A PSO/ACO Approach to Knowledge Discovery in a Pharmacovigilance Context

Margarita Sordo, PhD
Laboratory of Computer Science,
Massachusetts General Hospital,
Harvard Medical School
1 Constitution Center
Charlestown, MA. USA
1 (617) 643 5894
msordo@partners.org

Gabriela Ochoa, PhD
Automated Scheduling, Optimisation,
and Planning (ASAP) Group
School of Computer Science,
University of Nottingham
Nottingham, UK
44 (115) 85 15544
gxo@cs.nott.ac.uk

Shawn N. Murphy, MD, PhD
Laboratory of Computer Science,
Massachusetts General Hospital,
Harvard Medical School
1 Constitution Center
Charlestown, MA. USA
1 (617) 643 5894
snmurphy@partners.org

ABSTRACT

We propose and evaluate the use of a PSO/ACO methodology for classification and rule discovery in the context of medication postmarketing surveillance or pharmacovigilance. Our study considers a large data set of diabetic patients on two widely used antidiabetic drugs (rosiglitazone and pioglitazone), and the risk of myocardial infarction as an adverse effect. The goal is to determine the presence of previously undetected causal relationships between therapeutics, patient characteristics, and adverse medication outcomes. Since the proposed approach is able to discover classification rules, the elicited knowledge may suggest new hypotheses regarding associations between risk factors and an adverse event. Our classification results show high accuracy. Furthermore, several medication-related rules were discovered and analyzed. The elicited rules support previous studies from the medical literature. Moreover, one of the studied antidiabetic drugs (rosiglitazone) was found to have a significant higher risk of an adverse event on diabetic, hypertensive patients, as compared to the other drug. This last finding suggests that pioglitazone may have a protective effect against myocardial infarction on diabetic, hypertensive patients.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning – *Knowledge acquisition.*
I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search – *Heuristic methods*

General Terms

Algorithms, Measurement, Experimentation.

Keywords

Swarm Intelligence, Ant Algorithms, PSO/ACO, Knowledge Discovery, Genetic Based Machine Learning, Postmarketing Surveillance, Pharmacovigilance, Healthcare.

1. INTRODUCTION

The US Food and Drug Administration (FDA) defines an Adverse Drug Event (ADE) as "any incident where the use of a medication at any dose is suspected to have resulted in adverse outcome in a patient".

Given the limitations of premarketing trials, e.g. highly selected patient populations and limited duration of studies, often times unanticipated rare adverse events go undetected, and they only become more apparent when they reach the general population. Widely prescribed medications pose a substantial risk of previously undiscovered population-level effects during premarketing trials. Detection of adverse events relies mainly on three sources of information, namely a) data gathered from premarketing clinical trials; b) voluntary reporting of adverse events in postmarketing phase and; c) information gathered from postmarketing observational studies. For example, several postmarketing research studies have indicated a strong correlation between Cyclooxygenase-2 (COX-2) selective inhibitors, a class of non-steroidal anti-inflammatory drugs (NSAIDs), with an increase in the risk of myocardial infarction (MI) [1][5][11][16][24][29]. This was particularly true for Rofecoxib (Vioxx) which was withdrawn from the market in September, 2004 [9].

The work reported herein is part of an observational retrospective cohort study of patients on diabetic medications who may be at an increased risk of coronary heart disease (CHD) [3]. CHD is defined as acute myocardial infarction requiring hospitalization. We focus our analysis on two antidiabetic oral medications, rosiglitazone and pioglitazone.

The purpose of the present study is two-fold: First, we seek to evaluate potential benefits of PSO/ACO methodology as an adjunct to more traditional statistical methods for postmarketing surveillance or pharmacovigilance (hereafter we will use both terms indistinctively). The goal is to determine, using data from electronic medical records, the presence of previously undetected causal relationships between therapeutics, baseline characteristics (e.g. gender, race, age), comorbidities and adverse events. Second, use the elicited knowledge to develop potentially new hypotheses as to suspected associations between all risk factors involved that may play a critical role in the adverse outcome.

The remainder of this paper is organized as follows: Section 2 presents a brief overview of the rationale for a nation-wide and institution-wide pharmacovigilance efforts, in combination with

large electronic patient databases. Section 3 describes the patient data used for the current study, while section 4 briefly describes the PSO/ACO2 algorithm. In sections 5 and 6 we present our results and findings. Finally, in section 7 we draw some conclusions and discuss possible future work.

2. BACKGROUND

In 2006, the Institute of Medicine (IOM) issued a report, entitled *The Future of Drug Safety—Promoting and Protecting the Health of the Public* [20]. Among other suggestions, the IOM report recommended that the FDA identify ways to access other health-related databases and create a public-private partnership to support safety and efficacy studies. As a result, the FDA has been fostering public-private collaborations, leveraging increasingly available large electronic patient databases and exploring new, emerging technologies to further advance the safety and quality of all realms of healthcare. Partners Healthcare System is one of five institutions which, in conjunction with the eHealth Initiative (eHI) and the FDA, is collaborating in a nation-wide effort to develop novel health information technology tools to create an active drug safety surveillance system across the U.S.

Independent from nation-wide efforts, Partners Healthcare System has been carrying out patient-population pharmacovigilance research using patient data from the Research Patient Data Registry (RPDR). Partners Healthcare System is a non-profit, integrated health system that includes Brigham and Women's Hospital and Massachusetts General Hospital. The RPDR is a centralized data warehouse containing clinical data such as patient demographic information, dates, medication, diagnosis information, and discharge summaries.

3. DATA

Institutional Review Board approval was obtained prior to selecting a group of 2,185 diabetic patients on rosiglitazone or pioglitazone monotherapy from a cohort study of 34,252 patients on diabetic medications. These two medications belong to the thiazolidinediones (TZs) class of oral hypoglycemic medications. Both rosiglitazone and pioglitazone were introduced to the market in 1999. Both medications have shown to increase the risk of congestive heart failure (CHF) [30], and rosiglitazone has been associated with increased risk of myocardial infarction when compared to control groups [24][27].

Only patients over 18 years of age with at least one record of prescription as an outpatient, or dispensation as an inpatient, of either rosiglitazone (n = 1594) or pioglitazone (n= 591) between January 1st, 2000 and December 31st, 2006 were included in the study. Selected patients should be on monotherapy (single antidiabetic medication) for the whole duration of the study. Our definition of monotherapy was more stringent than the definition used in the larger cohort study in order to limit confounding factors between medication intake and a possible adverse event. The outcome of the study was the incidence of acute MI (identified by ICD9 code of 410.x) requiring hospitalization. Characteristics of rosiglitazone and pioglitazone users were similar in demographics and risk factors (Table 1).

Data for each patient consists of the following thirteen potential risk factors. All data are nominal values:

i) Age at time of enrollment within one of the following 10-year intervals starting at 20 years of age: {[20-30), [30-40), [40-50), [50-60), [60-70), [70-80), [80-)};

ii) Gender (M/F);

iii) Race as one of {White, Black, Hispanic, Asian, Native American, Multiracial, Unknown};

iv) Medication (Rosiglitazone / Pioglitazone);

v) History of cardiovascular disease (Y/N). Diagnoses considered are coronary artery disease, angina, congestive heart failure, cerebrovascular accident, percutaneous coronary intervention, coronary artery bypass graft surgery;

vi) Prior MI (Y/N);

vii) HBA1C as indicator for disease management (Y: patient has been monitored/ N: patient has not been monitored);

viii) HBA1C > 8 as indicator of poor glycemic control and disease severity (Y/N);

ix) Creatinine as indicator of disease management (Y/N);

x) Creatinine > 2 as indicator of chronic renal insufficiency and disease severity (Y/N);

xi) Hypertension (Y/N) indicated by use of any hypertensive drugs (See Table 1);

xii) Hyperlipidemia (Y/N) indicated by use of any anti-hyperlipidemic medications (See Table 1);

xiii) Hospitalizations/ED visits as proxy for severity of disease (Y/N).

Table 1. Characteristics table for patients on rosiglitazone or pioglitazone monotherapy. Values are number (%) unless otherwise indicated.

	Rosiglitazone (n=1594)		Pioglitazone (n=591)	
Age AVG(SD)	64.50	(11.33)	63.81	(11.58)
Gender (female)	759	(47.61)	274	(46.36)
MI outcome	257	(16.12)	58	(9.81)
Prior MI	207	(12.98)	68	(11.50)
Prior CVD	525	(32.93)	184	(31.13)
Hypertension	1427	(89.52)	511	(86.46)
Hyperlipidemia	1238	(77.66)	422	(71.40)
Chronic renal insufficiency (Cr>2)	295	(18.50)	86	(14.55)
Concomitant Therapy	1366	(85.69)	490	(82.91)
Antihyperlipidemic	1116	(70.01)	376	(63.62)
Combination	32	(2.00)	13	(2.20)
Fibrates	149	(9.34)	49	(8.29)
Statins	1075	(67.44)	357	(60.40)
Antihypertensive	1277	(80.11)	452	(76.48)
ACE Inhibitors	887	(55.64)	308	(52.11)
Angiotensin-II antagonists	326	(20.45)	114	(19.28)
Beta Blockers	877	(55.01)	274	(46.36)
Calcium Channel Blockers	453	(28.41)	137	(23.18)
Combinations	215	(13.48)	81	(13.70)

Alpha-Beta	138	(8.65)	33	(5.58)
Potassium Sparing Diuretics	5	(0.31)	3	(0.50)
Unclassified Combinations	10	(0.62)	3	(0.50)
ED Visits / hospital-izations AVG(SD)	1.82	(3.27)	1.814	(3.39)

4. PSO/ACO2

In this section we present a brief overview of the particle swarm optimization/ant colony optimization (PSO/ACO2) algorithm proposed by Holden and Freitas [17]. Both PSO and ACO algorithms mimic a population of decentralized, self-organized individuals that collectively work towards finding best solution(s) through an iterative searching process. Convergence to an optimal or near optimal solution is reached by social interaction amongst individuals, either by exchanging information with local neighbors –in the case of particles in PSO- or by updating a pheromone trail –in the case of ants in ACO.

PSO/ACO2 has been mainly used to discover classification rules in the context of data mining. This algorithm is capable of handling nominal attribute values without converting them into numbers, as well as continuous data values.

```
Set rule set to empty  RS= ∅
FOR EACH class C
   add all training examples to TS
   WHILE (number of uncovered examples belonging to
      class C > MaxUncovExampPerClass)
         Discover best nominal rule Rule for the class C
         When applicable, add continuous terms to Rule
         Return best discovered rule BestRule
         Prune BestRule
         Add BestRule to rule set:  RS = RS ∪ BestRule
         Update TS by removing correctly classified
            examples by discovered rule:
            TS = TS – {correctly classified examples}
   END WHILE
END FOR
Order rules in RS by descending quality
```

Figure 1. Sequential algorithm used by the PSO/ACO2 for knowledge discovery – in pseudocode (from [17][18]).

Discovered knowledge is represented in the form of rules, where each rule consists of a set of one or more <*attribute operator value*> triplets (antecedent) and a consequent indicating the class to which the classified object belongs. For nominal values, the operator used is "=", whereas for continuous attribute values "<=", ">" are used:

IF <*attrib op value*> AND…AND <*attrib op value*> THEN <*class*>

PSO/ACO2 algorithm depicted in Figure 1 (based on [17][18]) carries out the knowledge discovery process starting with an empty rule set (*RS*), sequentially searching the space of possible solutions to discover one classification rule at a time as follows: For each class *C*, the algorithm iterates through a set of training examples (*TS*) from which rules will be created. In a first step, only rules with nominal attributes are evaluated and the discovered rule (*Rule*) is returned. If there are continuous attributes in the *TS*, then the created rule is not yet complete and the algorithm performs a second step where it checks those attributes with continuous values. It is worth remembering that for nominal attributes the comparison operator used is "=" whereas, for continuous attributes both "<=" and ">" are used to define the upper and lower bounds of the range of possible values for this attribute. The best discovered rule (*BestRule*) is pruned and added to the rule set. An example that satisfies all the triplets <*attribute operator value*> in the antecedent of the rule and belongs to the class assigned by the rule is considered correctly classified. These correctly classified examples are removed from the *TS*. The iteration process continues until the number of unclassified examples for the current class *C* falls below a predefined threshold (*MaxUncovExampPerClass*). Once this threshold is reached, all the removed training examples are returned to *TS*, and the algorithm continues execution for the next class C_i.

This section presented a brief description of the PSO/ACO algorithm. For a more comprehensive description, see [17][18].

5. RESULTS

For our experiments, we used a freely available Java implementation of the PSO/ACO2 v1.0 rule induction algorithm [19] and the previously described data set of 2,185 diabetic patients on rosiglitazone or pioglitazone monotherapy (section 3). We used the standard 10-fold cross validation, precision fitness function, PSO continuous optimizer, and 200 iterations for all four experiments. The only varying parameter between experiments was the number of particles, which was set to 10^2, 15^2, 20^2 and 25^2 for each experiment respectively.

It can be seen from Table 2 that the classification accuracy for the PSO/ACO2 algorithm is very similar for all four configurations, with 20^2 being slightly better than the rest in terms of avg. classification rate and standard deviation.

Table 2. Classification accuracy on diabetes dataset for the PSO/ACO2 algorithm with varying number of particles.

Particles	**10^2**	**15^2**	**20^2**	**25^2**
Accuracy (Avg±SD)	85.72 ±2.41	86.09 ±2.42	86.09 ± 2.17	86.04 ± 2.25

Given the context of our study, we focused our analysis of elicited knowledge on medication-related rules, since one of our goals is to elucidate whether a) there are causal relationships between therapeutics (medications), comorbidities, and baseline characteristics, and adverse events and b) we can detect such signals with the PSO/ACO2 algorithm. The next section presents an analysis of elicited rules.

6. KNOWLEDGE DISCOVERY

In this section we present our analysis of discovered rules using the results produced by PSO/ACO2 with 20^2 particles. We analyzed all medication-related rules found. These rules are listed in Table 3.

We validated the discovered rules by a) providing references to literature supporting similar findings and/or b) performing a crude relative risk analysis of variables in each rule.

The relative risk (RR) estimates the magnitude of an association between potential factors and an adverse event. It measures the

incidence of the event in the exposed group compared with the non-exposed group. A relative risk of 1.0 indicates that the incidence rates in both groups are identical and there is no association between the potential factors and the outcome. A relative risk of less than 1.0 indicates a negative association, or protective effect between potential factors and the outcome under study, while a relative risk greater than 1.0 indicates a positive association or an increased risk of an adverse event [15].

In the following section we will see that some of the causal relationships have a truly protective effect, that is, a relative risk of less than 1.0 (rules 1-3 in Table 3, analyzed in section 6.1.1), while others, may show a protective effect in comparison – as in the case of rule 5 in Table 3, analyzed in section 6.1.2.

Table 3. List of discovered medication-related rules. Medication = R indicates Rosiglitazone, and Medication = P indicates Pioglitazone

If Medication = P then no_event
If Medication = P and hasHBA1C = Y then no_event
If Medication = P and Hospitalizations/ED = N then no_event
If Medication = P and PriorMI = N then no_event
If Medication = R and Hypertension = N then no_event
If Medication = R and PriorMI = N and hasCreatGt2 = N then no_event
If Medication = R and Gender = M and Age_Range = 50-60 then has_event
If Medication = R and Age_Range = 60-70 and hasHBA1C = Y then has_event

6.1 Analysis of Discovered Rules

6.1.1 Pioglitazone-Related Rules

There are four pioglitazone-related rules (Table 3). The first rule refers to the administration of pioglitazone with attributes taken into consideration. The rule suggests no association between pioglitazone and the possibility of an adverse event (MI). Our calculations indicate that the relative risk (RR) of having an adverse event if a patient is on pioglitazone compared to a patient on rosiglitazone is 0.6086 (Confidence Interval (CI) 0.467 – 0.7933). This is consistent with reports from [8] indicating that pioglitazone may have a neutral to favorable effect towards cardiovascular adverse events. Similarly, [10] and [31] have reported that rosiglitazone may have a higher risk of cardiovascular events compared to pioglitazone.

The second rule in Table 3: "*If Medication = P and hasHBA1C = Y then no_event*" indicates that if a patient is taking pioglitazone (Medication = P) and patient's HBA1C has been monitored (hasHBA1C = Y) then there is no event. It is worth remembering that HBA1C is used as a proxy for glycemic control and disease management. hasHBA1C = Y indicates that the patient's blood sugar has been monitored. A relative risk of 0.65 (CI 0.42 – 0.99) indicates that a patient on pioglitazone with monitoring of HBA1C is less likely to have an adverse event than patients on pioglitazone with no monitoring of HBA1C [21] [26].

The third rule in Table 3: "*If Medication = P and Hospitalizations/ED visits = N then no_event*" indicates that a patient on pioglitazone with no hospitalizations or emergency department (ED) visits is less likely to have an event with a RR = 0.423; CI 0.269 – 0.665). Hospitalizations/ED visits is used as a proxy for severity of disease, so having no visits indicates that the patient is relatively healthy for his/her condition and there are no contributing factors that may increase the risk of an adverse event.

The fourth rule Table 3: "*If Medication = P and PriorMI = N then no_event*" indicates that a patient on pioglitazone is at lesser risk of having and adverse event if there is no prior myocardial infarction (MI). A myocardial infarction may compromise the function of the heart and may increase the risk of subsequent events. This is particularly true in patients with diabetes [6] [13]. Patients on pioglitazone with evidence of having an MI prior to the study had a 2.68 (CI 1.61 – 4.46) risk of having an event when compared to patients on pioglitazone who did not have an MI prior to Jan 1st, 2000. For patients on rosiglitazone there is a slightly higher risk of having an MI if the patient has had a prior event (RR 3.61, CI 2.98 – 4.37).

6.1.2 Rosiglitazone-Related Rules

Rules five to eight in Table 3 depict causal relationships between rosiglitazone and an adverse event. Rule five "*If Medication = R and Hypertension = N then no_event*" indicates that a patient on Rosiglitazone with no hypertension may have a lesser risk of having an adverse event. Our calculations indicate that hypertensive patients on rosiglitazone have a relative risk of 9.90 (CI 8.46 – 11.60) of having an adverse event, compared to hypertensive patients on Pioglitazone (RR 2.87; CI 2.02 – 4.07). Given the fact that coronary artery disease and hypertension are common risk factors in patients with diabetes [14][22], patients presenting these conditions may be more susceptible of having an adverse event [23].

Rule six "*If Medication = R and PriorMI = N and hasCreatGt2 = N then no_event*" indicates that patients on rosiglitazone with no prior MI and creatinine levels within normal values are at considerably lower risk of having an event (RR = 0.44; CI 0.37 – 0.53) when compared to patients on rosiglitazone with abnormal creatinine levels (RR 2.253; CI 1.828 – 2.777)). [1][23] indicate that abnormal kidney function increases the risk of myocardial infarction and death.

Rule seven "*If Medication = R and Gender = M and Age_Range = 50-60 then has_event*" indicates that a 50-60 y/o male patient on rosiglitazone is slightly more likely to have an adverse event (RR = 1.735; CI 1.155 – 2.606) than a female patient in the same age group. This is consistent with reports from [2] indicating that although both diabetic men and women are at higher risk of an adverse cardiovascular event than non-diabetic patients, diabetic men in this age range are at a higher risk than diabetic women.

The last rule in Table 3 "*If Medication = R and Age_Range = 60-70 and hasHBA1C = Y then has_event*". The rule itself indicates that patients within this age range and with HBA1C monitored may have an event. Our calculations indicate that patients on rosiglitazone within this age range and with monitored HBA1C do not seem to be particularly at risk of having an event (RR= 0.70; CI 0.517 – 0.948) when compared to all patients on rosiglitazone and HBA1C = Y (RR= 0.769; CI 0.639 – 0.925). This suggests that: a) patients in this particular age range may not be at a higher risk of an adverse event and; b) appropriate glycemic monitoring may reduce the overall risk of an adverse event and improve disease management [25].

6.2 Visualizing Causal Relationships

We explored the applicability of a heatmap to visualize causal relationships in an easy-to-interpret manner. Similar to a weather map, where temperature is encoded by color, in a heatmap we depict the potential risk of an adverse event given specific combination of factors (e.g. medication, comorbidities, baseline characteristics) in terms of 'temperature', where 'cold temperatures' represent a low relative risk, and 'hot temperatures' represent a high relative risk of an adverse event.

We display potential causal relationships in a two-dimensional heatmap where the color of a cell in the *x*,*y* position depicts the relative risk for patients on medication *x* who had a comorbidty *y* of having an adverse event. The first two rows in Figure 2 depict the relative risk of having an adverse event for patients on rosiglitazone (R) and pioglitazone (P). The third row depicts the risk of having an event regardless of the medication (overall). For example, the cell in position row 2, column 3 in Figure 2 is a color-coded representation of the (low) relative risk of having an adverse event for patients on pioglitazone who had HBA1C levels monitored, as detected by rule 2 in Table 3.

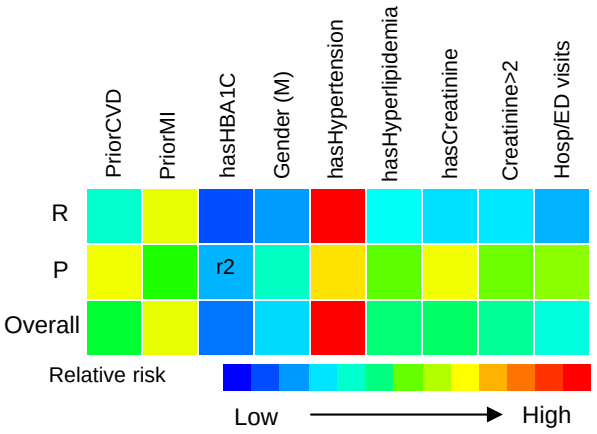

Figure 2. Heatmap depicting relative risk of an adverse event given a risk factor for patients on rosiglitazone (R), pioglitazone (P) and regardless of medication (overall).

Although exploratory, we have found this strategy of representing the strength of causal relationships by colors extremely powerful. Since the color of a cell depicts a quantitative risk relative to other cells in the same column, it is possible for users to identify potential trends and outliers in data. For example, column 5 in Figure 2 depicts the risk of an adverse event in hypertensive patients on rosiglitazone (row 1), pioglitazone (row 2) and overall (row 3). It shows that hypertensive patients on rosiglitazone may have a potentially higher risk of having an adverse events compared to hypertensive patients on pioglitazone. Further, since hypertensive patients on pioglitazone seem to have a lower relative risk, this could be interpreted as pioglitazone having a protective effect against myocardial infarction on hypertensive patients.

This section presented the analysis of medication-related elicited rules. In the following section we analyze our findings and discuss future work.

7. DISCUSSION AND FUTURE WORK

Overall, elicited causal relationships between therapeutics, patient characteristics and an adverse event were consistent with findings in medical literature. These findings support our initial assumptions as to the suitability of PSO/ACO as an alternative to more traditional methods for knowledge discovery in a pharmacovigilance context. Furthermore, due to the inherent nature of the applied algorithm, elicited rules were seamlessly coupled into a visual display easy to understand, thereby increasing the applicability and understandability of these findings.

We expect that the elicited knowledge may provide critical insight into potentially worrisome combination of factors that may increase the risk of an adverse event. For example, depicted in Figure 2, column five, we see that diabetic hypertensive patients on rosiglitazone are at higher risk for an adverse event while pioglitazone seems to have a protective effect on diabetic hypertensive patients. This observation prompts two hypotheses worthy of further investigation: a) is this increased risk due to the fact that thiazolidinediones in general can among other things, increase fluid retention, and hence increase blood pressure [22]? Is this particularly true for rosiglitazone but not for pioglitazone? Or b) could this be due to a possible drug-drug interaction between rosiglitazone and antihypertensive drugs?

In Summary, our findings are by no means exhaustive, but demonstrate that the potential benefits of PSO/ACO for knowledge elicitation are many-fold: The approach itself is capable of discovering causal relationships in the form of rules from patient data extracted from electronic medical records; discovered rules can be easily mapped into heatmaps – or any other visual aid - to provide users with 'at-a-glance' immediate interpretation of findings; rules could be seamlessly incorporated into monitoring systems [12]; elicited knowledge may serve to develop new hypotheses as to suspected associations between all risk factors involved that may play a critical role in the adverse outcome.

Important directions for future work include: a) extending our analysis of discovered rules; b) further investigate findings of diabetic hypertensive patients and their use of antihypertensive medications; c) improve visual display of results; d) explore other knowledge discovery methods and compare results and; e) extend our model to include other possible adverse events.

8. ACKNOWLEDGEMENTS

We would like to thank Drs. Mena Macedo and Margarita Sanchez for their advice in writing this manuscript.

9. REFERENCES

[1] Beddhu S, Allen-Brady K, Cheung AK, Horne BD, Bair T, Muhlestein JB and Anderson JL. Impact of renal failure on the risk of myocardial infarction and death. *Kidney International* (2002) 62, 1776–1783.

[2] Booth GL, Kapral MK, Fung KT, Jack V. Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *The Lancet* 2006; 368(9529):29-36.

[3] Brownstein JS, Murphy SN, Goldfine AB, Grant RW, Sordo M, Gainer V, Colecchi J, Dubey A, Nathan DM, Glaser JP, Kohane IS. Real-time surveillance of cardiovascular risk associated with diabetic medications from electronic medical records.*Tech Report* 2009

[4] Brownstein JS, Sordo M, Kohane IS, Mandl KD. The tell-tale heart: population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction. *PLoS ONE* 2007;2(9):e840.

[5] Caldwell B, Aldington S, Weatherall M, Shirtcliffe P, Beasley R (2006) Risk of cardiovascular events and celecoxib: a systematic review and meta-analysis. *J R Soc Med* 99: 132–140.

[6] Canto JG, Shlipak MG, Rogers WJ, Malmgren JA, Frederick PD, Lambrew CT, Ornato JP, Barron HV, Kiefe CI. (2000). Prevalence, clinical characteristics, and mortality among patients with myocardial infarction presenting without chest pain. *JAMA*. 2000;283(24): 3223–9.

[7] Carvalho DR, Freitas AA, Ebecken N. Evaluating the Correlation Between Objective Rule Interestingness Measures and Real Human Interest. *PKDD 2005, LNAI 3721,* pp 453-461.

[8] Dormandy JA, Charbonnel B, Eckland DJ, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. *Lancet* 2005;366(9493):1279-89.

[9] FDA issues public health advisory on Vioxx as its manufacturer voluntarily withdraws the product. *FDA News*. 2004 Sep 30. http://www.fda.gov/bbs/topics/news/2004/new01122.html

[10] Gerrits CM, Bhattacharya M, Manthena S, Baran R, Perez A, Kupfer S. A comparison of pioglitazone and rosiglitazone for hospitalization for acute myocardial infarction in type 2 diabetes. *Pharmacoepidemiol Drug Saf* 2007;16(10):1065-71.

[11] Graham DJ, Campen D, Hui R, Spence M, Cheetham C, et al. (2005) Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *Lancet* 365: 475–481.

[12] Greenes RA, Sordo M, Zaccagnini D, Meyer M, Kuperman GJ. Design of a Standards-Based External Rules Engine for Decision Support in a Variety of Application Contexts: Report of a Feasibility Study at Partners HealthCare System. Proceedings of Medical Informatics Association (MedInfo), September, 2004.

[13] Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med* 1998;339(4):229-34.

[14] He J, Ogden LG, Bazzano LA, et al. Risk factors for congestive heart failure in U.S. men and women: NHANES I epidemiologic follow-up study. *Arch Intern Med.* 2001; 161: 996–1002.

[15] Hennekens CH, Buring J. Epidemiology in Medicine. Little Brown and Co. 1987.

[16] Hippisley-Cox J, Coupland C (2005) Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *BMJ* 330: 1366.

[17] Holden N, Freitas AA. A Hybrid PSO/ACO algorithm for classification. *Proceedings of the 9th Genetic and Evolutionary Computation Conference Workshop on Particle Swarms: The Second Decade (GECCO '07)*, pp. 2745-2750, ACM Press, London, UK, July 2007.

[18] Holden N, Freitas AA. A Hybrid PSO/ACO Algorithm for Discovering Classification Rules in Data Mining. *Journal of Artificial evolution and Applications.* Vol 2008, Article id 316145, 11 pages.

[19] Holden N, Freitas AA. PSO/ACO2 Java application. Downloads at http://sourceforge.net/projects/psoaco2/

[20] Institute of Medicine, *The Future of Drug Safety- Promoting and Protecting the Health of the Public*, September 22, 2006, http://www.iom.edu/.

[21] S. B. King III. Acute myocardial infarction: are diabetics different? *J. Am. Coll. Cardiol*., May 1, 2000; 35(6): 1513 - 1515.

[22] Nesto R, Libby P. Diabetes mellitus and the cardiovascular system. In: Braunwald E, Zipes DP, Libby P, eds. Heart Disease: A Textbook of Cardiovascular Medicine. 6th ed. Philadelphia, Pa: W.B. Saunders Co; 2001: 2133–2150.

[23] Nichols GA, Hillier TA, Erbey JR, et al. Congestive heart failure in type 2 diabetes: prevalence, incidence, and risk factors. *Diabetes Care.* 2001; 24: 1614–1619

[24] Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes [published correction appears *N Engl J Med*. 2007;357(1): 100]. *N Engl J Med* 2007;356(24):2457-2471.

[25] Qaseem A, Vijan S, Snow V, Cross T, Weiss KB, Owens DK. Glycemic Control and Type 2 Diabetes Mellitus: The Optimal Hemoglobin A1C Targets. A Guidance Statement from the American College of Physicians. *Ann Inter Med* 2007; 147: 417-422.

[26] Singer DE, Nathan DM, Anderson KM, Wilson PW, Evans JC. Association of HbA1c with prevalent cardiovascular disease in the original cohort of the Framingham Heart Study. *Diabetes*, Vol 41, Issue 2 202-208.

[27] Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. *JAMA*. 2007;298(10):1189-1195.

[28] Solomon DH, Schneeweiss S, Glynn RJ, Kiyota Y, Levin R, et al. (2004) Relationship between selective cyclooxygenase-2 inhibitors and acute myocardial infarction in older adults. *Circulation* 109: 2068–2073.

[29] Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, et al. (2005) Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 352: 1071–1080.

[30] Yki-Jarvinen H. Thiazolidinediones. *N Engl J Med*. 2004; 351(11):1106-1118.

[31] Winkelmayer WC, Setoguchi S, Levin R, Solomon DH. Comparison of Cardiovascular Outcomes in Elderly Patients With Diabetes Who Initiated Rosiglitazone vs Pioglitazone Therapy. *Arch Intern Med* 2008;168(21):2368-75.