Predictors of Co-occurring Cardiovascular and Gastrointestinal Disorders among Elderly with Osteoarthritis

Jayeshkumar Patel a,*, Amit Ladani b, Nethra Sambamoorthi c, Traci LeMasters a, Nilanjana Dwibedi a, Usha Sambamoorthi b, c, d, e

a Department of Pharmaceutical Systems and Policy, West Virginia University, Morgantown, USA
b Rheumatology, West Virginia University Medicine, Morgantown, WV, USA
c Masters in Data Science Program, School of Professional Studies, Northwestern University, Chicago, IL, USA
d Department of Pharmacotherapy, University of North Texas Health Sciences Center College of Pharmacy, USA
e Vashisht Professor, Health Disparities, Health Education, Awareness & Research in Disparities Scholar, Texas Center for Health Disparities, HSC, USA

ARTICLE INFO
Keywords:
Osteoarthritis
Cardiovascular disease
Gastrointestinal disorders
Machine learning
Predictive modeling
Random forest
Older adults

ABSTRACT
Objective: To identify the leading predictors of co-occurring cardiovascular or gastrointestinal disorders (CV-GID) in a real-world cohort of elderly with osteoarthritis (OA).

Method: An observational retrospective cohort study using data from Optum’s deidentified Clinfoinformatics® Data Mart was conducted. Elderly with OA were identified in 2015 and were followed for two years to identify co-occurring CV-GID including ischemic heart disease, stroke, heart failure, dyspepsia, gastroesophageal reflux disorder, and peptic ulcer disease. Random Forest (RF) and Partial Dependence Plots (PDP) were used to identify the leading predictors of CV-GID and to examine their associations. Multivariable logistic regression was also used to examine the association of the leading predictors with CV-GID.

Results: Our study cohort consisted of 45,385 elderly with OA (mean age 76.0 years). CV-GID were present in 59% of elderly. Using RF, age was found to be the strongest predictor of CV-GID followed by cardiac arrhythmia, duration of opioid use, number of orthopedist or physical therapy visits, number of intra-articular corticosteroid injections, polypharmacy, duration of non-selective nonsteroidal anti-inflammatory drugs or oral corticosteroids, and hypertension. The PDPs demonstrated that higher age, cardiac arrhythmia, longer durations of opioid or oral corticosteroids, higher number of physical therapy visits or intra-articular corticosteroid use, polypharmacy, and hypertension were associated with a higher risk of CV-GID.

Conclusion: CV-GIDs are common among elderly with OA and can be predicted based on certain clinical factors. Machine learning methods with PDPs can be used to improve the interpretability and inform decision-making.

1. Introduction

Osteoarthritis (OA) is the most common chronic joint disease characterized by joint pain and an increasing prevalence with age [1]. The elderly with OA are at a higher risk for certain chronic conditions compared to those without OA [2]. Of these conditions, cardiovascular disease (CVD) and gastrointestinal (GI) disorders are of particular importance for several reasons. Evidence from recent studies suggest that the risk and mortality of CVD are higher in individuals with OA [3–5]. Both OA and CVD share a few common risk factors including physical inactivity, age, and obesity among others [1,6]. An increasing body of evidence has shown that OA, although traditionally thought to be a noninflammatory condition, is associated with chronic low-grade inflammation, which may increase the risk for CVD [6,7]. Emerging evidence also suggests that an inflammatory response may lead to GI complications such as gastroesophageal reflux disease (GERD) [8]. Furthermore, some pain medications commonly used to treat adults with OA can also lead to an increased risk of CVD and upper GI adverse events (AEs). The CV and GI AEs of non-steroidal anti-inflammatory drugs (NSAIDs) are well documented [9,10]. There is also some evidence on the CV risk of opioid analgesics [11,12].

Published studies, that are not specific to OA, have provided evidence...
on the association of many factors with CVD or GI disorders (CV-GID) using statistical \cite{13,14}, or machine learning methods \cite{15,16}. These factors include biomarkers (e.g. high cholesterol), clinical (e.g. hypertension, diabetes, concomitant medications), and behavioral factors (e.g. smoking, physical inactivity). However, to date, there are no published studies on a comprehensive list of factors associated with CV-GID. The examination of CV-GID is important because of a documented link between these two disorders \cite{17,18}. Several risk factors for CVD have also been found to be associated with an increased risk for GI disorders \cite{19}. Additionally, the use of low-dose aspirin for primary or secondary prevention of CVD may also increase an individual’s risk for GI complications \cite{20,21}. Specifically, the identification of the leading predictors of CV-GID among elderly with OA holds significance because the prevalence of OA rises from under 15% in those younger than 60 to over 30% in individuals over 60 years of age \cite{22}. Older age and greater use of pain medications can also increase the risk for CV-GID \cite{23,24}. Moreover, elderly with OA have higher rates of multimorbidity \cite{25}, including mental health conditions \cite{26} and polypharmacy \cite{27}. These factors have been found to be associated with a higher risk for CV-GID perhaps due to drug-drug, drug-disease, and disease-disease interactions \cite{25,28,29}. The identification and knowledge of the leading predictors of CV-GID can help clinicians and healthcare providers identify certain subgroups of patients requiring closer monitoring or treatment changes.

Therefore, the objective of the current observational retrospective cohort study was to identify the leading predictors of CV-GID in elderly with OA using machine learning approaches. We also used an interpretable machine learning technique to gain insights into the associations of the leading predictors to CV-GID.

2. Methods

2.1. Study design

We employed an observational retrospective cohort study design using data from two years. The study sample consisted of elderly (age $\geq 65$ years) with OA identified during 2015 and followed until the end of 2016.

2.2. Data source

The study cohort was selected from a 10% random sample of all individuals in the Optum De-identified Clininformatics® Data Mart (Optum) \cite{30}. The data consist of commercial and Medicare Advantage claims from enrollees. The data are geographically representative of the commercially insured United States (US) population and contain comprehensive information on demographics, prescription medications, diagnoses, and medical services used by insured individuals. As data are deidentified, informed consent from patients was not required. Clinical characteristics and medications were identified from claims using the International Classification of Diseases-9th and 10th Revision (ICD-9/10) codes, American Hospital Formulary Services (AHFS) codes, Healthcare Common Procedure Coding System (HCPCS) codes, and National Drug Codes (NDC).

2.3. Study cohort

Our study cohort comprised of 45,385 elderly (age $\geq 65$ years) with OA who met the study inclusion criteria: 1) age $\geq 65$ years; 2) one inpatient or two outpatient claims (30 days apart) with OA diagnosis codes - ICD-9 (715.xx) or ICD-10 (M15. x, M16. x, M17. x, M18. x, M19. x) in 2015 \cite{31}, and 3) continuous enrollment in Medicare Advantage plan with both medical and pharmacy benefits during 2015 and 2016.

2.4. Measures

2.4.1. Target variable: cardiovascular or gastrointestinal disorders (CV-GID)

CVD included in our study consisted of ischemic heart disease, stroke, and congestive heart failure. The GI disorders included dyspepsia, gastroesophageal reflux disorder (GERD), and uncomplicated and complicated peptic ulcer disease (PUD). The target variable was a binary variable indicating the presence or absence of at least one of the above-mentioned conditions. The ICD-9 and ICD-10 diagnosis codes used to identify these conditions from inpatient or outpatient health insurance claims are presented in Table 1 in appendix.

2.4.2. Features/predictors

Features included in our analyses were guided by the modified Determinants of Health Outcomes and Chronic Disease Model \cite{32} and were derived using published literature and study cohort’s most common diagnoses during 2015. Information regarding all features was obtained using data from 2015. Based on the model, the features included age and sex (biological factors); the type of health insurance coverage (health maintenance organization (HMO) vs other, the number of rheumatology and orthopedic doctor visits (access to care); duration and type (selective or non-selective) of NSAIDs, the duration (in days) of prescription opioids or duloxetine, the number of intra-articular corticosteroid injections, the duration (in days) of other medications (oral corticosteroids, or SSRI) known to be the risk factors for GI disorders, and polypharmacy (concomitant use of $\geq 6$ different classes of medications) (pharmacological treatment factors); the number of physical therapy visits (non-pharmacological treatment factor); co-occurring chronic conditions including known CVD (e.g. hypertension, hyperlipidemia, obesity, and diabetes) and GI risk factors (Helicobacter pylori infection), chronic pain conditions (back or neck pain, headache, migraine, other arthritis, and neuropathy), OA-related pain (pain in hip or knee joints), mental health conditions (anxiety disorders, depression, dementia, substance use disorders), and others (urinary tract infection, chronic kidney disease, sleep disorders, Vitamin D deficiency, difficulty in walking) (health status factor); obesity and tobacco use (lifestyle); and region (community resources).

For all medications, except intra-articular corticosteroids, NDC codes corresponding to only oral formulations were used to identify their claims. Chronic medical conditions were identified using both ICD-9 and ICD-10 codes. HCPCS codes were used to identify intra-articular corticosteroid use.

2.5. Data analysis

We evaluated the univariate relationship between the features and the target variable using two-tailed chi-square tests and independent t-tests, depending on the type of feature (e.g. continuous vs categorical). We used both standard statistical (multivariable logistic regression) and machine learning algorithms based on decision trees (Random Forest) to identify the leading predictors of CV-GID. We chose the machine learning approach for their ability to accommodate interactive effects \cite{33}. On the other hand, estimating interactive effects in multivariable logistic regression framework with many predictors is challenging and often inaccurate \cite{34,35}.

The Random forest (RF) is an ensemble learning method based on classification and regression trees (CART) \cite{36}. It is a supervised learning method that creates multiple decision trees simultaneously into a “forest”. Each tree is constructed from a bootstrapped sample of all observations in a dataset. A random subset of independent variables is selected at each node of each tree and the split is made such that the prediction accuracy is maximized. The prediction for an observation is then made based on the highest number of times it is classified (majority voting from all trees) belonging to the CV-GID class (Yes vs No). We used the randomForest package of R software suite version 3.6.3 (R Development Core Team; Vienna, Austria) to build and run RF models with 41 predictors. We used the best practices while building the machine learning models.
For example, we randomly split our dataset into the train and test subsets consisting of 80% and 20% of total observations, respectively. The RF model was built using train data and its performance was evaluated using test data. The combination of the Number of Trees (ntree) at 500 and the number of randomly sampled independent variables at each split at 3 (mtry) gave the best performance in terms of out-of-bag error [36,37]. To derive the leading predictors of CV-GID, we relied on the mean decrease in GINI coefficient.

Unlike statistical learning methods, the machine learning approaches do not provide a measure that summarizes both the direction and magnitude of the associations between the leading predictors and the target variable (example: adjusted odds ratios (AOR) in a multivariable logistic regression). To overcome this, in our study, we used Partial Dependence Plots (PDP) [38], a “model-agnostic” technique that graphically summarizes such relationships. PDPs display how the model predictions partially depend on a specific predictor and whether the relationship between the predictor and the target variable is linear, non-linear, or complex. The dependence is based on the average marginal contribution of each value of the predictor to the predicted probabilities of CV-GID. PDPs were built using the pdp package in R [39]. Additionally, multivariable logistic regression was used to compare its predictive accuracy with the machine learning algorithm (using area under the receiver operating curve (AUROC)), and to examine the association between the leading predictors and the target variable (using adjusted odds ratios (AORs) and 95% confidence intervals (CI)) while adjusting for other variables. There was no indication of multicollinearity between the predictors included in logistic regression based on variance inflation factor (VIF).

### 3. Results

Baseline characteristics of selected variables by CV-GID are presented in Table 1. Co-occurring CVD was identified in 36.9% (n = 16,762), GI disorders in 39.6% (n = 17,967), and CV-GID was identified in 59.0% (n = 26,769) individuals of the study cohort during two years of follow-up. Ischemic heart disease was the most common CVD (28.3%) followed by heart failure (13.3%) and stroke (8.4%). GERD was the most common GI (36.3%) followed by PUD (6.3%) and dyspepsia (2.2%). Overall, a higher percentage of elderly age 80 and over (65.8% vs 51.4% in 65–69 age group), males (61.2% vs 57.8%), elderly with diabetes (63.4% vs 56.2%), hyperlipidemia (63.4% vs 49.4%), hypertension (63.7% vs 42.1%), obesity (64.4% vs 57.8%), *H. pylori* infection (85.1% vs 58.8%), polypharmacy (67.3% vs 47.4%), OA-related pain (61.6% vs 55.0%), and cardiac arrhythmia (76.9% vs 53.1%) had CV-GID compared to elderly without these characteristics (Table 1). The elderly with CV-GID had, on average, shorter duration of non-selective NSAID use (28.2 days vs 30.4 days), and longer durations of selective NSAID (6.2 Days vs 5.1 days), opioid (47.7 days vs 30.5 days), gastroprotective agents (109.5 days vs 26.1 days) and oral corticosteroid use (55.7 days vs 48.5 days) (Table 1). The average number of physical therapy (4.8 vs 3.6), orthopedic doctor (1.7 vs 1.4), and rheumatologist (0.3 vs 0.2) visits, and the number of intra-articular corticosteroid injections (0.9 vs 0.7) were also higher among elderly with CV-GID.

#### 3.1. Predictive performance

Our tuned final random forest model consisted of 41 predictors, three of which were randomly selected for decision tree splits, and 500 trees. We observed 67.5% accuracy, 70.9% sensitivity, 62.6% specificity, and 72.8% positive predictive value (PPV) using test dataset. The performance metrics from multivariable logistic regression were 68.2% accuracy, 78.3% sensitivity, 53.8% specificity, and 70.5% PPV obtained from the standard multivariable logistic regression. While the RF had better specificity, the logistic regression had better sensitivity. The AUROC values for RF and multivariable logistic regression models were 0.73 (95% CI, 0.72–0.74) and 0.74 (95% CI, 0.73–0.75), respectively.

| Variable | CVD or GI AEs | No CVD or GI AEs | p-value |
|----------|---------------|-----------------|---------|
| N        | N             | N               |         |
| Percentage | Percentage |                   |         |
| All      | 26,769        | 18,161          | 41.0    |
| Age      |               |                 | <0.001  |
| 65-69    | 5334          | 5039            | 48.6    |
| 70-74    | 6377          | 5004            | 44.0    |
| 75-79    | 5455          | 3574            | 39.6    |
| 80 and over | 9603      | 4999            | 34.2    |
| Sex      |               |                 | <0.001  |
| Female   | 17,408        | 12,687          | 42.2    |
| Male     | 9361          | 5929            | 38.8    |
| Insurance Type |         |                 | <0.001  |
| HMO      | 9369          | 7128            | 43.2    |
| Other    | 17,400        | 11,488          | 39.8    |
| Diabetes |               |                 | <0.001  |
| Yes      | 11,017        | 6350            | 36.6    |
| No       | 15,752        | 12,266          | 43.8    |
| Hypertension |           |                 | <0.001  |
| Yes      | 19,778        | 11,441          | 36.6    |
| No       | 6991          | 7175            | 50.6    |
| Obesity  |               |                 | <0.001  |
| Yes      | 22,609        | 12,888          | 36.3    |
| No       | 4160          | 5728            | 57.9    |
| H. pylori infection |    |                 | <0.001  |
| Yes      | 5346          | 2952            | 35.6    |
| No       | 21,423        | 15,664          | 42.2    |
| Polypharmacy<sup>a</sup> | |                 | 0.018    |
| Yes      | 275           | 48              | 14.9    |
| No       | 26,494        | 18,568          | 41.2    |
| OA-related pain |        |                 | <0.001  |
| Yes      | 17,759        | 8635            | 32.7    |
| No       | 9010          | 9981            | 52.6    |
| Cardiac arrhythmia (continued) | |                 | <0.001  |
| Yes      | 16,879        | 10,518          | 38.4    |
| No       | 9890          | 8098            | 45.0    |
| Continuous Variables | |                 | <0.001  |
| Mean     | 0.70          | 0.60            | 0.20    |
| SD       | 0.20          | 0.30            | 0.87    |
| Diagnosis of Care Index (FCI) | |                 | <0.001  |
| Duration of non-selective NSAID use (days) | |                 | 0.003   |
| 28.2     | 30.4          | 80.7            |
| Duration of selective NSAID use (days) | |                 | 0.002   |
| 6.2      | 5.1           | 36.1            |
| Number of IA corticosteroid injections | |                 | 0.009   |
| 0.9      | 0.7           | 1.3             |
| Duration of prescription opioids (days) | |                 | <0.001  |
| 47.7     | 30.5          | 79.6            |
| Duration of GDA use (days) | |                 | <0.001  |
| 109.5    | 26.1          | 90.1            |
| Number of physical therapy visits | |                 | <0.001  |
| 4.8      | 3.6           | 10.3            |
| Number of orthopedic doctor visits | |                 | <0.001  |
| 1.7      | 1.4           | 2.8             |
| Number of rheumatologist visits | |                 | 0.001   |
| 0.3      | 0.2           | 1.1             |
| Duration of oral corticosteroid (days) | |                 | 0.002   |
| 55.7     | 48.5          | 97.5            |

Note: The sample includes 45,385 elderly with osteoarthritis from De-identified Optum Clininformatics Data Mart (2014–2015).

<sup>a</sup>Excludes opioid and NSAID combinations.

CVD—cardiovascular disease; GI—gastrointestinal; GDA—gastroprotective agents;
3.1. Leading predictors

The variable importance plot for the top 10 predictors of CV-GID based on mean decrease in GINI coefficient from random forest is presented in Fig. 1. The most important predictor of CV-GID was age followed by cardiac arrhythmia, duration of prescription opioid use, number of orthopedic doctor visits, number of physical therapy visits, number of IA corticosteroid injections, polypharmacy, duration of non-selective NSAID use, duration of corticosteroid use, and hypertension. On the other hand, anxiety, difficulty in walking, and neuropathy were among the least important predictors of CV-GID. The variable importance plot shows that without the top three predictors (i.e. age, and cardiac arrhythmia, and the duration), the performance of the random forest model would suffer significantly. The variable importance plot for all variables is provided in Fig. 1A in appendix.

3.1.2. Association of the leading predictors to CV-GID

The direction of the association of the top ten leading predictors to CV-GID are summarized with univariate partial dependence plots (PDPs) in Fig. 2. The x-axis in these plots represent the values of a particular predictor and the y-axis represent the centered logits of CV-GID. The PDP plots depict the direction of the associations, as well as, the type of the relationship (linear or non-linear) of the predictors with the target variable. For example, age had a positive and linear relationship with the predicted probability of CV-GID. Similarly, the predicted probability of CV-GID increased with cardiac arrhythmia, longer durations of prescription opioid use, higher number of physical therapy visits or intra-articular corticosteroid injections, polypharmacy, longer durations of corticosteroid use, or hypertension. On the other hand, the number of orthopedic doctor visits had a complex relationship with the risk of CV-GID wherein ten or less annual visits were associated with a higher predicted probability of CV-GID and 20 or more visits were associated with a lower predicted probability of CV-GID compared to those without any visits. The duration of non-selective NSAID had a linear and negative relationship with the risk of CV-GID.

For the ease of comparison with other published studies and to examine the association of the leading predictors with CV-GID, we also present the adjusted odds ratios (AORs) and 95% confidence intervals from a multivariable logistic regression in Fig. 3. Although the regression model was adjusted for all 41 variables, the results for only the top 10 predictors are presented. The direction of the associations from logistic regression were similar to those obtained using the PDP plots. However, the associations between the duration of opioid use, the number of orthopedic doctor visits, and the duration of oral corticosteroids were not statistically significant.

4. Discussion

In the present study, we examined the leading predictors of CV-GID and their associations with the predicted probability of CV-GID in a heterogeneous real-world population of elderly with OA. This is an important group of patients as they often have multiple clinical and biological risk factors that place them at a higher risk for CVD or GI disorders. While traditional risk factors for CVD and GI disorders are known, the possible roles of disease-disease, drug-disease, and drug-drug interactions of OA and its treatment with combined CV-GID are not studied before. In our study cohort, 59% elderly with OA had at least one CV-GID. Using a random forest classifier, we predicted CV-GID with high sensitivity and specificity, which were comparable to the traditional multivariable logistic regression. We also evaluated the relationships between the leading predictors and predicted probability of CV-GID by using an interpretable machine learning technique called PDP.

We observed that age was the strongest predictor of CV-GID risk in our study cohort. The results from the PDP plots and multivariable logistic regression showed that increasing age was associated with an increase in the risk of CV-GID. This finding was anticipated as age is an independent risk factor for both CVD and GI disorders [23, 24]. As the risk of CVD and GI disorders is higher in individuals with OA [2], primary prevention efforts may need to be targeted towards those OA patients with pre-existing shared risk factors for these conditions. Such risk factors may include long-term use of NSAIDs, physical inactivity due to OA-related joint pain, or obesity.

The second leading predictor of CV-GID was cardiac arrhythmia. It is well-known that cardiac arrhythmia can increase the risk of stroke and other CVD [40]. Moreover, just like atherosclerosis, some medications (e.g. antiplatelets, anticoagulants) used to reduce the risk of cardiovascular consequences from cardiac arrhythmias have been shown to increase the risk of GI disorders like GI bleeding [41, 42]. Therefore, the concomitant use of NSAIDs and these medications should be re-evaluated in elderly with OA.

One interesting finding from our study was that longer or greater use of certain medications (prescriptions opioids, IA and oral corticosteroids) used to manage pain in OA was associated with a higher risk of CV-GID. While there is some evidence linking the use of oral corticosteroids and opioids to both an increased risk of CVD and GI disorders [11, 42], such evidence for IA corticosteroids is very limited. An alternative explanation could be that patients requiring more pain medications including IA corticosteroid injections may have more severe OA that could increase the risk of CV-GID [43]. Our findings suggest a need for a closer examination of CV-GID risk of patients requiring frequent treatment with these agents. On the other hand, it was not clear why the higher number of physical therapy visits was associated with a higher risk of CV-GID in our study cohort. Our finding is in contrast to the findings of a study by Yeh et al. [44] which observed that physical therapy reduced the risk of ischemic heart disease and dyslipidemia. However, that study was conducted in adults with newly diagnosed OA and our study included all prevalent OA cases. Moreover, our study was only limited to elderly who may be at a higher risk for CV-GID. More research investigating the link between the frequency of physical therapy sessions and the risk of CV-GID is needed.

Our results also indicated that the number of orthopedic doctor visits had a complex relationship with the risk of CV-GID. While annual visits below 10 were associated with a higher risk for CV-GID, visits above 20
were associated with a lower risk for CV-GID. As there is no research examining the possible link between the number of orthopedic doctor visits and the risk of CV-GID, we speculate that patients requiring more visits to orthopedic doctors could be the ones more likely to receive joint replacements. Previous research has shown that joint replacement can help reduce the risk of CVD by up to 50% possibly due to an increase in physical activity [45]. We also found that polypharmacy was associated with a higher risk of CV-GID. This finding is supported by previous studies that found polypharmacy to be associated with both CVD and GI disorders [46,47]. Unlike other previous studies [9,10] that found a higher risk associated with NSAID use, we observed that the longer duration of non-selective NSAIDs was associated with a lower risk of CV-GID. However, our results could be confounded by over-the-counter NSAID use which is common among elderly. Moreover, it is likely that only those patients considered to be at a lower risk for CV-GID were prescribed NSAIDs because of the well-known CV and GI adverse events of these drugs. Lastly, hypertension, which is known to be a significant risk factor for CVD itself [48] or GI disorders through its treatment [49], was also among the top ten predictors of CV-GID in our study. As hypertension is common among elderly, 4 in 5 elderly with OA in our study had hypertension, NSAIDs and other drugs with a potential to lead to its exacerbation should be used with caution in this patient population.

Our study has many strengths. It is the first study to identify the leading predictors of CV-GID in elderly with OA using machine learning with a comprehensive list of predictors available from medical claims. The use of real-world data in the form of insurance claims allowed us to identify important variables relevant for clinical decision-making. We also used interpretable machine learning method (e.g. PDP) to identify the relationship of the top predictors with model predictions. However, the results from our study should be viewed in the light of its following limitations. Being an observational retrospective cohort study, we cannot infer causality. We did not have information on some variables including severity of OA, pain, lifestyle factors, and over-the-counter medications which can affect various treatment and outcome related factors.

Fig. 2. Univariate Partial Dependence Plots for Top 10 Predictors from Random Forest. Note: The sample includes 45,385 older adults with osteoarthritis from De-identified Optum Clinformatics Data Mart (2014–2015). The x-axis represents values of the predictor. The y-axis represents logits for CV-GID. nsNSAI – non-selective nonsteroidal anti-inflammatory drugs.

Fig. 3. Results from Multivariable Logistic Regression. Note: The sample includes 45,385 older adults with osteoarthritis from De-identified Optum Clinformatics Data Mart (2014–2015). The model was adjusted for all 39 predictors of CV-GID. CI—confidence interval; CV-GID—cardiovascular or gastrointestinal disorders; IA—intra-articular; NSAID—nonsteroidal anti-inflammatory drugs.
Administrative claims data could suffer from misdiagnosis and inconsistent coding of some conditions (e.g. pain) that do not affect reimbursement. As we used data from commercial insurance claims from one large insurance provider, the results from our study cannot be generalized to all elderly in the US. Additionally, results from the PDPs may be inaccurate in the presence of variable interactions which were not evaluated as part of this study. The follow-up in our study was only limited to two years making it difficult to examine the association between the variables of interest over a longer duration.

5. Conclusion

CVD and GI disorders are common among elderly with OA. Our findings suggest that these conditions can be predicted using certain biological, health-, and treatment-related factors that are readily available in a clinical setting, such as age, cardiac arrhythmia, opioids, corticosteroids, polypharmacy, NSAIDs, and hypertension. Our study demonstrates that machine learning methods can be used to identify the leading predictors of a binary variable as well as to study their relationships. Future research that includes OA severity and biomarkers are needed to improve the prediction accuracy, as well as, to confirm some of the predictors of CV-GID identified in our study.

Author contributions

All authors contributed to the conception and design of the research. JP and US conducted the statistical analyses. JP wrote the first draft. All authors worked on successive iterations.

Funding source

Research reported in this publication was not supported by any grant or funding.

Declaration of competing interest

Dr. Patel has nothing to disclose; Dr. Ladani has nothing to disclose; Dr. N Sambamoorthi has nothing to disclose; Dr. LeMasters has nothing to disclose; Dr. Ladani has nothing to disclose; Dr. Dwibedi has nothing to disclose; Dr. U Sambamoorthi has nothing to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocarto.2021.100148.

Appendix

Table 1

Diagnosis codes used to identify cardiovascular and gastrointestinal disorders (CV-GID).

| Condition                | ICD-9 Codes | ICD-10 Codes |
|--------------------------|-------------|--------------|
| Angina                   | 413.x       | I20.x        |
| Myocardial infarction    | 410.xx      | I21.xx-I23.x |
| Other ischemic heart disease | 411.xx      | I24.x, I25.xx |
| Stroke                   | 430.x, 431.x, 433.x-436.x, 362.3 | I60.x, I61.x, I63.x, I64.x, H34.1, G45.x |
| Congestive heart failure | 398.91, 402.x1, 402.x3, 404.x1, 404.x3, 422.90, 425.4, 425.9, 428.xx | I50.xx |
| Dyspepsia                | 536.8       | K30          |
| GERD                     | 530.11, 530.81 | K21.0, K21.9 |
| Uncomplicated PUD        | 531.3, 531.7, 531.9, 532.3, 532.7, 532.9, 533.3, 533.7, 533.9, 534.3, 534.7, 534.9 | K25.3, K25.7, K25.9, K26.3, K26.7, K26.9, K27.3, K27.7, K27.9, K28.3, K28.7, K28.9 |
| Complicated PUD          | 531, 531.1, 531.2, 531.4, 531.5, 531.6, 532, 532.1, 532.2, 532.4, 532.5, 532.6, 533, 533.1, 533.2, 533.4, 533.5, 533.6, 534, 534.1, 534.2, 534.4, 534.5, 534.6, 535.01, 535.11, 535.21, 535.31, 535.41, 535.51, 535.61, 578, 578.1, 578.9 | K25.0-K25.6, K26.1-K26.2, K26.4-K26.6, K27.1-27.2, K27.4-K27.6, K28.1-28.2, K28.4-K28.6, K29.01, K29.21, K29.41, K29.51, K29.61, K29.71, K29.81, K29.91, K92.0-92.2 |
Fig. 1A. Variable Importance Plot from Random Forest – All Predictors. Note: The sample includes 45,385 older adults with osteoarthritis from De-identified Optum Clinformatics Data Mart (2014–2015). Variable importance is based on mean decrease in GINI coefficient. COPD – chronic obstructive pulmonary disorder; IA – intra-articular; NSAID – nonsteroidal anti-inflammatory drugs; OA – osteoarthritis.

References

[1] G.S. Fernandes, A.M. Valdes, Cardiovascular disease and osteoarthritis: common pathways and patient outcomes, Eur. J. Clin. Invest. 45 (4) (2015) 405–414, https://doi.org/10.1111/eci.12413.

[2] S. Swain, A. Sarmanova, C. Coupland, M. Doherty, W. Zhang, Comorbidities in osteoarthritis: a systematic review and meta-analysis of observational studies, Arthritis Care Res. 72 (7) (2020) 991–1000, https://doi.org/10.1002/acr.24008.

[3] G.A. Hawker, R. Croxford, A.S. Bierman, P.J. Harvey, B. Ravi, I. Stanaitis, et al., All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a population based cohort study, PloS One 9 (3) (2014), https://doi.org/10.1371/journal.pone.0091286.

[4] I.K. Haugen, V.S. Ramachandran, D. Misra, T. Neogi, J. Niu, T. Yang, et al., Hand osteoarthritis in relation to mortality and incidence of cardiovascular disease: data from the Framingham heart study, Ann. Rheum. Dis. 74 (1) (2015) 74–81, https://doi.org/10.1136/annrheumdis-2013-203789.

[5] H. Wang, J. Bai, B. He, X. Hu, D. Liu, Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep39672.

[6] C. Camille, J. Martel-Pelletier, J. Sellam, J.-P. Pelletier, J.-P. Raynauld, F. Roubille, Osteoarthritis and cardiovascular diseases: is meta-inflammation the missing link? J. Cardiol. Ther. 2 (3) (2015) 315–320.

[7] F. Berenbaum, Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!), Osteoarthritis Cartilage 21 (1) (2013) 16–21, https://doi.org/10.1016/j.joca.2012.11.012.

[8] K. Pekhuri, H. Vapaatalo, R. Korpela, Even low-grade inflammation impacts on small intestinal function, World J. Gastroenterol. 16 (9) (2010) 1057–1062, https://doi.org/10.3748/wjg.v16.i10.1057.

[9] Coxib and traditional NSAID Trialists’ (CNT) Collaboration, N. Bhala, J. Emberson, A. Merhi, S. Abrahamson, N. Arber, et al., Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials, Lancet 382 (9984) (2013) 769–779, https://doi.org/10.1016/S0140-6736(13)60900-9.

[10] N. Moore, Coronary risks associated with diclofenac and other NSAIDs: an update, Drug Saf. 43 (4) (2020) 301–318, https://doi.org/10.1007/s40264-019-00900-8.

[11] Y. Khodneva, P. Muntner, S. Ketke, M.M. Safford, Prescription opioid use and risk of cardiovascular disease among older adults from a community sample, Drug Alcohol Depend. 140 (2014) e103, https://doi.org/10.1016/j.drugalcdep.2014.02.297. Complete.

[12] A. Chen, M.A. Ashburn, Cardiac effects of opioid therapy, Pain Med. 16 (Suppl 1) (2015) S27–S31, https://doi.org/10.1111/pme.12915.

[13] R.S. Mhaskar, I. Ricard, A. Adiyatir, R. Laxminarayan, B. Amol, W. Santosh, et al., Assesment of risk factors of Helicobacter pylori infection and peptic ulcer disease, J. Global Infect. Dis. 5 (2) (2013) 60–67, https://doi.org/10.4103/0974-777X.112288.

[14] L.M. Thor, M. Fordblom, J. Wadén, M. Sarabemble, N. Tolonen, K. Hietala, et al., Metabolic syndrome as a risk factor for cardiovascular disease, mortality, and progression of diabetic nephropathy in type 1 diabetes, Diabetes Care 32 (5) (2009) 950–952, https://doi.org/10.2337/dc08-2022.

[15] R. Nishenametla, T. Manesh, S. Hafeez, H. Krishna, Prediction of heart disease using machine learning algorithms, Int. J. Eng. Technol. 7 (2018) 363–366, https://doi.org/10.14419/ijet.v7i2.15714.

[16] A.R. Gaite-Escamila, A. Hajar El Hassani, E. Andrevs, Classification models for heart disease prediction using feature selection and PCA, Info. Med. Unlock. 19 (2020) 100300, https://doi.org/10.1016/j.imu.2020.100300.

[17] A.R. Brook, N.L. Kleinman, R.S. Choung, J.E. Smeeding, Excess comorbidity prevalence and cost associated with functional dyspepsia in an...
