AN EXPLORATION OF MACHINE LEARNING TO PREDICT MEDICATION WASTE AMONGST RHEUMATOID ARTHRITIS PATIENTS

Katrina Alexandra Tan Go
Student ID: 5847761

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
Science and Business Management

Utrecht University
Graduate School of Life Sciences
Department of Pharmaceutical Sciences

Supervisors: Dr. C.L. Bekker and prof. Dr. B.J.F. Bemt
Examiner: prof. Dr. A.C.G. Egberts

ABSTRACT

Background: In case of treatment failure, rheumatoid arthritis (RA) patients switch their relatively expensive biological disease- modifying anti-rheumatic drug (bDMARD) therapy, which could lead to waste. Machine learning has potential to be used in pharmacy to predict medication waste.

Aim: To explore the application of machine learning to identify patterns in patient, clinical and medication factors that lead to medication waste.

Methods: In a retrospective cohort study, patient, clinical and medication data was collected from a Dutch outpatient pharmacy and hospital information system of patients (≥ 18 years) who received at least one bDMARD prescription, dispensed between January 2015 and December 2020. Medicine waste was defined as a treatment switch before its expected end date. A random forest model was used to identify predictors.

Results: The database included 1996 patients, of which 285 wasted at least 1 syringe of bDMARD. During the five-year study period, a total of 719 units were unused, with an economic value of €237,692. Out of 32,484 prescriptions, 324 lead to waste. The random forest model had a positive prediction value of 0.21, with total cost of a prescription, age, disease duration, as the highest predictors for medication waste.

Conclusion: bDMARD waste occurs when rheumatoid arthritis patients switch therapies. Machine learning has the potential to be used in waste preventing activities. With improvements to the model, such as down sampling, reducing features and correcting for correlations, it can be used to identify patterns which lead to medication waste.

INTRODUCTION

Rheumatoid arthritis (RA) is an auto-immune disease, characterized by inflammation of the joints or connective tissue which causes severe pain for those who suffer from it. When left untreated, it can cause damage to these structures. Biological disease modifying anti-rheumatic drugs (bDMARDs) have shown to reduce and even halt the progression of the disease, thereby substantially improving quality of life (1–4).

Although bDMARDs have positive effects for patients with RA, these bDMARDs are expensive and frequently remain unused (5). Main reasons thereof include medication passing the expiry date, therapy change, condition resolving and adverse side effects (6,7). Bekker et al. showed that over 50% of patients who discontinue bDMARDs therapies, end up with unused medication, worth about €180,000 annually in the Netherlands. (8). Not only is this a financial burden, it also leads to environmental pollution when improperly disposed of by patients.

To optimize pharmacological treatment for patients with RA and reduce this medication waste, patients at risk for having unused medications could be identified. Machine learning (ML) is a statistical technique which uses big data to train models and is considered to be the most precise prediction model (9–11). Within healthcare, the most common use of machine learning is in precision medicine; predicting the outcome of treatment protocols based on patient characteristics and context. The majority of these applications require a known outcome variable (e.g. onset of a disease) (12,13). Hospital and pharmacy information systems are potentially valuable sources of information regarding unused medication and therefore, machine learning could be applied to identify patterns within patients who waste.

This study aims therefore to explore the application of machine learning in identifying patterns in patient, clinical and medication factors for medication waste.

METHODS

Design and setting
This retrospective cohort study was conducted in the outpatient pharmacy of the Sint Maartenskliniek, Nijmegen, in the Netherlands. This clinic is specialized in diseases which affects movement, including rheumatology. In the Netherlands, bDMARDs are dispensed by hospital-based outpatient pharmacies in the Netherlands.

Ethical considerations
Patient data was handled confidentially and according to the Dutch law 'Protection of Personal Data' for medical research. The study protocol was reviewed and approved by the ethical review committee of Radboud University Medical Center (CMO-reference number 2021-7341).

Study population
Patients diagnosed with rheumatoid arthritis, aged ≥ 18 years who had received a bDMARD from the outpatient pharmacy for at least one prescription between January 2015 until

December 2020, either as a first or repeated supply, were considered eligible for study inclusion. Patients included have complete pharmacy data containing waste information.

Data collection

Information from included patients was extracted from the hospital information system, which included patient characteristics (gender, age, disease duration) and clinical measures: date of measurement, disease activity score (DAS), number of swollen and tender joints, patient global assessment (visual analogue scale (VAS)), erythrocyte sedimentation rate, C-Reactive Protein concentration (CRP), hemoglobin (Hb) and creatinine levels, anti- Cyclic Citrullinated Peptide (CCP) and Rheumatoid Factor (RF). The following information about patient's dispensed medications was extracted from the pharmacy information; dispensing date, medication name, strength, anatomical therapeutic chemical (ATC) classification, cost price (per prescription), administration form, dispensed amount and prescribed daily dosage.

Measurements

Firstly, medication waste needs to be identified before being applied to the machine learning model. Medication waste was assumed to occur when a patient was left with unused bDMARD syringes when switching to a different strength of the same medication or to another bDMARD type before the theoretical end date of the prescription. For patients who had switched their medication, the theoretical end date was calculated by adding the dispensed quantity (i.e. number of dispensed units multiplied by the prescribed daily dosage) to the dispensing date. Unused medication was calculated from the difference in days between the theoretical end date and the dispensing date of the subsequent newly prescribed medication, divided by the prescribed daily dosage.

$$Theoretical\ end\ date\ =\ start\ date\ +\ (units\ *\ daily\ dosage)$$

$$Medicine\ waste\ (units) = \frac{(theoretical\ end\ date - date\ of\ switch\ in\ medications)}{daily\ dosage}$$

The economic value was also calculated by multiplying the cost price of a singular unit of medication with the number of unused units (i.e. number of syringes remaining unused).

Data analysis

Data were descriptively analyzed in STATA 13. Averages were expressed as means with standard deviations (SD) or as medians with interquartile ranges (IQR) if non-normally distributed. Proportions were expressed as percentages.

Machine learning: random forest

A random forest (RF) model was built based on the dataset of medication waste positive (prescription leading to waste) and negative events (prescriptions that did not result into waste). For every event, treatment and clinical characteristics were assessed and potential changes thereof that could potentially be determined which lead to waste.

Since random forest algorithm is not commonly used in pharmacy, this section is dedicated to explain the fundamental basics. The random forest algorithm is considered to be the best

method used for classification as it takes the most important variables for classification and creates an unbiased model from the original data by introducing randomness (9).

Random forest is an example of machine learning which uses many randomized decision trees based on classification and regression trees method (14). The algorithm creates various random individual decision trees which each produce a class prediction (Figure 1A). The class prediction with the most votes by majority voting, becomes the models' prediction (Figure 1B).

Each decision tree has a standard flow-chart structure in which the internal splitting nodes test variables and branch out. The tree building uses a top-down, greedy approach known as binary recursive splitting that chooses the variable with the lowest impurity score at each node. It is top-down as the tree begins at the top and successfully splits into two new branches further along. The term greedy is used showing that at every step the best split is chosen at that particular stage, not looking ahead (15).

The Gini-index is a well-known measure of node impurity and depicts the most important variables for classification (9,14). The Gini-index measures the probability that a randomly chosen variable gets wrongly classified when following the class distribution on a scale of 0 to 1. Zero indicates that all elements belong to one class, or that only one class exists, which means that it is pure, and 1 meaning that all variables are randomly distributed. This means that 0.5 indicates an equal distribution between two classes. As the tree traverses down, the node impurity score gets closer to 0 because the groups become smaller and a better classification is made. The tree growing process is repeated indefinitely until no further improvements can be made to the node impurity (16).

Random forest incorporates two types of randomness; 1. each tree is built based on a random bootstrapped sample of the original data (Figure 1A), 2. at each splitting node, a subset of variables is randomly chosen amongst all available variables to decide the node splitting. The best node split in variables is based on the lowest node impurity score.

In this study, the learning process was carried out with the randomForest package for R software environment. The final predictive model of the random forest includes a sensitivity, specificity and positive prediction value (PVV) of the model. To demonstrate the potential use of machine learning, the relation between the DAS and medication waste were taken as example to demonstrate the feasibility in a clinical setting as the DAS is often used as an empirical indication of disease progression (17,18).
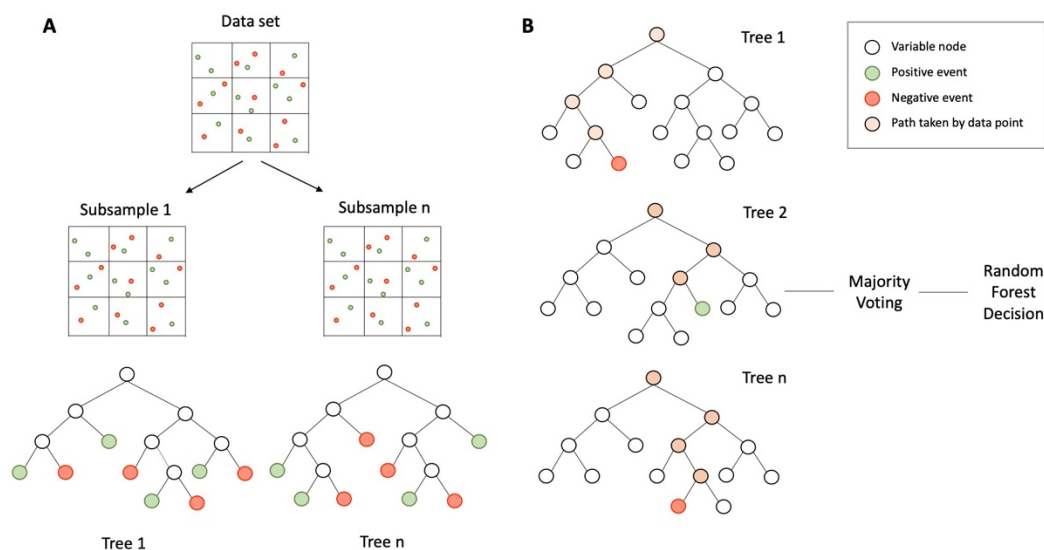
Figure 1. **Random forest model.** Example of training and classification process of a random forest model. A) Each tree is built based on a random bootstrapped sample of the original data. B) Every data point starts at the top of the decision tree and makes it way down. At every split, the data point tests the variable values. This process is repeated until it reaches the end node where a class is assigned. At the end, each tree casts a vote for the preferred class and the tree with the most votes is chosen for the final prediction.

Variables

Certain variables including disease activity score, number of swollen and tender joints and patient global assessment can change over time. The model takes these changes over time into account by creating new variables for each time point. Additional features were added to the data set for the relation to time. These features include, difference between the DAS score over time, amount of days since last prescription, amount of times waste has occurred before, amount of types of medication a patient has been prescribed, amount of prescriptions in total and of a particular medication.

RESULTS

A total of 2427 patients were eligible for inclusion. Of these, 347 were excluded due to missing prescription data which likely resulted from having their medication dispensed by another pharmacy. A further 84 patients were excluded due to not fulfilling the inclusion criteria. This can be found in Figure 2. In total 1996 patients were included, with a mean age of 58.8 SD ± 13.8 years and 69.7% was female (Table 1).

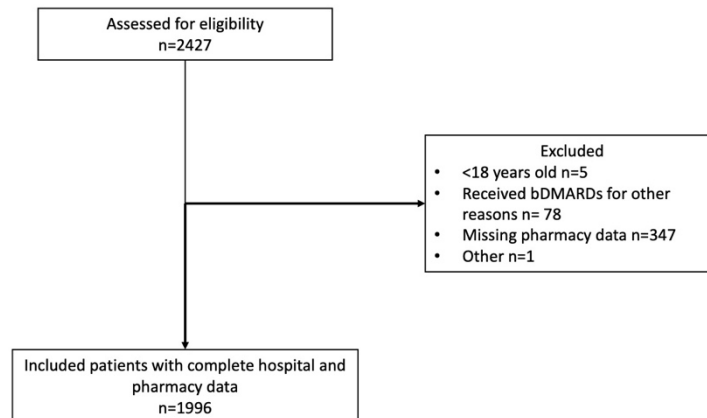**Rheumatoid Arthritis patients using bDMARDs**

Figure 2. Flow diagram of included patients with RA in the study.

In total, 32,484 prescriptions were dispensed by the pharmacy, with an average of 16.2 prescriptions per patient. There were 324 positive events which lead to medication waste (1% of 32,484 prescriptions). The most prescribed bDMARDs included etanercepts (47.6%) and adalimumab (32.8%) (Table 1.)

Table 1. Patient characteristics of patients who have received a bDMARD

	Total patients n=1996
Age (years), mean (SD)	58.8 (13.8)
Gender, (female) n (%)	1392 (69.7)
Disease duration (years), median (range)	7.2 (1.8 – 15.1)
Prescriptions per patient, n, mean (SD)	16.3 (10.9)
Total prescriptions, n, (%)	32,484
Etanercept	15,468 (47.6)
Adalimumab	10,663 (32.8)
Abatacept	1,716 (5.28)
Tocilizumab	2,794 (8.60)
Certolizumab	613 (1.89)
Golimumab	911 (2.80)
Sarilumab	193 (0.59)
Secukinumab	52 (0.16)
Anakinra	68 (0.21)
Ustekinumab	6 (0.02)

Quantity and economic value of unused bDMARDs

A total of 719 syringes were left unused, on average 2 per patient (IQR 1-3). These unused bDMARDs had a total economic value of €237,692, with a median value of €285 (IQR €256 - €476) per unit (Table 2). The majority of unused units included etanercept (56.6%) and adalimumab (19.7%) as its active ingredient.

Table 2. The quantity and economic value of unused bDMARDs during the study period 2015-2020

bDMARD	Quantity (no. of syringes), n (%) n=719	Total Economic Value of unused bDMARD €)	Median units per patient n (IQR)	Median value of unused bDMARD (€) (IQR)	Median value of a prescription leading to waste (€) (IQR)
Etanercept	407 (56.6)	99,501	1 (1-2)	293 (283-585)	1805 (1144-2313)
Adalimumab	142 (19.7)	75,578	2 (1-3)	856 (545-1428)	2855 (2106-3441)
Abatacept	71 (9.87)	19,477	3 (2-5)	825 (557-1392)	2740 (2169-3250)
Tocilizumab	55 (7.65)	16,765	1 (1-3)	315 (303-944)	2513 (1214-3576)
Certolizumab	32 (4.45)	16,232	1 (1-4)	534 (502-2003)	3004 (2009-3067)
Golimumab	6 (0.83)	6769	1.5 (1-2)	1693 (1130-2255)	3946 (2824-4510)
Sarilumab	4 (0.56)	2082	2 (1-3)	1040 (554-1527)	1873 (1527-2219)
Secukinumab	2 (0.28)	1287	2 (2-2)	1287 (1287-1287)	1930 (1930-1930)

Machine Learning: Random forest

The final random forest had a sensitivity of 0.54 and a specificity of 0.62. The Positive Prediction Value (PVV) of the model was 0.21. A variable importance analysis using the Gini index to assess the variable nodes' impurity was done. Figure 2 shows the results of this. The variables are ranked based on their importance. Only 11 variables are shown This indicates that the top five of most important variables that indicated risk for unused bDMARD included the total cost of a prescription, patient's age, time since diagnosis, disease activity score, and the number of previously received prescriptions.
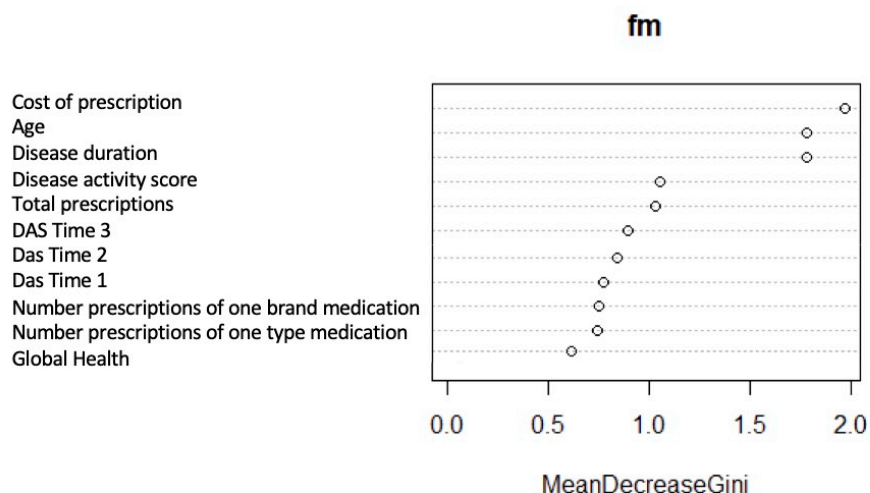

Figure 3. Variable importance analysis performed by Random Forest. Only the top 11 variables are shown, ranked by importance by the model.

Machine Learning: relation between DAS and medication waste

As an example, to highlight the use of machine learning, the relation between a change in DAS and occurrence of medication switch is used. RA activity is measured by the DAS score, indicating remission (score <2.3), low disease activity (score 2.3-3.8), medium disease activity (score 3.8-4.9), or high disease activity (score >4.9). Over time, the disease activity scores (DAS) can change. Based on rheumatologists, it can be expected that the medication switch depends highly on the DAS. With machine learning, it was possible to visualize if this was actually the case. In figure 4A. the DAS score initially increases, but no change in medication occurs. Once the score increases above the 3.2 line, the medication switches from Enbrel to Benepali (both brands with active ingredient etanercept). After this switch, the DAS is seen to decrease over the months. Similarly, in figure 4B. the DAS increases to the point of it being categorized as high; this is when a switch in medication occurs. While on this medication, the DAS seems to go into remission for a few years, until it rises again, resulting into another switch. For a few patients, a trend can be seen between a change in DAS and occurrence of unused bDMARD.

However, this trend cannot be seen for the majority of the patients. Figure 4C. depicts an overall low/remission disease activity score, which initially spiked, but still switched medications later on. Figure 4D. shows a spike in the DAS score, but returned to remission levels. This, however, still resulted into a switch in medications. Consequently, the DAS score continued to increase over time, but no switch occurred afterwards.
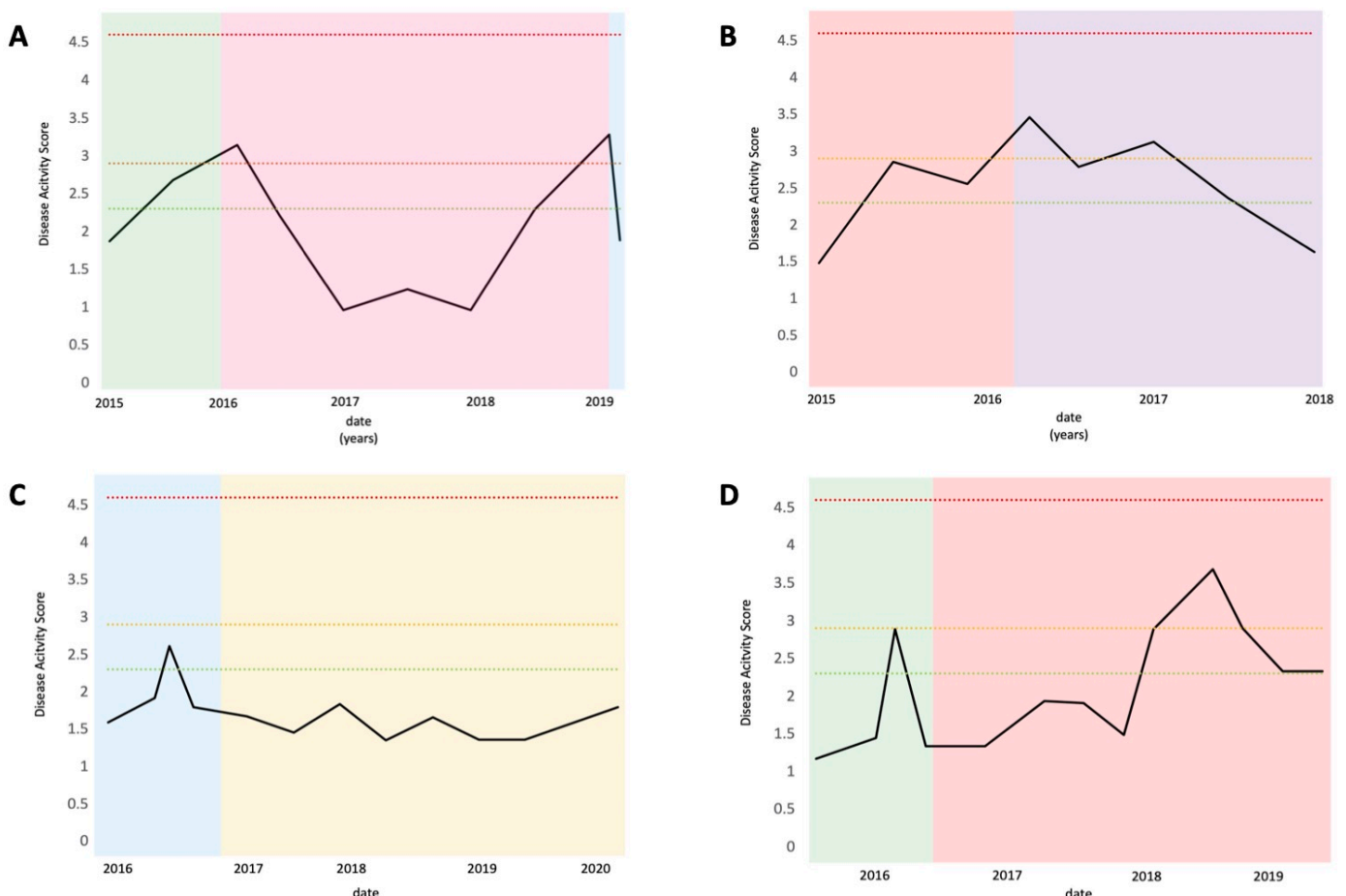

Figure 4. A visual example of the trend in Disease Activity Score of 4 different patients over time and their subsequent use of bDMARDs. The color change in the area indicates a switch in bDMARD used. The dotted lines represent the DAS: remission (<2.3), low (2.3-3.8), medium (3.8-4.6), or high (>4.6).

DISCUSSION

In this study, bDMARD waste resulting from a switch in therapy was calculated based on the data collected from the pharmacy information system. With this medication waste information, utilization of machine learning is evaluated to identify risk factors for waste. The random forest model used, shows that it is capable of recognizing patterns in prescribing data, looking at clinical, patient and medication characteristics. However, the strength of this model is quite weak, shown by the low sensitivity, specificity and positive predictive value.

Overall, it was estimated that around 64 million euros was spent on bDMARDs that were dispensed to 1996 rheumatoid arthritis patients during the inclusion period, of which 0.37% was wasted as patients had unused medication due to a switch in therapies. Annually, this equates to €41,286. This study indicates that only a very small quantity of dispensed bDMARDs are left unused. These finding correspond with those of a previous study done estimating the quantity of bDMARDs, which found that less than 0.8% spent on these medications were wasted (8,19). Both the previous study and this study similarly show that only a small quantity of dispensed bDMARDs remain unused. However, the current model is only based on medication waste resulting from a switch, which was calculated assuming that the dispensing date of the bDMARD is also the start date of the therapy. Additionally, there are many other sources of medication waste amongst patients who do not switch, such as discontinuation of treatment or non-adherence, which were not taken into account and therefore, the results are likely an underestimation of absolute quantity of unused bDMARDs (8). Also, this study is based on a single site, which is highly specialized in rheumatic diseases. This could indicate that waste because of a switch occurs less, due to specialized treatment protocols (20). These outcomes demonstrate that a significant amount of money is wasted on unused bDMARDs and that prevention is needed.

Noticeably, the prescriptions which lead to wasted bDMARD units had a higher cost when compared to the prescriptions which did not lead to waste. These findings correspond with those Bekker et al. found, namely that there is a higher risk of preventable waste when prescriptions are dispensed for a longer time (8). Dosage interval also plays a role in this, as medications with a longer dosage interval usually means a higher concentration of active ingredient, making this medication more expensive than those with a shorter dosage interval. As this was a retrospective study with an inclusion period of 5 years, the duration of treatment was not considered. However, there are various interventions that can be undertaken to reduce this medication waste. By limiting the prescriptions of first-use medications to 15 days, and follow up prescriptions of maximum of 1 month, the risk of waste is reduced and the waste that does occurs when switching will be minimal (3,21,22). However, this is not feasible for cheaper medication as the financial cost will be too high for pharmacies as it will increase the dispensing fee. Dispensing smaller amounts is mainly beneficial for expensive medication such as bDMARDs since this fee is only a fraction of the total price. Additionally, dispensing of smaller amounts of medication requires some cooperation from the patient as it requires more visits to the pharmacy. To lessen the burden on patients and on the pharmacy, being able to identify high risk patients who will waste medications, is beneficial for both parties, as it is more cost effective to dispense larger amounts.

Machine learning: random forest

Random forest is considered to be the most precise predictive machine learning model by incorporating randomness. The advantages that it brings is its ability to determine variable importance, ability to model interaction between variables and flexibility in performing statistical models including classification and unsupervised learning (9). When applied to this study, the random forest model had a positive predictive value of 21% and a sensitivity of 54% and a specificity of 62%. According to these percentages, the model is relatively low in predicting and would provide a high degree of false positives. This can be accounted by a variety of factors. This can be seen in the example used in this study which is the relation between the disease activity score and switching bDMARDs. The DAS is an empirical way of measuring inflammatory activity in rheumatoid arthritis and is used as a tool to monitor inflammation, support choice of specific DMARDs, adjust DMARD dosage, to understand if the chosen therapy is needed and effective and is the central determinant of a prescription (17). It is therefore expected to see a trend in the DAS and the degree of switching medications. For a few select patients, a trend can be seen. However, for the majority of the patients, a switch in medications occur without a supposed change in the DAS. Additionally, there were brand switches within one type of medication, for which a reason is hard to find in clinical data. It is suggested that the DAS alone cannot be used to determine prescriptions, and that there are other variables that it does not cover (23). This is also the case with other assessments of rheumatoid arthritis. Therefore, it could be said that there are so many other variables that go into determining treatment for rheumatoid arthritis and that recognizing a pattern is difficult, which would further explain the low predictive value of the model.

The final variables ranking in a descending order of importance, provided by the random forest variable importance analysis suggests that the main variables involved in bDMARD waste are related to medication factors, DAS28 score, and duration of disease as will be discussed later. Performing a feature selection has shown to reduce error by eliminating variables which do not contribute or has a minimal effect on the predictive performance of the model, thus improving the overall predictive performance and/or specificity and sensitivity (9,12). By comparing the different models with reduced features, the quality of the model can be evaluated and perhaps a more efficient model can be made (12). Studies demonstrated that feature selection did not change the predictive performance, but increased the specificity and sensitivity, making it more efficient and reducing the model's complexity. Additionally, new features could be added incorporating different scales to measure progression of disease, such as the Health Assessment Questionnaire (HAQ) and Rheumatoid Arthritis Disease Activity Index (RADAI) (23). The use of additional medications and comorbidities were also not incorporated as variables in the predictive model (24).

Regarding the results of variable importance, only the most relevant variables will be discussed. According to the model, the most predictive variable is the total cost of the prescription. As discussed earlier, the median price of a prescription leading to waste is higher than one that does not lead to waste. But interestingly, the age and disease duration of a patient are also indicators for medication waste. These variables are related to each other. Onset of rheumatoid arthritis is around 40 to 45 years old and Table 1 shows that the average age of the study population is approximately 60 years old with a median disease duration of 10 years. Studies have shown that 30% of RA patients enter remission within 3 years of onset suggesting that less medication switches occurred due to the effectiveness of their bDMARD

(25,26), further emphasizing that the medication waste noted in this study is an underestimation of the absolute quantity of unused bDMARDs amongst all rheumatoid arthritis patients. Additionally, studies have shown that even with a feature selection done, the presence of correlated variables can reduce the importance of causal variables in the random forest model (27,28). Therefore, it is suggested to also test and subsequently correct for correlations to get a more accurate predictive model.

Based on the classification results by the random forest, the model's performance is relatively low and the sensitivity and specificity is quite unbalanced. Most algorithms assume a balanced data set with equal amounts of positive and negative events (29). A problem occurs when an unbalanced data set is used. In this study, the model had to base the final predictive model on 324 positive events (1% of 32,484 prescriptions). When using an unbalanced set, the algorithms fails to distinguish distributive characteristics of the data and the affects the sensitivity of the prediction (14). Similarly, studies done on predicting varied outcomes also have unbalanced data sets (30). However, these studies have created a more balanced data set by down-sampling, which showed to have improved the sensitivity (12). Therefore, by implementing a pre-processing strategy by down sampling, the predictive performance of the model can be improved. In this study, the relatively low number of positive events possibly indicates that the problem of medication waste is not that severe and occurs rarely without a specific pattern (8).

Limitations

This study has several important limitations. Firstly, our current model is only based on medication waste resulting from a switch, which was calculated assuming that the dispensing date of the bDMARD is also the start date of the therapy. Additionally, there are many other sources of medication waste amongst patients who do not switch, such as discontinuation of treatment or non-adherence (3,31), which were not taken into account and therefore, the results are likely an underestimation of absolute quantity of unused bDMARDs. Secondly, data used for this study is from a single site and due to it being highly specific, the generalization of the results is limited. Additionally, the differences in prescribing, data kept in electronic records, patient populations, treatments on which this predictive random forest model is built on, could also make it difficult to apply to other institutions. However, the random forest approach is very general and could be retrained and adjusted to account for these differences. Thirdly, the data could include prescribing bias, as this study is based on data from a single site, namely a third line hospital, highly specified for rheumatic disorders, the method of prescribing medication differs when compared to other hospitals in the Netherlands. Therefore, it is also possible that the amount of waste is already minimized due to optimized methods. Finally, this study was done retrospectively looking at all RA patients which fell in the inclusion period. In this way it was possible to include more patients than by selecting patients and following their treatment plan since first diagnosis. According to Table 1, the average disease duration is 9.89 years, which indicates that the data is quite homogenous. This could be a reason why there is a low indication of waste, as RA patients often switch medications in the first few years of treatment (24).

In pharmacy, machine learning has been successfully used as a tool to predict outcomes of treatment protocols (12,13), and now the application of machine learning in predicting

medication waste has been evaluated. Studies have already shown the usefulness of random forest but to our knowledge, it has not been applied before in preventing medication waste. As the aim was to test for the possibility to apply machine learning in identifying risk factors in predicting medication waste, minimal editing was done to the data set. This provided the true potential of the application. There are ways to improve the sensitivity, specificity and predictive value of the model which can add more value in a clinical setting and successfully identify predictors for medication waste which can then be prevented.

While the results of the predictive model are poor, it can be concluded that the random forest prediction model is a promising approach for predicting medication waste. Moreover, this model demonstrated the use of observational data that is available at hospital and pharmacy data bases.

CONCLUSION

This study shows that machine learning has potential in preventing medicine waste related activities. Further improvements need to be made in the predictive model to increase its sensitivity, specificity and positive prediction value. This could be done by down sampling, reducing features and correcting for correlations. Only then, would machine learning have potential to be used in a clinical setting in pharmacy.

REFERENCES

1. Westhovens R, Cole JC, Li T, Martin M, MacLean R, Lin P, et al. Improved health-related quality of life for rheumatoid arthritis patients treated with abatacept who have inadequate response to anti-TNF therapy in a double-blind, placebo-controlled, multicentre randomized clinical trial. Rheumatology. 2006;45(10).

2. Gerhold K, Richter A, Schneider M, Bergerhausen HJ, Demary W, Liebhaber A, et al. Health-related quality of life in patients with long-standing rheumatoid arthritis in the era of biologics: Data from the German biologics register RABBIT. Rheumatology. 2015;54(10).

3. Strand V, Miller P, Williams SA, Saunders K, Grant S, Kremer J. Discontinuation of Biologic Therapy in Rheumatoid Arthritis: Analysis from the Corrona RA Registry. Rheumatol Ther. 2017;4(2).

4. Furst DE, Keystone EC, So AK, Braun J, Breedveld FC, Burmester GR, et al. Updated consensus statement on biological agents for the treatment of rheumatic diseases, 2012. Ann Rheum Dis. 2013;72(SUPPL. 2).

5. 2019 Drug Trend Report [Internet]. 2020 [cited 2020 Nov 20]. Available from: https://www.express-scripts.com/corporate/drug-trend-report#2019-by-the-numbers

6. Braund R, Chuah F, Gilbert R, Gn G, Soh A, Tan LY. Identification of the reasons for medication returns Original Scientific Papers Original. N Z Fam Physician. 2008;35(4):248–52.

7. Ekedahl ABE. Reasons why medicines are returned to Swedish pharmacies unused. Pharm World Sci. 2006;28:352–8.

8. Bekker CL, Melis EJ, Egberts ACG, Bouvy ML, Gardarsdottir H, van den Bemt BJF. Quantity and economic value of unused oral anti-cancer and biological disease-modifying anti-rheumatic drugs among outpatient pharmacy patients who discontinue therapy. Res Soc Adm Pharm. 2019;

9. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.

10. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Futur Healthc J. 2019;6(2):94–8.

11. Benzidia S, Makaoui N, Bentahar O. The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. Technol Forecast Soc Change. 2021;165(January):120557.

12. Machado G, Mendoza MR, Corbellini LG. What variables are important in predicting bovine viral diarrhea virus? A random forest approach. Vet Res. 2015;46(1).

13. Lee SI, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun. 2018;9(1).

14. Breiman L, Freidman J, Stone CJ, Olshen RA. Classification and Regression Trees. 1984.

15. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18–22.

16. Hastie T, Tibshirani R, Friendman J. The Elements of Statistical Learning. 2008.

17. van Riel PLCM, Renskers L. The Disease Activity Score (DAS) and the Disease Activity Score using 28 joint counts (DAS28) in the management of rheumatoid arthritis. Clin Exp Rheumatol. 2016;34(4):40–4.

18. Salaffi F, Ciapetti A. Clinical disease activity assessments in rheumatoid arthritis. Int J Clin Rheumtol. 2013;8(3):347–60.

19. Bekker CL, Gardarsdottir H, Egberts TCG, Bouvy ML, van den Bemt BJF. Redispensing of medicines unused by patients: a qualitative study among stakeholders. Int J Clin Pharm. 2017;39(1).

20. Carli C, Ehlin A, Klareskog L, Lindblad S, Montgomery S. Trends in disease modifying antirheumatic drug prescription in early rheumatoid arthritis are influenced more by hospital setting than patient or disease characteristics. Ann Rheum Dis. 2006;65:1102–5.

21. Larison B, Njabo KY, Chasar A, Fuller T, Harrigan RJ, Smith TB. Spillover of pH1N1 to swine in Cameroon: An investigation of risk factors. BMC Vet Res. 2014;10(1):1–8.

22. Preventing Waste by Limiting Prescription [Internet]. Ministry of Health Welfare and Sports. 2016 [cited 2021 Jul 2]. Available from: https://www.rijksoverheid.nl/ministeries/ministerie-van-volksgezondheid-welzijn-en-sport/ nieuws/2016/11/29/afspraak-verspilling-voorkomen-door-korter-voorschrijven

23. van der Heijde DM, van 't Hof M, van Riel PL, van Leeuwen MA, van Rijswijk MH, van de Putte LB. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. Ann Rheum Dis. 1992;51:177–81.

24. Ramiro S, Landewé R, Van Der Heijde D, Harrison D, Collier D, Michaud K. Discontinuation rates of biologics in patients with rheumatoid arthritis: Are TNF inhibitors different from non-TNF inhibitors? RMD Open. 2015;1(1).

25.   Ollier WER, Harrison B, Symmons D. What is the natural history of rheumatoid arthritis? Best Pract Res Clin Rheumatol. 2001;15(1):27–48.

26.   Frisell T, Baecklund E, Bengtsson K, Di Giuseppe D, Forsblad-D'elia H, Askling J. Patient characteristics influence the choice of biological drug in RA, and will make non-TNFi biologics appear more harmful than TNFi biologics. Ann Rheum Dis. 2018;77:650–7.

27.   Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet. 2018;(19).

28.   Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. Stat Comput. 2017;27(3):659–78.

29.   Piquemal J-P, Isayev O, Christensen AS, Banerjee P, Dehnbostel FO, Preissner R. Prediction Is a Balancing Act: Importance of Sampling Methods to Balance Sensitivity and Specificity of Predictive Models Based on Imbalanced Chemical Data Sets. Front Chem. 2018;1:362.

30.   He H, Garcia E. Learning from Imbalanced Data. In: IEEE Transactions on Knowledge and Data Engineering. 2009. p. 1263–84.

31.   Bell C, Petford S, Tibbetts L-M, Douglas KM. THU0182 Waste Not, Want Not: Is Anti-TNF Wastage Preventable? Ann Rheum Dis. 2013;72(Suppl 3).

LIST OF ABBREVIATIONS

ATC	Anatomical Therapeutic Chemical
bDMARDs	Biological Disease Modifying Anti-Rheumatic Drugs
CCP	Cyclic Citrullinated Peptide
CRP	C-Reactive Protein
DAS	Disease Activity Score
HAQ	Health Assessment Questionnaire
Hb	Hemoglobin
IQR	Interquartile Range
ML	Machine Learning
PVV	Positive Prediction Value
RA	Rheumatoid Arthritis
RADAI	Rheumatoid Arthritis Disease Activity
RF	Rheumatoid Factor
RF	Random Forest
SD	Standard Deviations
VAS	Visual Analogue Scale

APPENDIX

Appendix A. Biological disease-modifying anti-rheumatic drugs (bDMARDs) included in this study

Active ingredient	Brand bDMARD
Abatacept	Orencia 125mg/1ml Pen
	Orencia 125mg/1ml WWSP
Adalimumab	Humira 40mg/0,4ml injv pen
	Humira 40mg/0,4ml injv wwsp
	Humira 40mg/0,8ml injv pen
	Humira 40mg/0,8ml injv wwsp
Anakinra*	Kineret 100mg/wwsp injvlst
Certolizumab	Cimzia 200mg/1ml injv wwsp
	Cimzia 200mg/1ml injvl patr
	Cimzia 200mg/1ml injvls pen
Etanercept	Benepali 25mg/0,5ml ww injv
	Benepali 50mg/ml pen injvls
	Benepali 50mg/ml wwsp injvl
	Enbrel 25mg injpdr fl+sv+t
	Enbrel 25mg/0,5ml wwsp injv
	Enbrel 50mg/ml wwsp injvlst
	Enbrel myclic 50mg/pen inj
Golimumab	Simponi 50mg/0,5ml injv pen
Sarilumab	Kevzara 150mg/1,14ml wwsp
	Kevzara 200mg/1,14ml pen
	Kevzara 200mg/1,14ml wwsp
Secukinumab	Cosentyx 150mg/1ml injv pen
	Cosentyx 150mg/1ml injv wws
Tocilizumab	Roactemra 162mg/0,9ml in ww
	Roactemra 162mg/0,9ml pen
Ustekinumab*	Stelara 45mg/0,5ml injvl ws

*These medications were not wasted