

Detection of Long Term Adverse Drug Reactions in Electronic Health Care Data

D1: Protocol for the experiments

Dr. Martijn J. Schuemie

Dr. Gianluca Trifirò

Prof. Dr. Miriam Sturkenboom

Erasmus University Medical Center, Rotterdam, The Netherlands

Document History

Name	Date	Version	Description
Martijn Schuemie, Gianluca Trifirò	29-07-2011	1.0	First version

Table of Contents

Document History	2
1. Introduction	3
1.1. Definition of long-term adverse drug reactions	4
2. Real data evaluation set.....	4
2.1. Data sources.....	4
2.2. Event selection.....	4
2.3. Selection of true positives drug-long term event associations.....	5
2.4. Selection of true negative drug-long term event associations.....	6
3. Simulated data evaluation set	6
3.1. Risk function.....	7
3.2. Injection in OSIM II datasets	7
4. Protocol.....	7
4.1. Division into training and test sets	7
4.2. Baseline methods.....	8
4.3. Performance metrics.....	8
References	8
Appendix A: Long term ADR ground truth set	10
Appendix B: Long term ADR simulated positives.....	12
Appendix C: Random sample for training	13

1. Introduction

Spontaneous Reporting Systems (SRSs) depend on human recognition of potential cause and effect, and are therefore most likely to pick up acute effects where exposure to a drug is followed immediately or with only a short delay by the adverse reaction. Currently, observational pharmacovigilance initiatives such as OMOP and Sentinel in the US and EU-ADR in Europe also employ statistical methods that are most likely to detect these acute effects. These methods often use risk windows extending no more than 30 to 60 days beyond the original exposure, and methods that do include all time after exposure make no distinction between short and long exposure duration, effectively diluting the time at risk and thereby reducing the signal to noise ratio.

However, a substantial group of Adverse Drug Reactions (ADRs) is likely to have a much longer time between first exposure and adverse event occurrence, and the risk could actually be increasing with cumulative exposure. Well known examples are corticosteroids and fractures, dopamine agonists and cardiac valve fibrosis, and chronic effect of levodopa such as freezing during movement or dyskinesia. The availability of large amounts of longitudinal observational health care data with long follow-up now puts us in a unique position where we would be able to pick up these types of ADRs, but appropriate signal detection screening methods are lacking.

The objective of this research is to develop methods that can detect long term ADRs. These type of ADRs in general follow the stages depicted in **figure 1**. For instance, for the relationship between pergolide and cardiac valve fibrosis¹ there is a minimum amount of exposure required before histopathological manifestations of fibrosis start to occur. After that it can take a long time before first clinical symptoms start to manifest, and even longer before the disease is diagnosed. Although in practice the moment of biological disease onset is impossible to determine, in signal detection this concept can play an essential role: we can expect that with enough exposure, patient will have a higher risk of being diagnosed with the disease, even if the exposure has stopped before diagnosis. By taking the cumulative drug exposure into account, the risk window can be extended far beyond the immediately preceding drug exposure.

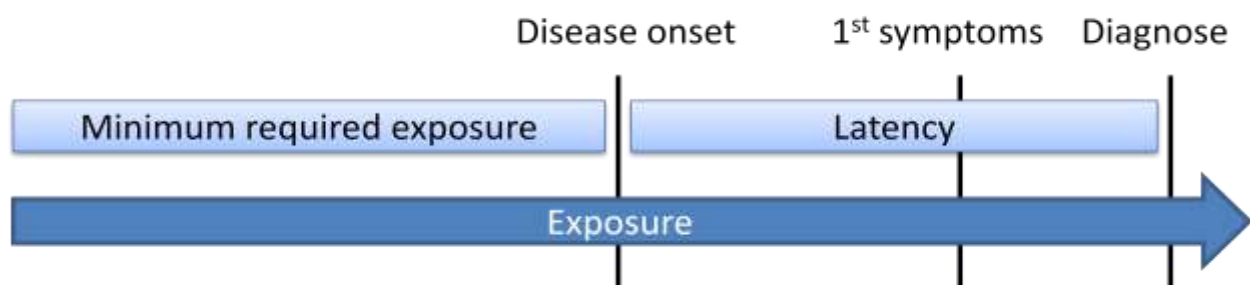


Figure 1. Long term ADR development.

1.1. Definition of long-term adverse drug reactions

Based on the timing of onset, adverse drug reactions may be classified as acute (immediate or very early onset after the start of drug exposure, i.e. few hours, days), delayed (days, weeks of drug exposure are required before the onset of adverse event) and chronic/long term (months of drug exposure are required)².

An official definition for long term adverse reaction does not currently exist. After reviewing carefully the literature, we considered as long term all the adverse reactions for which there is evidence that *at least 90 days* of drug exposure are necessary to be induced.

Examples of long term adverse drug reactions include:

- Adaptive changes (e.g. development of drug tolerance and physical dependence)
- Carcinogenesis
- Chronic organ damage due to deposit (e.g. pulmonary fibrosis)

Generally, these types of adverse reactions are rare and depend on the cumulative drug exposure.

The first step in developing methods to detect long term ADRs is to develop test data that can be used to evaluate these methods. Two types of data will be used: real data and simulated data.

2. Real data evaluation set

The real data will comprise the electronic health record data available in the OMOP Research Lab, and the ground truth will be known long term ADRs, as defined by experts. The ground truth set, shown in **appendix A**, was constructed using the following data sources and procedures:

2.1. Data sources

To explore if a drug was associated or not with specific long term events, leading drug-related information sources have been explored. In particular, Summary of Product Characteristics (SPC), as reported in Daily Med (<http://dailymed.nlm.nih.gov>), PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) for medical literature search and MICROMEDEX[®] 1.0 from Thomson (<http://www.thomsonhc.com/hcs>) have been taken into account, in addition to expert opinion.

2.2. Event selection

Long term adverse events have been selected among those events under study in the EU-ADR project (www.eu-adr-project.com)³ and OMOP (www.omop.fnih.org). As regard to events investigated in EU-ADR, a selection has been conducted among the 13 events for which a coding algorithm for the event data extraction from databases has been already generated (acute myocardial infarction, acute liver injury, acute renal failure, anaphylactic shock, bullous eruption, cardiac valve fibrosis, rhabdomyolysis, upper gastrointestinal bleeding, neutropenia, aplastic anemia, acute pancreatitis, hip fracture,

progressive multifocal leuko-encephalopathy). Some events may be drug-induced as both acute and delayed/chronic effect. We selected those events which are more likely to be induced through long term drug exposure, even if occasionally some of those events may also be induced as a result of drug exposure for less than 90 days.

Two medical experts identified the events that could be induced by long term drug exposure, and for which at least one scientific publication (e.g. case report, observational studies, and randomized clinical trials) indicated that the onset of the event was attributable to long term (i.e. at least 90 days) drug exposure.

2.3. Selection of true positives drug-long term event associations

A stepwise procedure was followed, similarly to what has been done for the creation of other validation sets in the EU-ADR project.

First, among the drugs that are captured in the GE database in the OMOP research lab, we identified all the drugs (N=154) for which at least 10,000 users were treated for ≥ 1 year during the study follow-up. From this restricted list of drugs, we selected the drugs for both the sets of true positive and true negative drug-event associations.

Second, from this restricted list of drugs, we identified for each event at least 5 drugs which are known to be associated with the event and for which there was evidence that at least 90 days of exposure were required to induce the corresponding adverse reaction. For some events, there were not enough drugs known to be associated; in this situation, we selected other drugs that were captured in OMOP database that did not fulfill the criterion of at least 10,000 users with ≥ 1 year of exposure.

A drug was considered to be associated with long term event if:

- a) the corresponding adverse drug reaction was described in Micromedex (section adverse drug reactions)
- b) the corresponding adverse drug reaction was described in the Summary of Product Characteristics
- c) there was at least one publication (e.g. case report, observational studies, and randomized clinical trials) describing the occurrence of the corresponding adverse drug as a result of at least 90 days of drug exposure.

With respect the literature search in PubMed, a tool that was developed in EU-ADR has been used to facilitate both the identification of the drugs with the largest number of publications concerning each potential long term event and the retrieval of the related publications. Subsequently, the publications were reviewed independently by two medically trained researchers (PC and GD). In case of disagreement a third expert (GT) (medical doctor, clinical pharmacologist and epidemiologist with expertise on drug safety) arbitrated.

Whenever possible, for each event drugs belonging to different classes (ATC 3rd level, pharmacological subgroup) were selected.

2.4. Selection of true negative drug-long term event associations

To identify true negative drug-event associations, a similar stepwise procedure was adopted. Five drugs per event were selected as described below:

Among the above-mentioned restricted list of OMOP drugs (N=154), for each event 5 drugs belonging to different classes (ATC 3rd level, pharmacological subgroup) were first randomly selected. Drug-event associations were considered as true negative associations if there was no mention of the corresponding adverse drug reaction in the SPC, Micromedex and scientific literature. In case of presence of mention of the drug-event association in at least one of these data sources, the drug-event associations was discarded and another drug for that specific event was randomly selected from the OMOP restricted list of drugs and investigated as possible true negative association.

For matter of efficiency, the scientific literature was initially checked through a tool developed in EU-ADR: drugs with no citations were further evaluated through manual inspection of PubMed. In addition, the website www.farmacovigilanza.org was also consulted. This is an Italian website about drug safety which is managed by the Italian Society of Pharmacology.

All the assessments have been independently conducted by two medically trained researchers (PC and GD). A third expert (GT) arbitrated in case of disagreement.

3. Simulated data evaluation set

Simulated data has the advantage that the underlying truth is known without doubt. The simulated data sets will be generated using OSIM II. In order to make OSIM II suitable for generating long-term ADRs according the definition used here, the accumulative exposure injection procedure was redefined as follows:

- The probability of a person getting an adverse event injected is a function of the accumulated exposure that that person has. The function is described below.
- The accumulation starts counting after a patient has min_onset of exposure to the drug, and stops counting after max_offset has been achieved.
- For all persons that are determined to get an adverse event using the rules mentioned above, the timing of the event is also determined by the risk function: events are more likely to occur later in time, when a person has more accumulated exposure.
- Each inserted event is delayed by a random number of days between min_delay and max_delay. This is done to simulate for instance the delay in time between onset of cancer (a single cell mutation), development of symptoms and diagnosis of the cancer.

3.1. Risk function

The risk function used is inspired by the risk function of lung cancer and accumulated exposure to asbestor⁴ as shown in **figure 2**. This function is approximated by a power function:

$$\text{Relative risk} = \text{Exposure-years}^{\text{Alpha}}$$

In this example, the alpha is approximately 0.4.

The power function is first used to determine whether a person receives an adverse event or not. The timing of the event is then randomly picked, where the random number generator function of SQL, which uses a uniform probability function, is transformed to the appropriate distribution over time using the integral of the inverse of the power function.

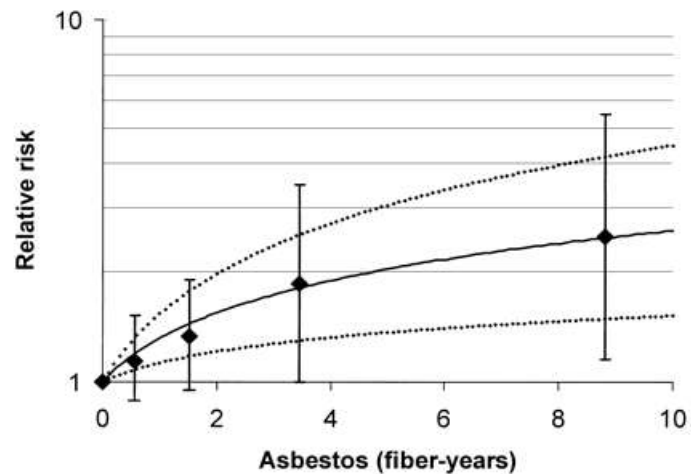


Figure 2. Accumulation of exposure and relative risk of lung cancer

3.2. Injection in OSIM II datasets

The ground truth set described earlier is used as the basis for the insertion of signals. The different parameters settings are selected to give a wide diversity of signal types. The list of injected signals and their parameters are described in **appendix B**.

4. Protocol

The methods developed in this project will be tested both on the simulated and real test data. The real data will be used as the golden standard.

4.1. Division into training and test sets

To overcome overfitting methods to the evaluation data, part of the data should be kept separately and only used for a final test, while the other data will be used to test improvements during development.

For training data, the following data will be used:

- The simulated data
- A small random sample of the ground truth set. For each outcome type, one positive and one negative is selected. The random sample is shown in **appendix C**.

The remainder of the ground truth set will be used for testing purposes. Testing will be done twice during this project: after development of the methods incorporating cumulative exposure, and after the development of methods for comparison group selection.

4.2. Baseline methods

The performance of the methods developed in this project will be compared to:

- Random baseline
- Incidence Rate Ratio with Mantel-Haenszel correction for age and gender, with varying carry-over periods (0, 30, 180, 365 days)
- The best performing method in the upcoming OMOP experiment

4.3. Performance metrics

The main performance metric will be the area under the ROC curve. For each method, the ROC curve itself will also be generated.

References

1. Schade, R., Andersohn, F., Suissa, S., Haverkamp, W. & Garbe, E. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med* **356**, 29-38 (2007).
2. Rawlins, M. & Thompson, J. in Textbook of Adverse Drug Reactions (ed. D. Davies) (Oxford University Press, Oxford 1977).
3. Trifiro, G., Pariente, A., Coloma, P.M. et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* **18**, 1176-1184 (2009).
4. Gustavsson, P., Nyberg, F., Pershagen, G., Scheele, P., Jakobsson, R. & Plato, N. Low-dose exposure to asbestos and lung cancer: dose-response relations and interaction with smoking in a population-based case-referent study in Stockholm, Sweden. *Am J Epidemiol* **155**, 1016-1022 (2002).
5. Bhattacharyya, S., Schapira, A.H., Mikhailidis, D.P. & Davar, J. Drug-induced fibrotic valvular heart disease. *Lancet* **374**, 577-585 (2009).
6. Lycette, J.L., Luoh, S.W., Beer, T.M. & Deloughery, T.G. Acute bilateral pulmonary emboli occurring while on adjuvant aromatase inhibitor therapy with anastrozole: Case report and review of the literature. *Breast Cancer Res Treat* **99**, 249-255 (2006).
7. Tchaikovski, S.N. & Rosing, J. Mechanisms of estrogen-induced venous thromboembolism. *Thromb Res* **126**, 5-11 (2010).
8. Drospirenone: high risk of venous thrombosis. *Prescrire Int* **20**, 43-45 (2011).
9. Adomaityte, J., Farooq, M. & Qayyum, R. Effect of raloxifene therapy on venous thromboembolism in postmenopausal women. A meta-analysis. *Thromb Haemost* **99**, 338-342 (2008).
10. Tripp, A.C. Nonfatal pulmonary embolus associated with clozapine treatment: a case series. *Gen Hosp Psychiatry* **33**, 85 e85-86 (2011).

11. Keene, D.L., Legare, C., Taylor, E., Gallivan, J., Cawthorn, G.M. & Vu, D. Monoclonal antibodies and progressive multifocal leukoencephalopathy. *Can J Neurol Sci* **38**, 565-571 (2011).
12. Bresalier, R.S., Sandler, R.S., Quan, H. et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* **352**, 1092-1102 (2005).
13. Nissen, S.E. & Wolski, K. Rosiglitazone Revisited: An Updated Meta-analysis of Risk for Myocardial Infarction and Cardiovascular Mortality. *Arch Intern Med* (2010).
14. Shapiro, S., Slone, D., Rosenberg, L., Kaufman, D.W., Stolley, P.D. & Miettinen, O.S. Oral-contraceptive use in relation to myocardial infarction. *Lancet* **1**, 743-747 (1979).
15. Orti, G., Mira, Y. & Vaya, A. Acute myocardial infarction associated with Yasmin oral contraceptive. *Clin Appl Thromb Hemost* **13**, 336-337 (2007).
16. Nakagawa, T., Yasuno, M., Tanahashi, H. et al. A case of acute myocardial infarction. Intracoronary thrombosis in two major coronary arteries due to hormone therapy. *Angiology* **45**, 333-338 (1994).
17. Kim, Y.W., Park, B.S., Ryu, C.H. et al. Allopurinol-induced aplastic anemia in a patient with chronic kidney disease. *Clin Nephrol* **71**, 203-206 (2009).
18. Handoko, K.B., Souverein, P.C., van Staa, T.P. et al. Risk of aplastic anemia in patients using antiepileptic drugs. *Epilepsia* **47**, 1232-1236 (2006).
19. Focosi, D., Kast, R.E., Benedetti, E., Papineschi, F., Galimberti, S. & Petrini, M. Phenobarbital-associated bone marrow aplasia: a case report and review of the literature. *Acta Haematol* **119**, 18-21 (2008).
20. Balamuthusamy, S. & Arora, R. Hematologic adverse effects of clopidogrel. *Am J Ther* **14**, 106-112 (2007).
21. Symeonidis, A., Kouraklis-Symeonidis, A., Seimeni, U. et al. Ticlopidine-induced aplastic anemia: two new case reports, review, and meta-analysis of 55 additional cases. *Am J Hematol* **71**, 24-32 (2002).

Appendix A: Long term ADR ground truth set

Table 1. List of drugs associated with long term adverse events under study in EU-ADR and/or OMOP (true positive associations)

Long Term Event	Drug	ATC**
CARDIAC VALVE FIBROSIS*	Phentermine ⁵	A08AA01
	Pergolide ^{1, 5}	N04BC02
	Cabergoline ^{1, 5}	N04BC06/G02CB03
VENOUS THROMBOSIS	Anastrozole ⁶	L02BG03
	Levonorgestrel ⁷	G03AA07
	Drospirenone ⁸	G03AA12
	Raloxifene ⁹	G03XC01
	Clozapine ¹⁰	N05AH02
PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY (PML)*	Alemtuzumab ¹¹	L01XC04
	Rituximab ¹¹	L01XC02
	Fludarabine ¹¹	L01BB05
	Natalizumab ¹¹	L04AA23
MYOCARDIAL INFARCTION	Rofecoxib ¹²	M01AH02
	Rosiglitazone ¹³	A10BG02
	Levonorgestrel ¹⁴	G03AA07
	Drospirenone ^{15 14}	G03AA12
	Medroxyprogesterone ¹⁶	G03AC06
PANCYTOPENIA	Allopurinol ¹⁷	M04AA01
	Valproic acid ¹⁸	N03AG01
	Phenobarbital ¹⁹	N03AA02
	Clopidogrel ²⁰	B01AC04
	Ticlopidine ^{21***}	B01AC05

As bold type, the drugs reported in the OMOP drug restricted list. All the other drugs are captured by OMOP database but do not have 10,000 users with ≥ 1 year of exposure

* For PML and cardiac valve fibrosis, it was not possible to identify additional true positive drug-event associations, concerning a drug that can be captured in OMOP database

** Some drugs have multiple ATC, with some of those corresponding to topical formulations. Make sure that these drugs are not considered for the true positive associations

*** In some case reports, aplastic anemia occurred earlier than three months after the start of the therapy with ticlopidine

Table 2. List of drugs that are NOT associated with potential long term adverse events under study in EU- ADR and/or OMOP (true negative associations)

Long Term Event	Drug	ATC
CARDIAC VALVE FIBROSIS	Fluconazole	J02AC01
	Nifedipine	C08CA05
	Allopurinol	M04AA01
	Celecoxib	M01AH01
	Zolpidem	N05CF02
VENOUS THROMBOSIS	Gemfibrozil	C10AB04
	Prednisone	H02AB15
	Tadalafil	G04BE08
	Esomeprazole	A02BC05
	Ciprofloxacin	S01AX13
PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY (PML)	Raloxifene	G03XC01
	Paroxetine	N06AB05
	Norgestimate	G03AA11
	Irbesartan	C09CA04
	Glyburide	A10BB01
MYOCARDIAL INFARCTION	Valacyclovir	J05AB11
	Topiramate	N03AX11
	Triamcinolone	A01AC01
	Rabeprazole	A02BC04
	Amoxicillin	J01CA04
PANCYTOPENIA	Risedronate	M05BA07
	Irbesartan	C09CA04
	Doxazosin	C02CA04
	Desloratadine	R06AX27
	Venlafaxine	N06AX16

Appendix B: Long term ADR simulated positives

Table 3. Accumulative signals injected in the OSIM II dataset.

drug concept id	drug name	condition name (concept id)	Relative risk	alpha	min_delay	max_delay
1189754	rofecoxib	Acute myocardial infarction (35205180)	1.41	0.4	0	7
1500211	Medroxyprogesterone		4	0.2	0	7
1512674	drospirenone		3	0.2	0	7
1547504	rosiglitazone		1.44	0.4	0	7
1589505	Levonorgestrel		1.26	0.4	0	7
734275	Phenobarbital	Aplastic anaemia (35104101)	2	0.4	0	30
745466	Valproate		1.5	0.4	0	30
1167322	Allopurinol		1.2	0.4	0	30
1302398	Ticlopidine		4	0.2	0	30
1322184	clopidogrel		3	0.2	0	30
800878	Clozapine	Embolism venous (37622456)	4	0.2	0	180
1348265	anastrozole		1.2	0.4	0	180
1512674	drospirenone		2	0.4	0	180
1513103	Raloxifene		3	0.2	0	180
1589505	Levonorgestrel		1.5	0.4	0	180
732309	Pergolide	Mitral valve disease (35205091)	2	0.4	0	90
735340	Phentermine		1.5	0.4	0	90
1558471	cabergoline		3	0.4	0	90
735843	natalizumab	Progressive multifocal leuko-encephalopathy (36110978)	3	0.2	30	365
1312706	alemtuzumab		1.2	0.4	30	365
1314273	rituximab		1.5	0.4	30	365
1395557	fludarabine		2	0.4	30	365

Appendix C: Random sample for training

Table 4. Random sample of positive signals for training.

Long Term Event	Drug	ATC
CARDIAC VALVE FIBROSIS	Cabergoline	N04BC06/G02CB03
VENOUS THROMBOSIS	Clozapine	G03AA12
PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY (PML)	Rituximab	L01XC02
MYOCARDIAL INFARCTION	Drospirenone	G03AA12
PANCYTOPENIA	Allopurinol	M04AA01

Table 5. Random sample of negative signals for training.

Long Term Event	Drug	ATC
CARDIAC VALVE FIBROSIS	Fluconazole	J02AC01
VENOUS THROMBOSIS	Tadalafil	G04BE08
PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY (PML)	Glyburide	A10BB01
MYOCARDIAL INFARCTION	Amoxicillin	J01CA04
PANCYTOPENIA	Desloratadine	R06AX27