Detect adverse drug reactions for drug Simvastatin

Yihui Li 
'Institute of Intelligent Information Processing 
Shandong Polytechnic University, China
UK Yihui.liu.2005@yahoo.co.uk

Uwe Aickelin
2Department of Computer Science, 
University of Nottingham,

II. FEATURE MATRIX AND FEATURE SELECTION

A. The Extraction of Feature Matrix

To detect the ADRs of drugs, first feature matrix is extracted from THIN database, which describes the medical events that patients occur before or after taking drugs. Then feature selection method of Student's t-test is performed to select the significant features from feature matrix containing thousands of medical events. Figure 1 shows the process to detect the ADRs using feature matrix. Feature matrix A describes the medical events for each patient during 60 days before they take drugs. Feature matrix B reflects the medical events during 60 days after patients take drugs. In order to reduce the effect of the small events, and save the computation time and space, we set 100 patients as a group. Matrix X and Y are feature matrix after patients are divided into groups.

Figure 1. The process to detect ADRs. Matrix A and B are feature matrix before patients take drugs or after patients take drugs. The time period of observation is set to 60 days. Matrix X and Y are feature matrix after patients are divided into groups. We set 100 patients as one group.

B. Medical Events and Readcodes

Medical events or symptoms are represented by medical codes or Readcodes. There are 103387 types of medical events in "Readcodes" database. The Read Codes used in general practice (GP), were invented and developed by Dr James Read in 1982. The NHS (National Health Service) has expanded the codes to cover all areas of clinical practice. The code is hierarchical from left to right or from level 1 to level 5. It means that it gives more detailed information from level 1 to level 5. Table 1 shows the medical symptoms based on Readcodes at level 3 and at level 5. 'Other soft tissue disorders' is general term using Readcodes at level 3. 'Foot pain', 'Heel pain', etc., give more details using Readcodes at level 5.
C. Feature Selection Based on Student’s t-test

Feature extraction and feature selection are widely used in biomedical data processing [11-18]. In our research we use Student’s t-test [19] feature selection method to detect the significant ADRs from thousands of medical events. Student's t-test is a kind of statistical hypothesis test based on a normal distribution, and is used to measure the difference between two kinds of samples.

Table 2 shows the top 30 detected results in ascending order of p value of Student's t-test, using Readcodes at level 1-5 and at level 1-3. The detected results are using p value less than 0.05, which represent the significant change after patients take the drug. Table 3 shows the results in descending order of the ratio of the number of patients after taking the drug to one before taking the drug. Table 4 shows potential ADRs related cancer for Simvastatin. The detected ADRs are based on our computerized method, further investigation is needed.

It is clear that our detected results are consistent with published side effects for statin drugs [21, 22]. Major ADRs of 'muscle and musculoskeletal' events for statin drugs are detected not only based on Readcodes at level 1-5, but also based on Readcodes at level 1-3.

IV. CONCLUSIONS

In this study we propose a novel method to successfully detect the ADRs using feature matrix and feature selection. A feature matrix, which characterizes the medical events before patients take drugs or after patients take drugs, is created from THIN database. The feature selection method of Student's t-test is used to detect the significant features from thousands of medical events. The significant ADRs, which are corresponding to significant features, are detected. Experiments are performed on the drug Simvastatin. Compared to other computerized method, our proposed method achieves good performance.

REFERENCES

[1] G. Severino, and M.D. Zompo, "Adverse dmg reactions: role of pharmacogenomics," Pharmacological Research, vol. 49, pp. 363-373, 2004.
[2] Y. Qian, X. Ye, W. Du, J. Ren, Y. Sun, H. Wang B. Luo, Q. Gao, M. Wu, and J. He, "A computerized system for signal detection in spontaneous reporting system of Shanghai China," Pharmacoeconomic and Drug Safety, vol. 18, pp. 154-158, 2009.
[3] K.S. Park, and D. Kwon, "The state of adverse event reporting and signal generation of dietary supplements in Korea," Regulatory Toxicology and Pharmacology, vol. 57, pp. 74-77, 2010.
[4] R. Kasliwal, L.V. Wilton, V. Cornelius, B. Aurich-Barrera, S.A.W. Shakir, "Safety profile of Rosuvastatin-results of a prescription-event monitoring study of 11680 patients," Drug Safety, vol. 30, pp. 157170, 2007.
[5] R.D. Mann, K. Kubota, G. Pearce, and L. Wilton, "Salmeterol: a study by prescription-event monitoring in a UK cohort of 15,407 patients," J Clin Epidemiol, vol. 49, pp. 247-250, 1996.
[6] R.H. Meyboom, M. Lindquist, A.C. Egberts, and I.R. Edwards, "Signal detection and follow-up in pharmacovigilance," Dmg Saf., vol. 25, pp. 459-465, 2002.
[7] A. Bate, M. Lindquist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner, and R.M. De Freitas, "A Bayesian neural network method for adverse drug reaction signal generation," Eur J Clin Pharmacol, vol. 54, pp.315-321, 1998.
[8] M. Hauben and A. Bate, "Decision support methods for the detection of adverse events in post-marketing data," Dmg Discovery Today, vol. 14, pp.343-357, 2009.
[9] Y. Koh, C.W. Yap, and S.C. Li, "A quantitative approach of using genetic algorithm in designing a probability scoring system of an adverse drug reaction assessment system," International Journal of Medical Informatics, vol. 77, pp. 421-430, 2008.
[10] C. Henegar, C. Bousquet, A.L. Lillo-Le, P. Degoulet, and M.C. Jaulet, "A knowledge based approach for automated signal generation in pharmacovigilance," Stud Health Technol Inform, vol. 107, pp. 626-630, 2004.
[11] Y. Liu, “Feature extraction and dimensionality reduction for mass spectrometry data,” Computers in Biology and Medicine, vol. 39, pp. 818-823, 2009.

[12] Y. Liu, "Detect key genes information in classification of microarray data," EURASIP Journal on Advances in Signal Processing, 2008. Available: doi:10.1155/2008/612397.

[13] Y. Liu and L. Bai, "Find significant gene information based on changing points of microarray data," IEEE Transactions on Biomedical Engineering, vol. 56, pp. 1108-1116, 2009.

[14] Y. Liu, M. Muftah, T. Das, L. Bai, K. Robson, and D. Auer, "Classification of MR Tumor Images Based on Gabor Wavelet Analysis," Journal of Medical and Biological Engineering, vol. 32, pp. 22-28, 2012.

[15] Y. Liu, "Dimensionality reduction and main component extraction of mass spectrometry cancer data," Knowledge-Based Systems, vol. 26, pp. 207-215, 2012.

[16] Y. Liu, "Wavelet feature extraction for high-dimensional microarray data," Neurocomputing, vol. 72, pp. 985-990, 2009.

[17] Y. Liu, W. Aubrey, K. Martin, A. Sparks, C. Lu, and R. King "The analysis of yeast cell morphology features in exponential and stationary Phase," Journal of Biological Systems, vol. 19, pp. 561-575, 2011.

[18] Y. Liu, "Prominent feature selection of microarray data," Progress in Natural Science, vol. 19, pp. 1365-1371, 2009.

[19] E. Kreyszig, Introductory Mathematical Statistics, John Wiley, New York, 1970.

[20] http://en.wikipedia.org/wiki/Simvastatin.

[21] http://www.statanswers.com/effects.htm

[22] http://www.drugs.com/sfx/simvastatin-side-effects.html

### TABLE II. THE TOP 20 ADRS FOR SIMVASTATIN BASED ON DESCENDING ORDER OF R1 VALUE

| Rank | Readcodes | Medical events | NB | NA | R1 | R2 |
|------|-----------|----------------|----|----|----|----|
| 1    | 11Z12.00  | Chronic kidney disease stage 3 | 185 | 1095 | 5.92 | 7.35 |
| 2    | M036000   | Cellulitis NOS | 98  | 503  | 5.13 | 3.37 |
| 3    | F4C0.00   | Acute conjunctivitis | 113 | 525  | 4.65 | 3.52 |
| 4    | N131.00   | Cervicalgia - pain in neck | 140 | 609  | 4.35 | 4.09 |
| 5    | 11060000  | Chest infection NOS | 284 | 1201 | 4.23 | 8.06 |
| 6    | N143.00   | Sciatica | 83  | 366  | 4.41 | 2.46 |
| 7    | F46.00    | Cataract | 40  | 312  | 7.80 | 2.09 |
| 8    | 1M10.00   | Knee pain | 198 | 762  | 3.85 | 5.11 |
| 9    | A53.11    | Shingles | 41  | 262  | 6.39 | 1.76 |
| 10   | C34.00    | Cough | 107 | 381  | 3.56 | 2.56 |
| 11   | A55.00    | Dysuria | 70  | 308  | 4.40 | 2.07 |
| 12   | N245.17   | Shoulder pain | 185 | 717  | 3.88 | 4.81 |
| 13   | F45.00    | Glaucoma | 20  | 148  | 7.40 | 0.99 |
| 14   | K190.00   | Urinary tract infection, site not specified | 128 | 607  | 4.74 | 4.07 |
| 15   | F501.00   | Infective otitis externa | 89  | 372  | 4.18 | 2.50 |
| 16   | 1D14.00   | C/O: a rash | 152 | 689  | 4.53 | 4.62 |
| 17   | N094K12   | Hnp pain | 96  | 461  | 4.80 | 3.09 |
| 18   | 1B32.11   | Ankle swelling symptom | 34  | 190  | 5.59 | 1.27 |
| 19   | 1L9.00    | Sore throat symptom | 97  | 410  | 4.23 | 2.75 |
| 20   | M03z000   | Cellulitis NOS | 98  | 503  | 5.13 | 3.37 |

Variable NB and NA represent the numbers of patients before or after they take drugs. Variable R1 represents the ratio of the numbers of patients after taking drugs to the numbers of patients before taking drugs. Variable R2 represents the ratio of the numbers of patients after taking drugs to the numbers of patients before taking drugs for having one particular medical event.

### TABLE III. THE TOP 20 ADRS FOR SIMVASTATIN BASED ON DESCENDING ORDER OF R1 VALUE

| Rank | Readcodes | Medical events | NB | NA | R1 | R2 |
|------|-----------|----------------|----|----|----|----|
| 1    | 1Z1E.00   | Chronic kidney disease stage 3 | 185 | 1095 | 5.92 | 7.35 |
| 2    | M038000   | Cellulitis NOS | 98  | 503  | 5.13 | 3.37 |
| 3    | C106.00   | Diabetes mellitus with neurological manifestation | 295 | 1054 | 2.98 | 7.07 |
| 4    | Eu320000  | [X]Mild depressive episode | 140 | 609  | 4.35 | 4.09 |
| 5    | SK171000  | Other leukaemia | 140 | 609  | 4.35 | 4.09 |
| 6    | N143.00   | Sciatica | 83  | 366  | 4.41 | 2.46 |
| 7    | F46.00    | Cataract | 40  | 312  | 7.80 | 2.09 |
| 8    | 1M10.00   | Knee pain | 198 | 762  | 3.85 | 5.11 |
| 9    | A53.11    | Shingles | 41  | 262  | 6.39 | 1.76 |
| 10   | C34.00    | Cough | 107 | 381  | 3.56 | 2.56 |
| 11   | A55.00    | Dysuria | 70  | 308  | 4.40 | 2.07 |
| 12   | N245.17   | Shoulder pain | 185 | 717  | 3.88 | 4.81 |
| 13   | F45.00    | Glaucoma | 20  | 148  | 7.40 | 0.99 |
| 14   | K190.00   | Urinary tract infection, site not specified | 128 | 607  | 4.74 | 4.07 |
| 15   | F501.00   | Infective otitis externa | 89  | 372  | 4.18 | 2.50 |
| 16   | 1D14.00   | C/O: a rash | 152 | 689  | 4.53 | 4.62 |
| 17   | N094K12   | Hnp pain | 96  | 461  | 4.80 | 3.09 |
| 18   | 1B32.11   | Ankle swelling symptom | 34  | 190  | 5.59 | 1.27 |
| 19   | 1L9.00    | Sore throat symptom | 97  | 410  | 4.23 | 2.75 |
| 20   | M03z000   | Cellulitis NOS | 98  | 503  | 5.13 | 3.37 |

Variable NB and NA represent the numbers of patients before or after they take drugs for having one particular medical event.

Variable R1 represents the ratio of the numbers of patients after taking drugs to the numbers of patients before taking drugs. Variable R2 represents the ratio of the numbers of patients after taking drugs to the number of the whole population.
TABLE W. THE POTENTIAL ADRS RELATED TO CANCER FOR SIMVASTATIN BASED ON P VALUE OF STUDENT’S T-TEST.

| Rank | Readcodes | Medical events                                      | NB   | NA  | R1   | R2   |
|------|-----------|-----------------------------------------------------|------|-----|------|------|
| 1    | B93..00   | Other malignant neoplasm of skin                    | 46   | 241 | 5.24 | 1.02 |
| 2    | B34..00   | Malignant neoplasm of female breast                 | 8    | 72  | 9.00 | 0.48 |
| 3    | B76..00   | Benign neoplasm of skin                             | 75   | 240 | 3.20 | 1.61 |
| 4    | BB5..00   | [M]Adenomas and adenocarcinomas                     | 7    | 69  | 9.86 | 0.46 |
| 5    | BB2..00   | [M]Papillary and squamous cell neoplasms            | 7    | 66  | 9.43 | 0.44 |
| 6    | B46..00   | Malignant neoplasm of prostate                      | 23   | 105 | 4.57 | 0.70 |
| 7    | B22..00   | Malignant neoplasm of trachea, bronchus and lung    | 5    | 47  | 9.40 | 0.52 |
| 8    | B32..00   | Malignant melanoma of skin                          | 1    | 22  | 22.00| 0.15 |
| 9    | B70..00   | Suspected malignancy                                | 4    | 33  | 8.25 | 0.22 |
| 10   | BB3..00   | [M]Basal cell neoplasms                             | 4    | 30  | 7.50 | 0.20 |
| 11   | B8..00    | Carcinoma in situ                                   | 12   | 41  | 3.42 | 0.28 |
| 12   | B57..00   | Secondary malig neop of respiratory and digestive systems | 0   | 15  | 15.00| 0.10 |
| 13   | B83..00   | Carcinoma in situ of breast and genitourinary system| 4    | 29  | 7.25 | 0.19 |
| 14   | B13..00   | Malignant neoplasm of colon                         | 7    | 30  | 4.29 | 0.20 |
| 15   | B17..00   | Malignant neoplasm of pancreas                      | 0    | 13  | 13.00| 0.08 |
| 16   | B62..00   | Other malignant neoplasm of lymphoid and histiocytic tissue | 2   | 17  | 8.50 | 0.11 |
| 17   | B11..00   | Malignant neoplasm of stomach                       | 1    | 12  | 12.00| 0.08 |
| 18   | B49..00   | Malignant neoplasm of urinary bladder               | 3    | 19  | 6.33 | 0.13 |
| 19   | B14..00   | Malignant neoplasm of rectum, rectosigmoid junction and anus | 5   | 24  | 4.80 | 0.16 |
| 20   | B63..00   | Multiple myeloma and immunoproliferative neoplasms   | 0    | 8   | 8.00 | 0.05 |
| 21   | B59..00   | Malignant neoplasm of unspecified site              | 1    | 11  | 11.00| 0.07 |
| 22   | B10..00   | Malignant neoplasm of oesophagus                    | 3    | 12  | 4.00 | 0.08 |
| 23   | BB4..00   | [M]Transitional cell papillomas and carcinomas      | 6    | 17  | 2.83 | 0.11 |
| 24   | B1..00    | Malignant neop of ill-defined sites digestive tract/peritoneum | 0   | 5   | 5.00 | 0.03 |
| 25   | B58..00   | Secondary malignant neoplasm of other specified sites | 2   | 9   | 4.50 | 0.06 |
| 26   | B44..00   | Malignant neoplasm of ovary and other uterine adenxa | 0   | 4   | 4.00 | 0.03 |
| 27   | B81..00   | Carcinoma in situ of respiratory system             | 0    | 4   | 4.00 | 0.03 |
| 28   | B56..00   | Secondary and unspecified malignant neoplasm of lymph | 0   | 4   | 4.00 | 0.03 |
| 29   | B11..00   | Benign neoplasm of other parts of digestive system   | 10   | 23  | 2.30 | 0.15 |
| 30   | BBQ..00   | [M]Germ cell neoplasms                              | 0    | 3   | 3.00 | 0.02 |
Detected adverse drug reactions for drug Simvastatin

Liu, Y. & Aickelin, U. (2012). Detect adverse drug reactions for drug Simvastatin. 2012 FOURTH INTERNATIONAL CONFERENCE ON MULTIMEDIA INFORMATION NETWORKING AND SECURITY (MINES 2012), pp.246-249. IEEE. https://doi.org/10.1109/MINES.2012.101.