Applying Anatomical Therapeutic Chemical (ATC) and Critical Term Ontologies to Australian Drug Safety Data for Association Rules and Adverse Event Signalling

Gary Saunders **Sasa Ivkovic** **Ranadhir Ghosh** **John Yearwood**

University of Ballarat, Mount Helen Victoria 3353, Australia,
Email: g.saunders@ballarat.edu.au

Abstract

The sparse nature of voluntarily reported drug safety data benefits from a system that consolidates the massive amount of data into a manageable format for analysis. This has been done for Australian drug safety data by the Australian Adverse Drug Reaction Advisory Committee (ADRAC) for reactions using the systems organ class (SOC) ontology. There has long been a need for a similar kind of grouping to apply to drugs in this type of data. In ADRAC, drugs are currently listed by trade-name, where only some of these trade-names were assigned anatomical-therapeutic-chemical classification (ATC) codes. We assigned an ATC code for each ADRAC trade-name and show that this ontology facilitates the detection of drug class / reaction class associations at various levels of specificity. This allows different views of these associations (even very rare ones) and their significance measured for the development of more sensitive signal detection methods. We report that this ATC classification enables both the grouping of association rule approach that is useful for studying rare associations, and the development of an adverse reaction signal detection method.

1 Introduction

1.1 ADRAC Data

The Australian Adverse Drug Reaction Advisory Committee (ADRAC) database has been developed and maintained by the Therapeutic Goods Administration (TGA) with the aim to detect signals from adverse drug reactions (ADRs) as early as possible. The ADRAC data contain 137,297 voluntarily reported adverse drug reaction records involving 5057 different drugs, based on the 'drug dictionary' used by ADRAC of 7416 available drug terms, and 868 different reactions, based on 1392 available reaction terms. There are many fields in these data, such as patient information (age, weight, and height), expert information (causality, outcome, and suspected drug). There are two groups of ten fields, which are allocated for drug terms and reaction terms. That is, ADRAC data can only accommodate a maximum of ten terms for drugs and reactions. A more detailed account of the ADRAC database is given in Mamedov *et al.* (2002). These data are derived from voluntary reports, some of the problems and advantages of such a reporting system are discussed in Bates (2003), Heeley (2001), Pinkston and Swain (Pinkston, V. & Swain, E. J. 1998), Pirmohamed *et al.* (Pirmohamed, M., Breckenridge, A. M., Kitteringham, N. R. & Park,B. K. 1998), Troutman and Doherty (Troutman, W. G. & Doherty, K. M. 2003), van der Heijden *et al.* (van der Heijden, P. G. M., van Puijenbroek, E. P., van Buuren, S. & van der Hofstede, J. W. 2002), van Puijenbroek *et al.* (van Puijenbroek, E. P., Diemount, W. L. & van Grootheest, K. 2003).

1.2 Data Consolidation

The biggest challenge in summarizing safety data is the need to consolidate the massive amount of data into a manageable format. One way is to group the safety data into K classes characterized by body systems and determined in conjunction with underlying disease and treatments involved. Such pooling of data through coding is especially helpful for rare events Chuang-Stein (1998), Pinkston and Swain (1998). This has been done for Australian drug safety data by ADRAC for ADR terms using the body systems organ class (SOC) grouping Saunders *et al.* Mamedov *et al.* (2003), (Mamedov, M. & Saunders, G. 2004), (Mamedov, M. & Saunders, G. 2002). For more details of SOC information in ADRAC see Saunders (2004).

There has been a need for a similar kind of grouping to apply to drugs in these data, which are currently listed by trade-name Mamedov *et al.* (2003). ADRAC had assigned some of these trade-names anatomical-therapeutic-chemical classification (ATC) codes, but in cases where the trade-name had more than one ingredient, ATC codes were not assigned. This paper reports application the classification described in Saunders (2004), where each trade-name code was assigned a corresponding ATC code.

1.3 Critical Terms

From the third quarter of 1998, a new field has been added at the end of the WHO-ART file which indicates Critical Terms. Critical terms are a subset of adverse reaction terms referring to, or possibly being indicative of, serious disease states, which have been regarded as particularly important to follow up. WHO (2002), (World Health Organization 2002) This component of expert knowledge has been incorporated into ADR signal detection. In this report we weight critical reaction terms to highlight associations which are likely to be of greater interest and particularly for those that are rare.

2 Methods

2.1 Basis for drug classification

The ADRAC database uses the "drug trade-name" field as the main field for drugs. This is not satisfactory because the same drug can have more than one trade-name, which dilutes the information for that drug. Also it is difficult to group drugs using this field. The classification system implemented was the WHO Collaborating Centre for Drug Statistics Methodology ATC codes in order to enable: (i) the grouping of the same substance into one code, and (ii) the grouping of related substances into more general categories. To this purpose the ATC codes for drugs have been purchased by us from Ms Kirsten Myhr, RELIS Øst, Ullevål University Hospital, 0407 OSLO (http://www.whocc.no/).

2.2 ATC embedded code

The ATC encoding system implements an an embedded encoding system, which employs a seven character coding system. As an example, the first member of the system will be used. For level 0, the character is 'A', level 1 is 'A01', level 2 is 'A01A' , level 3 is 'A01AA' and level 4 is 'A01AA01'. The embedded coding can be utilized to simplify database queries and algorithm coding. Employing this embedded coding, which resides in the ATC code string, can be now utilized to aggregate the data to the required level. There was a need, for the purposes of associating ADRAC and ATC ingredient terms, to go to level 4 (7 character). The ATC embedded code has been utilized in our algorithms to traverse the ATC hierarchy (tree). For details of the ATC hierarchy – see Saunders (2004).

2.3 Classification of ADRAC data

All trade-name instances that had ingredients that matched ones in the WHO ATC file were assigned an ATC code. In cases where more than one code was available for a particular ingredient, the most generic one was chosen. The remaining trade-name codes that had not been assigned an ATC code were classified as described in Saunders (2004). For the purposes of this present investigation, in cases where the original drug code had more than one ingredient, a unique code was assigned from ATC codes for combinations of ingredients. This was done in order to preserve a one-to-one correspondence between original drug codes and the ATC codes facilitating comparison of the two drug classification schema. This resulted in 1806 ATC drug terms from the 5081 drug trade-name terms.

2.4 Association Rules

The application of the ATC classification to the ADRAC drugs effectively causes a dramatic reduction in number of drug terms. We had previously tried to apply the methods for grouping association rules Agrawal *et al.* (1993), which are described in Ivkovic (2004), (Ivkovic, Sasa, Yearwood, John & Stranieri, Andrew 2002), (Ivkovic, Sasa, Yearwood, John & Stranieri, Andrew 2003), on ADRAC data and found the number of drug terms were excessive. Now we report that the association rule methods can be usefully applied to the ADRAC data by exploiting the hierarchical structures of the ATC and SOC classifications to produce signals from associations between individual drug terms and reaction terms.

The convention used for the association rules (also see Saunders *et al.* (2005)) is given by:

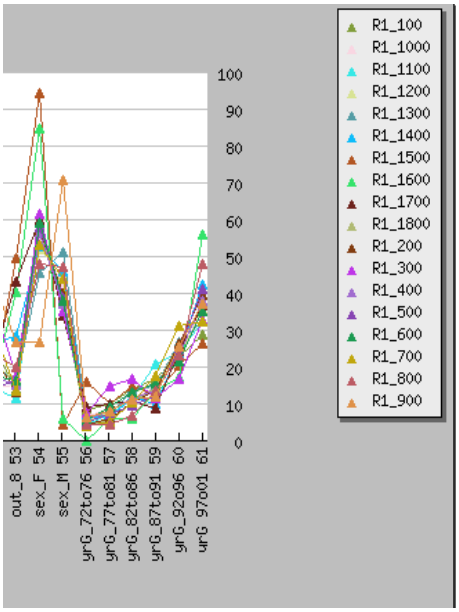

Figure 1: Reaction Classes differences and similarities.

Definition 2.1 *Association Rule Definition: If antecedent ⇒ consequent.*

Example from Figure 1 – highest point for rule 54: $R1\_1500 \Rightarrow sex\_F \quad 94\%$

if first reaction is 1500 ($R1\_1500$*) then records containing sex female (sex_F), confidence = 94%;*

as conditional probability $\mathcal{P}$*:* $\mathcal{P}(sex\_F \mid R1\_1500) = 0.94$;

that is, 94% of records with first reaction (field) being reaction 1500, were female.

2.5 Adverse Event Signalling

Since the ATC ontology made our association rule method useful for investigating the ADRAC data, we decided to develop an adverse event signalling method which also exploited the ATC hierarchy. This method begins at an ATC level specified by the user. It then examines the children of this node and performs a χ^2 test. If any of these children have a significantly higher number of reactions than expected (based on the hypothesis that the children are the same) then their children are tested. The search ends when there are no more children (leaf node), that is the lowest ATC level. Thus drugs are found that have significantly greater numbers of reactions. As well as this the report dates are grouped by month. For a given drug or drug class another algorithm examines the first three month period of reports for that drug. A χ^2 test is applied to the reaction frequencies for these periods. If the χ^2 test is not significant, the next month is added and the test repeated. This is repeated until the last month. To add to this, reactions carry an extra weighting if they belong to the critical terms. In this way the seriousness of the reaction is also evaluated for signal production.

3 Results

3.1 Investigation using Association Rules

Applying a drug classification to ADRAC drug terms makes these data more amenable to analysis by methods that work best with a limited number of variables. The ability to reduce the number of drug terms using the ATC hierarchy facilitates the application of the

association rules data mining algorithms (1993). We report here some preliminary results from grouping association rules by content using ADRAC data.

To demonstrate some of the features of this method we will show some association rules looking at all 18 SOC reaction classes in reaction field 1. In the legend Figure 1 "R1_SOC" **antecedent** represents the SOC of the first reaction. and characteristic number 54 (sex_F – **consequent**) represents sex female. It can be seen that: (i) Some reaction classes contain a many female patients, (ii) others have a very few, (iii) among all patient who reported foetal disorders (reaction class 1500) 94% are female, (iv) among all patient who reported endocrine disorders (reaction class 900) more than 70% are males.

We can filter the graph output by showing very few or many characteristics Figure 2 shows 99 characteristics. Reactions are listed in the legend, where the top entry, R1_100 means the first reaction field is SOC 100 (skin and appendage disorders – see Saunders (2004)); the **consequent** on the x-axis, for example, D1_A02 is the first drug field is drug class A02 (drugs for acid related disorders – see Saunders (2004)), D2_0 means no drug in second drug field, R3_1000 is third reaction field is SOC 1000 (cardiovascular) and R3_0 no reaction in that field, ageG_13to20 means age group 13–20 years, nreac_1 means number of reactions in record equals 1, out_4 outcome code 4 (death as a reaction – see Mamedov *et al.* (2002)), sex_X – sex not recorded, yrG_72to76 means year group 1972–1976. In this graph we can select for age group differences and similarities between reaction classes – see Figure 3.

In Figure 4 we show SOC 1600 (neonatal and infancy disorders) and focus on a rare group of patients with this disorder – 52 patients (only 0.04%). We are interested to explore further the characteristics of 13 to 20 years old patients with reaction 1600, displayed in Figure 5, where 33% of patients with neonatal and infancy disorder are teenagers – 17 patients. All teenage patients with this disorder are female (17/17 – 100%). Almost half recovered without sequel (7/17 – 41%). 30% of these patients took drug class N06 (psychoanaleptics). 30% of these patients took drug class N02 (analgesics). 5% took NO2 as a second drug.

Thus this method is able to 'drill down' with the aid of the drug and reaction ontologies to reach fine level associations. In the following section we apply the drug and reaction ontologies, along with the critical term ontology to develop a more sensitive adverse event signalling method than presently exists.

3.2 Ontologies to facilitate a more sensitive adverse event signalling method's development

We are currently developing a more sensitive adverse event signalling method, which is described in more detail in Ivkovic *et al.* (2005). To illustrate our general approach we select "N" (nervous system) for drug class and cardiovascular system for SOC class. The algorithm traverses down the ATC tree and finds the drug Clozapine (N05AH02). It has found the drug having the most cardiovascular reactions. The individual reactions for this drug are illustrated in Figure 6, the blue gives the frequency of the reaction and the red indicates the critical term status of the particular reaction (10 if critical, 1 otherwise). The inclusion of critical terms provides a means of assessing the seriousness of reactions, which in turn can be used to give the level of warning indicated. We also group report date as already described in Section 2.5 and this is illustrated in Figure 7 showing the monthly reaction counts for this drug. Here the signalling algo-

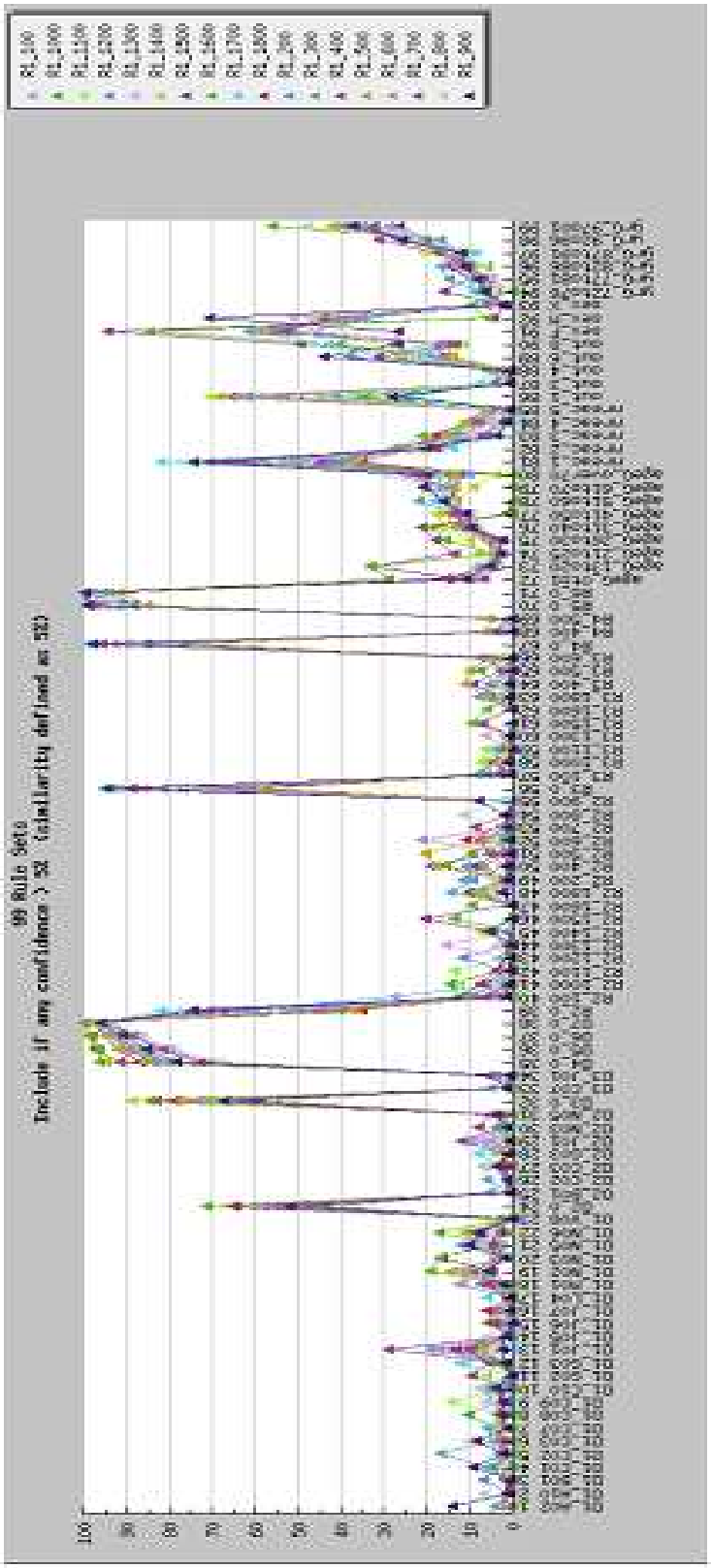

Figure 2: Graph Output Filtering - reaction characteristics (R1_SOC_code – SOC of first reaction field; D1_ATC_code – ATC class of first drug field).

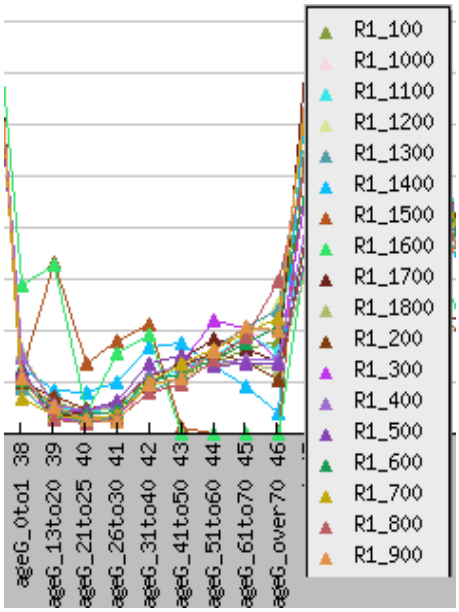

Figure 3: Age group characteristics - closer view.

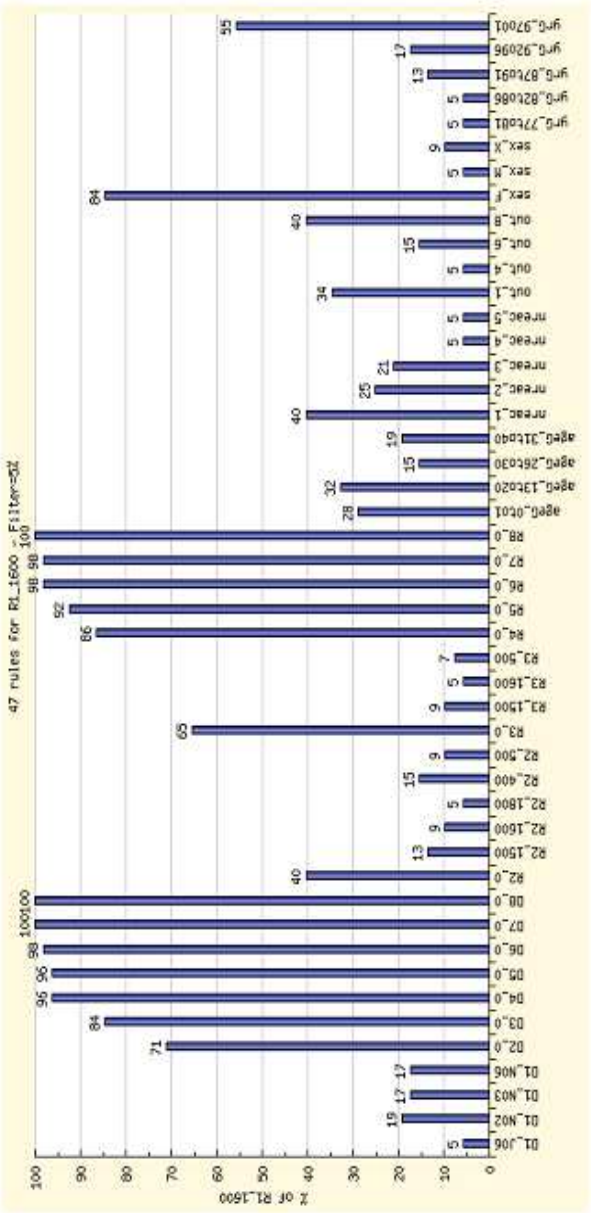

Figure 4: SOC 1600 Neonatal and infancy disorders.

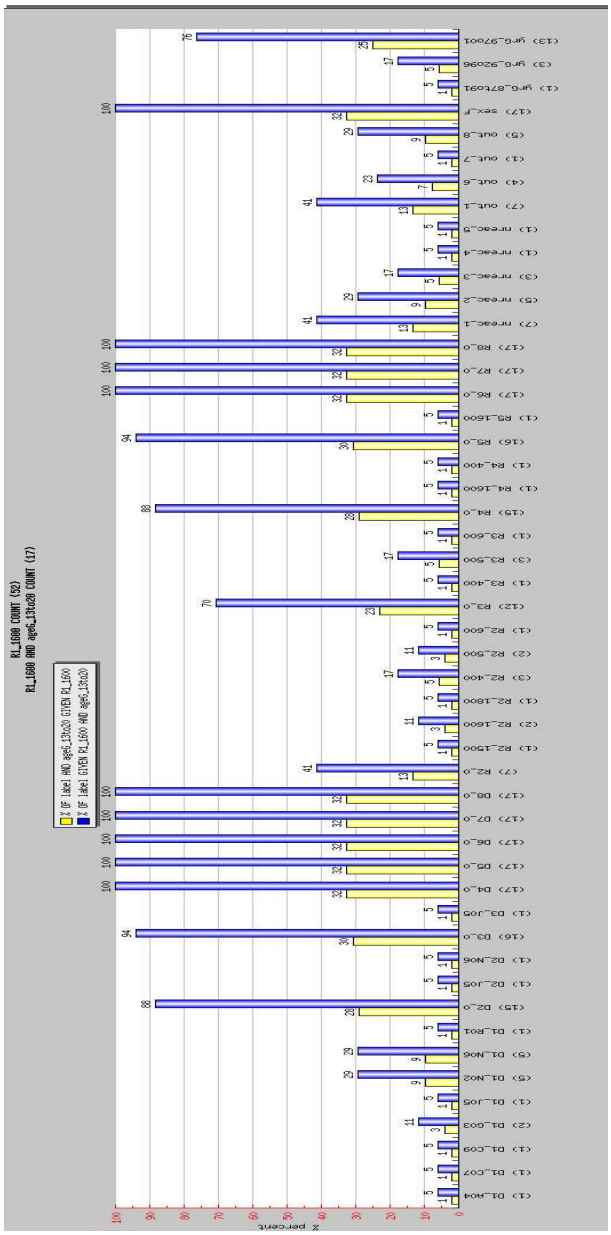

Figure 5: Teenagers with Neonatal disorders.

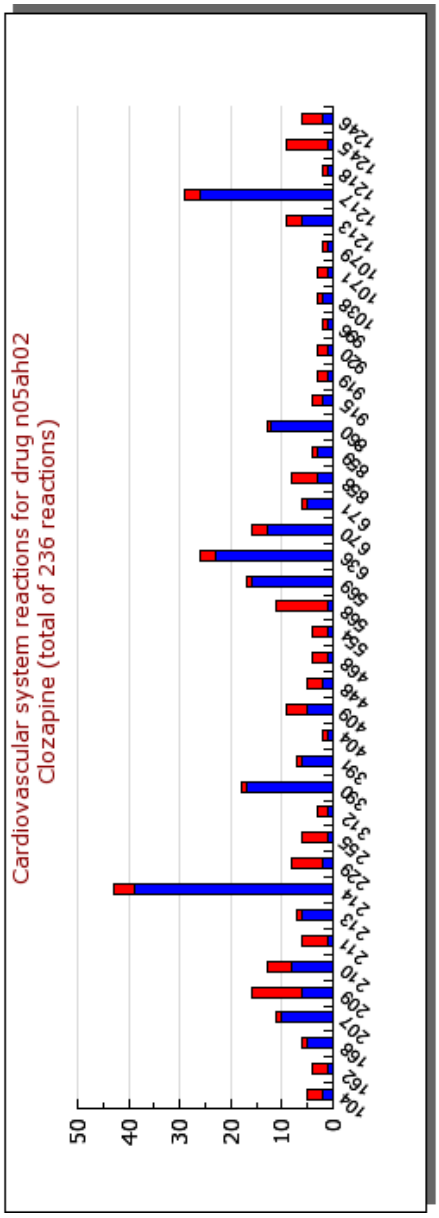

Figure 6: Reactions for Clozapine (N05AH02) for reaction class 1000 – Cardiovascular system.

rithm evaluated this drug and gives a report: "Null hypothesis: There are no ADR signals in this time span is REJECTED... ALERT...There is signal on 199908 (position 34)." This drug has produced a signal in August 1999. This algorithm has traversed all the way down the ATC tree and produced a warning with its time of occurrence.

4 Conclusions

The ATC ontology enables pooling of all data relating to a single ingredient, rather than having it split between different trade-names. The fact that there are several levels of granularity in this classification allows analysis at different levels. We have demonstrated that, using the ATC and SOC ontologies our association rule methods can explore even very rare associations in the ADRAC data. The tree structure of the ATC ontology was exploited to develop a adverse event signalling method that can 'drill down' the ATC tree to find individual drugs that have significantly more reactions in a given SOC. Then, using the critical term ontology, monthly reaction frequencies are examined to produce warnings when reac-

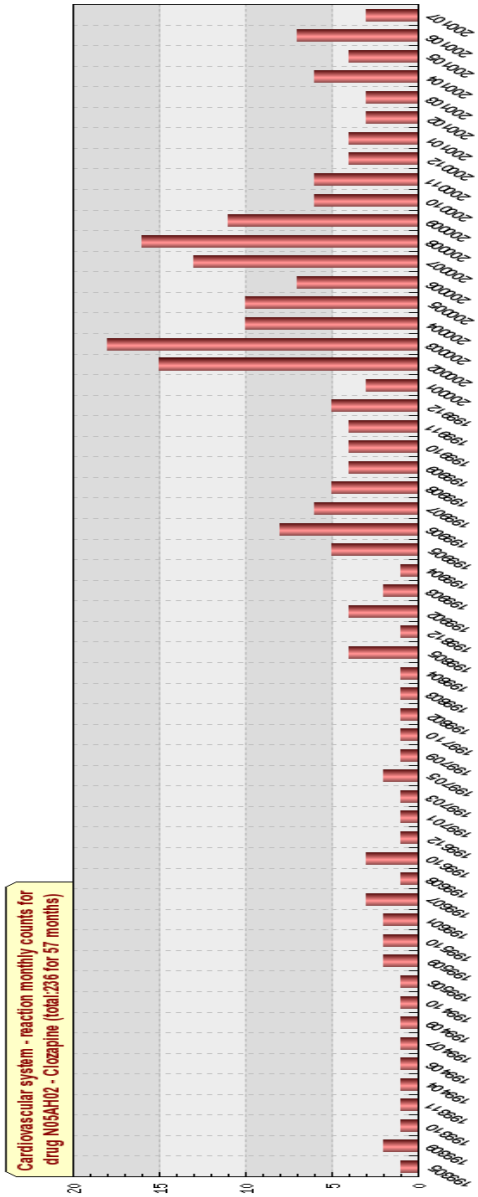

Figure 7: Timing of reactions Clozapine (N05AH02) for reaction class 1000 – Cardiovascular system.

tions rise to a significant level. Thus the application of ontologies to drug safety data enables a significant advance in adverse event signalling methodology.

Acknowledgements

The work reported in this paper has been supported by The Australian Research Council, the TGA (Therapeutic Goods Administration), the Ballarat Division of General Practice, the Medical Software Industry Association, the Therapeutic Guidelines Pty. Ltd. We are indebted to Patrick Purcell from the TGA for providing expert comments and Frank De Luca, in this school, for assistance with database issues. I would also like to thank Andrew Bate, of the World Health Organization, Uppsala Monitoring Centre (UMC), for making available for my research the World Health Organization Adverse Reaction Terminology (WHO-ART) Critical Term List WHO (2002), (World Health Organization 2002).

References

Agrawal, R., Imielinski, T. & Swami, A. (1993), Mining association rules between sets of items in large databases, *in* 'ACM SIGMOD International Conference on Management of Data', Vol. 22, ACM Press, Washington DC, USA, pp. 207–216.

Bates, D. W., Evans, R., Murff, H., Stetson, P. D. & Pizziferri, L. (2003), Detecting Adverse Events Using Information Technology, *J Am Med Inform Assoc* **10** (2) February, 115–128.

Chuang-Stein, C. (1998), Statistics for Safety Data. Detection of New Adverse Drug Reactions. 271–279, eds. M. D. B. Stephens J. C. C. Talbot P. A. Routledge, Macmillan Reference Ltd, London

Heeley, E., Riley, J., Layton, D., Wilton, L. V. & Shakir, S. A. (2001), Prescription-event monitoring and reporting of adverse drug reactions. Lancet **358** (9296) 1872–1873.

Ivkovic, Sasa, Saunders, G., Ghosh, R. & Yearwood, J. (2005), Using association and overlapping time window approach to detect drug reaction signals. International Coference on Computational Intelligence for Modelling Control And Automation – CIMCA'2005, November Vienna, Austria.

Ivkovic, Sasa, Ghosh, Ranadhir, Yearwood, John (2004), The Automated Identification and Description of Clusters for Categorical Data using Conditional Probabilities. International Coference on Computational Intelligence for Modelling Control And Automation – CIMCA'2004, July 214–223 Gold Coast, Australia.

Ivkovic, Sasa, Yearwood, John, Stranieri, Andrew (20042), Discovering Interesting Association Rules from Legal Databases. Information and Communication Technology Law **1** (8) 35–47.

Ivkovic, Sasa, Yearwood, John & Stranieri, Andrew (2003), Visualizing Association Rules for feedback within the legal system. Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003, June 214–223 Edinburgh, Scotland, UK.

Mamedov, M. & Saunders, G. (2004), A Fuzzy Derivative Approach to Classification of outcomes from the ADRAC database. International Trans Operational Research **11** (2) March 169–180.

Mamedov, M. A. & Saunders, G. W. (2002), An Analysis of Adverse Drug Reactions from the ADRAC Database. Part1: Cardiovascular group. University of Ballarat School of Information Technology and Mathematical Sciences Research Report, February **02/01** 1–48. Available at: http://www.ballarat.edu.au/ard/itms /publications/researchPapers/papers2002.shtml

Mamedov, M. A., Saunders, G. W. & Dekker, E.(2003), A Study of Drug-Reaction Relationships in Australian Drug Safety Data. Proceedings of the 2nd Australian Data Mining Workshop (ADM03). December 143–162 eds. S. J. Simoff and G. J. Williams and M. Hegland, Canberra, University of Technology Sydney.

Pinkston, V. & Swain, E. J. (1998), Management of Adverse Drug Reaction and Adverse Event Data through Collection, Storage and Retrieval. Detection of New Adverse Drug Reactions. 281–296 eds. M. D. B. Stephens, J. C. C. Talbot and P. A. Routledge, Macmillan Reference Ltd, London.

Pirmohamed, M., Breckenridge, A. M., Kitteringham, N. R. & Park,B. K. (1998), Adverse drug reactions. BMJ **316** (7140) April (1998) 1295–1298.

Saunders, Gary W. (2004), Draft anatomical therapeutic chemical ATC classification of Australian drug safety data. University of Ballarat School of Information Technology and Mathematical Sciences Research Report, September **04/22** (2004) 1–10. Available at: http://www.ballarat.edu.au/ard/itms/publications /researchPapers/Papers_2004/04-22.pdf

Saunders, G. W., Mamedov, M. & Ivkovic, Sasa (2005), Using anatomical therapeutic chemical (ATC) classification to stratify Australian drug safety data. 16th Australasian Workshop on Combinatorial Algorithms (AWOCA 2005), September Ballarat, Victoria, Australia.

Troutman, W. G. & Doherty, K. M. (2003), Comparison of voluntary adverse drug reaction reports and corresponding medical records. Am. J Health-Syst Pharm **60** (6) 572–575. ISSN: 1079-2082

van der Heijden, P. G. M., van Puijenbroek, E. P., van Buuren, S. & van der Hofstede, J. W. (2002), On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. Statist Med **21** 2027–2044.

van Puijenbroek, E. P., Diemount, W. L. & van Grootheest, K. (2003), Application of Quanitative Signal Detection in the Dutch Spontaneous Reporting System for Adverse Drug Reactions. Drug Saf **26** (5) 293–301.

World Health Organization (2002), WHO-ART Critical Term List, *Current Critical Term List*, World Health Organization, Uppsala Monitoring Centre (UMC), 2002.

World Health Organization (2002), The WHO Adverse Reaction Terminology, *Technical Report on Reaction Terminology*, 2002. url: http://www.umc-products.com/graphics/3149.pdf