Social and Administrative Pharmacy*. 2019;15(6):801-5.
10. Nair NP, Chalmers L, Peterson GM, Bereznicki BJ, Castelino RL, Bereznicki LR. Hospitalization in older patients due to adverse drug reactions—the need for a prediction tool. *Clinical interventions in aging*. 2016;11:497.
11. Domchek SM, Eisen A, Calzone K, Stopfer J, Blackwood A, Weber BL. Application of breast cancer risk prediction models in clinical practice. *Journal of Clinical Oncology*. 2003;21(4):593-601.
12. Tammemägi MC. Application of Risk Prediction Models to Lung Cancer Screening: A Review. *Journal of Thoracic Imaging*. 2015;30(2).
13. Ten Haaf K, Jeon J, Tammemägi MC, Han SS, Kong CY, Plevritis SK, et al. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. *PLoS medicine*. 2017;14(4):e1002277.
14. Lee J-w, Lim H-s, Kim D-w, Shin S-a, Kim J, Yoo B, et al. The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record. *Computer Methods and Programs in Biomedicine*. 2018;153:253-7.
15. Kessler RC, Hwang I, Hoffmire CA, McCarthy JF, Petukhova MV, Rosellini AJ, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *International Journal of Methods in Psychiatric Research*. 2017;26(3):e1575.
16. Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, et al. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. *Journal of the American Medical Informatics Association*. 2020;27(9):1383-92.
17. Müller BS, Uhlmann L, Ihle P, Stock C, von Buedingen F, Beyer M, et al. Development and internal validation of prognostic models to predict negative health outcomes in older patients with multimorbidity and polypharmacy in general practice. *BMJ open*. 2020;10(10):e039747.

18. Parameswaran Nair N, Chalmers L, Connolly M, Bereznicki BJ, Peterson GM, Curtain C, et al. Prediction of hospitalization due to adverse drug reactions in elderly community-dwelling patients (the PADR-EC score). *PLoS One*. 2016;11(10):e0165757.
19. Deawjaroen K, Sillabutra J, Poolsup N, Stewart D, Suksomboon N. Clinical usefulness of prediction tools to identify adult hospitalized patients at risk of drug-related problems: A systematic review of clinical prediction models and risk assessment tools. *British Journal of Clinical Pharmacology*. 2022;88(4):1613-29.
20. Stevenson JM, Williams JL, Burnham TG, Prevost AT, Schiff R, Erskine SD, et al. Predicting adverse drug reactions in older adults; a systematic review of the risk prediction models. *Clinical interventions in aging*. 2014;9:1581.
21. Høj K, Pedersen HS, Lundberg ASB, Bro F, Nielsen LP, Sædder EA. External validation of the Medication Risk Score in polypharmacy patients in general practice: A tool for prioritizing patients at greatest risk of potential drug-related problems. *Basic & Clinical Pharmacology & Toxicology*. 2021;129(4):319-31.
22. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European Heart Journal*. 2017;38(23):1805-14.
23. Kostick-Quenet KM, Cohen IG, Gerke S, Lo B, Antaki J, Movahedi F, et al. Mitigating Racial Bias in Machine Learning. *Journal of Law, Medicine & Ethics*. 2022;50(1):92-100.
24. Aronowitz SV, McDonald CC, Stevens RC, Richmond TS. Mixed studies review of factors influencing receipt of pain treatment by injured black patients. *Journal of Advanced Nursing*. 2020;76(1):34-46.
25. Nayar SK, Marrache M, Ali I, Bressner J, Raad M, Shafiq B, et al. Racial disparity in time to surgery and complications for hip fracture patients. *Clin Orthop Surg*. 2020;12(4):430.
26. Ali I, Vattigunta S, Jang JM, Hannan CV, Ahmed MS, Linton B, et al. Racial disparities are present in the timing of radiographic assessment and surgical treatment of hip fractures. *Clinical orthopaedics and related research*. 2020;478(3):455.
27. Schnitzer K, Merideth F, Macias-Konstantopoulos W, Hayden D, Shtasel D, Bird S. Disparities in Care: The Role of Race on the Utilization of Physical Restraints in the Emergency Setting. *Academic Emergency Medicine*. 2020;27(10):943-50.
28. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight – Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine*. 2020;383(9):874-82.
29. Qi M, Cahan O, Foreman MA, Gruen DM, Das AK, Bennett KP. Quantifying representativeness in randomized clinical trials using machine learning fairness metrics. *JAMIA open*. 2021;4(3):ooab077.
30. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. 2019;170(1):51-8.
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*. 2015;102(3):148-58.
32. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. *New England Journal of Medicine*. 2011;365(21):2002-12.
33. Centers for Disease Control and Prevention. Adverse Drug Events from Specific Medicines 2019 [Available from: <https://www.cdc.gov/medicationsafety/adverse-drug-events-specific-medicines.html>].
34. Budnitz DS, Shehab N, Lovegrove MC, Geller AI, Lind JN, Pollock DA. US Emergency Department Visits Attributed to Medication Harms, 2017-2019. *JAMA*. 2021;326(13):1299-309.
35. Hohl CM, Karpov A, Reddekopp L, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. *Journal of the American Medical Informatics Association*. 2014;21(3):547-57.
36. Grant SW, Collins GS, Nashef SAM. Statistical Primer: developing and validating a risk prediction model†. *European Journal of Cardio-Thoracic Surgery*. 2018;54(2):203-8.

37. Mate KS, Berman A, Laderman M, Kabcenell A, Fulmer T. Creating Age-Friendly Health Systems – A vision for better care of older adults. *Healthcare*. 2018;6(1):4-6.
38. Institute for Healthcare Improvement. *Age-Friendly Health Systems: Measures Guide*. Boston, MA; 2020.
39. Pharmacy Quality Alliance. *PQA Quality Measures Alexandria, VA, USA2022* [Available from: <https://www.pqaalliance.org/pqa-measures>].
40. Borrelli EP, Bratberg J, Hallowell BD, Greaney ML, Kogut SJ. Application of a diazepam milligram equivalency algorithm to assess benzodiazepine dose intensity in Rhode Island in 2018. *Journal of Managed Care & Specialty Pharmacy*. 2022;28(1):58-68.
41. Pharmacy Quality Alliance. *Opioid Measures Alexandria, VA2022* [Available from: <https://www.pqaalliance.org/opioid-measures>].
42. Pharmacy Quality Alliance. *Polypharmacy: use of multiple anticholinergic medications in older adults Alexandria, VA2021* [Available from: <https://www.pharmacyquality.com/wp-content/uploads/2021/05/PQA29Polypharmacy2021.pdf>].
43. Pharmacy Quality Alliance. *Polypharmacy: CNS-Active Medications in Older Adults Alexandria, VA2021* [Available from: <https://www.pharmacyquality.com/wp-content/uploads/2021/05/PQA30Polypharmacy-CNSActive-Medications.pdf>].
44. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40(5):373-83.
45. Reuben DB, Hackbarth AS, Wenger NS, Tan ZS, Jennings LA. An Automated Approach to Identifying Patients with Dementia Using Electronic Medical Records. *Journal of the American Geriatrics Society*. 2017;65(3):658-9.
46. Grodstein F, Chang C-H, Capuano AW, Power MC, Marquez DX, Barnes LL, et al. Identification of Dementia in Recent Medicare Claims Data, Compared With Rigorous Clinical Assessments. *The Journals of Gerontology: Series A*. 2022;77(6):1272-8.
47. Dreisbach AW, Lertora JJJ. The effect of chronic renal failure on drug metabolism and transport. *Expert opinion on drug metabolism & toxicology*. 2008;4(8):1065-74.
48. Kim WR, Mannalithara A, Heimbach JK, Kamath PS, Asrani SK, Biggins SW, et al. MELD 3.0: The Model for End-Stage Liver Disease Updated for the Modern Era. *Gastroenterology*. 2021;161(6):1887-95.e4.
49. Mintzer DM, Billet SN, Chmielewski L. Drug-induced hematologic syndromes. *Advances in hematology*. 2009;2009.
50. Kanjanarat P, Winterstein AG, Johns TE, Hatton RC, Gonzalez-Rothi R, Segal R. Nature of preventable adverse drug events in hospitals: a literature review. *American Journal of Health-System Pharmacy*. 2003;60(17):1750-9.
51. Zopf Y, Rabe C, Neubert A, Hahn EG, Dormann H. Risk Factors Associated with Adverse Drug Reactions Following Hospital Admission. *Drug Safety*. 2008;31(9):789-98.
52. Goldstein BA, Bhavsar NA, Phelan M, Pencina MJ. Controlling for informed presence bias due to the number of health encounters in an electronic health record. *American journal of epidemiology*. 2016;184(11):847-55.
53. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. 1999;94(446):496-509.
54. Pavlou M, Ambler G, Seaman SR, Guttman O, Elliott P, King M, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ : British Medical Journal*. 2015;351:h3868.
55. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. *International journal of medical informatics*. 2017;97:120-7.
56. Ebinger J, Wells M, Ouyang D, Davis T, Kaufman N, Cheng S, et al. A machine learning algorithm predicts duration of hospitalization in COVID-19 patients. *Intelligence-based medicine*. 2021;5:100035.

57. Suzuki S, Yamashita T, Sakama T, Arita T, Yagi N, Otsuka T, et al. Comparison of risk models for mortality and cardiovascular events between machine learning and conventional logistic regression analysis. *PLoS One*. 2019;14(9):e0221911.
58. Muhlestein WE, Akagi DS, Chotai S, Chambless LB. The impact of presurgical comorbidities on discharge disposition and length of hospitalization following craniotomy for brain tumor. *Surgical neurology international*. 2017;8.
59. Gartner J. A New Metric for Quantifying Machine Learning Fairness in Healthcare: ClosedLoop.ai; 2020 [updated March 2, 2020. Available from: <https://www.closedloop.ai/post/a-new-metric-for-quantifying-machine-learning-fairness-in-healthcare>.
60. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine*. 2019;170(1):51-8.
61. Connolly A, Gaehl E, Martin H, Morris J, Purandare N. Underdiagnosis of dementia in primary care: variations in the observed prevalence and comparisons to the expected prevalence. *Aging & mental health*. 2011;15(8):978-84.
62. Katz A, Chateau D, Enns JE, Valdivia J, Taylor C, Walld R, et al. Association of the Social Determinants of Health With Quality of Primary Care. *The Annals of Family Medicine*. 2018;16(3):217.