Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
22(1) 97–112

DOI: 10.1177/0962280211403599
smm.sagepub.com
SAGE

Performance of a semi-automated approach for risk estimation using a common data model for longitudinal healthcare databases

Hoa Van Le,[1,2] Kathleen J Beach,[2] Gregory Powell,[2] Ed Pattishall,[2] Patrick Ryan[1,2] and Robertino M Mera[2]

Abstract

Different structures and coding schemes may limit rapid evaluation of a large pool of potential drug safety signals using multiple longitudinal healthcare databases. To overcome this restriction, a semi-automated approach utilising common data model (CDM) and robust pharmacoepidemiologic methods was developed; however, its performance needed to be evaluated. Twenty-three established drug-safety associations from publications were reproduced in a healthcare claims database and four of these were also repeated in electronic health records. Concordance and discrepancy of pairwise estimates were assessed between the results derived from the publication and results from this approach. For all 27 pairs, an observed agreement between the published results and the results from the semi-automated approach was greater than 85% and Kappa coefficient was 0.61, 95% CI: 0.19–1.00. Ln(IRR) differed by less than 50% for 13/27 pairs, and the IRR varied less than 2-fold for 19/27 pairs. Reproducibility based on the intra-class correlation coefficient was 0.54. Most covariates (>90%) in the publications were available for inclusion in the models. Once the study populations and inclusion/exclusion criteria were obtained from the literature, the analysis was able to be completed in 2–8 h. The semi-automated methodology using a CDM produced consistent risk estimates compared to the published findings for most selected drug-outcome associations, regardless of original study designs, databases, medications and outcomes. Further assessment of this approach is useful to understand its roles, strengths and limitations in rapidly evaluating safety signals.

Keywords

common data model, drug safety, observational database, pharmacovigilance, propensity score, safety signal, risk estimation

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC, USA
[2]GlaxoSmithKline, Research Triangle Park, NC, USA

Corresponding author:
Hoa Van Le, Department of Epidemiology, UNC Gillings School of Global Public Health, 2101 McGavran-Greenberg Hall, CB #7435, Chapel Hill, NC 27599-7435, USA
Email: hoa@unc.edu

1 Introduction

Drug safety signals arise from many sources, including spontaneous adverse event reports, published case reports or case series and clinical trial data. Historically, US drug safety surveillance has been a passive process, relying on voluntary reporting to the FDA Adverse Event Reporting System and mandatory reporting by industry, with well-recognised deficiencies.[1–4] Recently, the Institute of Medicine recommended that new methods and approaches of safety surveillance should be implemented, including data mining in an active safety surveillance system.[5–8] Sources of information may include longitudinal electronic health records (EHRs), and patient information derived from health insurance claims.[9] The Observational Medical Outcomes Partnership, an initiative to test and standardise analytical methods, and the FDA Sentinel Initiatives are working towards these objectives.[10]

An active surveillance system may generate hundreds of potential drug safety signals, which emphasises the need for a rapid approach to evaluate those signals that incorporate robust pharmacoepidemiologic methods.[10] Conducting risk analysis in multiple disparate data sources simultaneously would be a powerful addition towards the interpretation of a signal. However, databases have different structures and use different coding schemes; therefore, adjustments for these and other differences are necessary to successfully compare results across databases.[10,11] One approach would be to normalise the data sources into a common data model (CDM) to enable comparable questions to be asked of disparate databases.[12] In addition, an automated process for risk estimation might enable timely evaluation of numerous safety signals. The objective of this study was to compare the risk estimates generated by a semi-automated method utilising a CDM to the published pharmacoepidemiologic results.

2 Methods

2.1 Common data model

A CDM developed by GlaxoSmithKline was applied to two databases: PharMetrics (PM) and General Electric Centricity (GE). PM is a US administrative health claims database containing data for over 31 million patients with an average of 24 months of coverage and contains all reimbursed claims for each enrollee. Drug utilisation is extracted from dispensed medications and medical conditions captured *via* diagnosis codes represented in the databases as National Drug Code and International Classification of Diseases, 9th Revision (ICD-9), respectively. GE is a US EHR providing health information for approximately 8.9 million patients. Drug utilisation is extracted from two sources: (1) prescriptions written by the provider and (2) medication history lists. Medical conditions are captured from a problem list of diagnoses, symptoms and other components of medical history. Drug and medical event codes are represented as Generic Product Identifier and ICD-9, respectively.

The databases were transformed into a common framework that enables standardised analyses across sources followed by integration of analytical methods for risk estimation. Three primary categories of data were extracted from the database: patient, drug and medical conditions. The CDM used biomedical ontologies to normalise reference vocabularies to drugs and medical conditions.[12,13] During data transformation, ICD-9 codes were mapped to the MedDRA ontology[14] and medications (at the brand name level) were mapped to an enhanced SNOMED drug ontology.[15] Drug eras represented exposure time a given person uses a given drug concept. Once normalised, drugs were represented by SNOMED drug concepts and could be aggregated at product name, generic name or higher level drug classes. Similarly, medical conditions were created

after mapping ICD-9 diagnosis codes to MedDRA and were aggregated to preferred terms and then to high-level group terms.

2.2 Semi-automated methodology

Using the retrospective cohort design, the propensity score (PS)[16] was used to make exposure cohorts more comparable, as described elsewhere.[17,18] A list of medical conditions and medications occurring in a specified period prior to the drug exposure was automatically generated. A univariate odds ratio (OR) depicted the association between a particular drug or condition and the exposure. Covariates were suggested by the process based on an OR estimate above a desired threshold, although any covariate with an $OR > 0.01$ is available. Conditions and drugs were finally chosen taking into account biological rationale, confounding by indication and detection bias using the available literature and clinical knowledge.[18,19] Logistic regression was used to compute the PS and utilised to match with replacement on a 1:1 ratio within each stratum.[18] Using the matched data set, a separate model was built where the dependent variable was the outcome. For this outcome model, covariates were automatically generated based on association between the outcome and every condition and medication for the groups to be compared. Selection of outcome model covariates was similar to the above, with consideration of the number of events per covariate.[20,21] Adjusted incidence rate ratios (IRR) were computed using multivariate Poisson regression.

2.3 Selection of observational drug safety publications

Observational studies were found by searching PubMed/Medline, the Cochrane Database of Systematic Reviews, Google Scholar and peer-reviewed journals for either retrospective cohort or case–control studies using observational/administrative data sources in the US, UK or Canada to evaluate drug safety questions reported from 2000 to 2008. Selection was guided by several criteria, including English language, sufficient details of study design for reproduction, availability of database variables in the CDM and designated important medical events. A wide spectrum of populations, exposures and outcomes was sought for inclusion. Studies reporting results inconsistent with a body of evidence supporting an association were rejected. Studies with complex design algorithms, unique populations, requirements for procedure or laboratory values were considered if an appropriate alternative could be constructed.[22–41]

2.4 Reproduction of publications

For each drug-condition association, similar inclusion/exclusion criteria, target and comparator drugs, study start and stop dates, time at risk, outcomes and covariates of the respective publication were considered while constructing study designs (Table 1 and additional details available upon request). A retrospective cohort design with PS matching with replacement within each stratum was used for all comparisons.[18] Modifications to the study designs were permitted when necessary. For example, if PS-balanced populations differed in medically important ways to the study population, e.g. basic demographics, co-morbidities or co-medications, or too few exposures or outcomes, alternatives (i.e. creating a comparator for a non-user cohort, restricting the population to mimic the original database, substitution of medical diagnoses for laboratory values, extension of the exposure period to increase the number of exposed patients or increase the

Table 1. Details of the publications included in comparison to the semi-automated approach

Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	Data source	Study design	Exposure period	Age	Gender	Implementation differences between the approach and publication
1	Haerian et al.[22]	Hyperglycaemia	Ciprofloxacin Gatifloxacin	VA hospital	Case–control	1995–2002	18 and older	M/F	
2	Nordstrom et al.[23]	Pneumonia	Oseltamivir Non-oseltamivir	United Healthcare	Retrospective cohort	1/31/1999 to 3/31/2002	All ages	M/F	Anti-pyretic users instead of non-users
3	McAFee et al.[24]	Renal dysfunction	Rosuvastatin Other statins	Ingenix	Retrospective cohort	09/01/2003 to 02/29/2004	18 and older	M/F	Extended: 09/01/2003 to 05/13/2005 (the day of first warning from FDA) to have enough outcome
4	Wang et al.[25]	Hypertension	Celecoxib Non-selective NSAID	GE	Retrospective cohort	01/01/1999 to 06/30/2004	18 and older	M/F	
4EHR			Celecoxib Non-selective NSAID						
5	Gill et al.[26]	Ischaemic strokes	Atypical anti-psychotics Typical anti-psychotics	Ontario administrative healthcare databases	Retrospective cohort	4/4/1997 to 12/31/2003	65 and older	M/F	Not excluding non-oral anti-psychotics
6	L'Allier et al.[27]	Atrial fibrillation	Calcium channel blockers ACE inhibitors	US Claims, otherwise not specified	Retrospective cohort	01/01/1995 to 06/30/1999	18 and older	M/F	
7	Seeger et al.[28]	Tendon rupture	Fluoroquinolone Non-fluoroquinolone	Ingenix	Case–control	1/1/1997-6/30/2001	18 and older	M/F	
8	Cheng et al.[29]	Acute pancreatitis	ACE inhibitors Warfarin	Administrative Healthcare databases Ontario	Retrospective cohort	1/1/1994 to 3/31/2000	66 and older	M/F	1/1/1994 to 3/31/2008, 65 years and older
9	Suissa et al.[30]	Interstitial lung	Leflunomide Methotrexate	PharMetrics	Nested case–control	09/01/1998 to 12/31/2003	18 and older	M/F	Unable to exclude prior exposures to injectable drugs

(continued)

Table 1. Continued

Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	Data source	Study design	Exposure period	Age	Gender	Implementation differences between the approach and publication
9EHR			Leflunomide Methotrexate						09/01/1995 to 12/31/2007
10	Ray et al.[31]	Hip fracture	Statins Other lipid-lowering agents	Tenn Medicaid	Retrospective cohort	1/1/89 to 12/31/98	50 and older	M/F	1/1/1989 to 12/31/2003
11	Cadarette et al.[32]	Non-vertebral fractures	Calcitonin Alendronate	Medicare	Retrospective cohort	04/01/2000 to 12/31/2005	65 and older	M/F	
11EHR			Calcitonin Alendronate						
12	Schade et al.[33]	Cardiac valve regurgitation	Pergolide Cabergoline	GPRD	Retrospective cohort	1/1/1988 to 08/31/2005	40-80	M/F	1/1/1988 to 08/31/2005
13	Johannes et al.[34]	Allergic reactions	Levofloxacin Moxifloxacin	Ingenix	Retrospective cohort	7/1/2000 to 6/30/2004	All ages	M/F	
14	Abraham et al.[35]	Myocardial infarction	Rofecoxib Naproxen	VA Medicare	Retrospective cohort	1/1/2000 to 12/31/2002	65 and older	M/F	
15	Smitten et al.[36]	Herpes zoster	Oral Steroids traditional DMARDs	PharMetrics	Retrospective cohort	1/1/1998 to 12/31/2002	18 and older	M/F	1/1/1998 to 12/31/2005
15EHR			Oral steroids Traditional DMARDs						1/1/1995 to 12/31/2007
16	McAFee et al.[24]	Rhabdomyolysis	Rosuvastatin Other statins	Ingenix	Retrospective cohort	09/01/2003 to 02/29/2004	18 and older	M/F	09/01/2003 to 05/13/2005
17	Duncan et al.[37]	Diabetes-level hyperglycaemia	Olanzapine Risperidone	Veterans Integrated Service Network	Retrospective cohort	10/01/1998 to 06/30/2003	18 and older	M/F	
18	Cole et al.[38]	Venous thromboembolism	Norgestimate patch Norgestimate OC	UnitedHealthcare	Nested case–control	4/1/2002 to 12/31/2004	15–44	F	

(continued)

Table 1. Continued

Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	Data source	Study design	Exposure period	Age	Gender	Implementation differences between the approach and publication
19	Patterson et al.[39]	Peptic ulcer and bleeding	Naproxen Celecoxib	UnitedHealthcare	Case–control	1/1/1999 to 12/31/2003	18 and older	M/F	
20	van Staa et al.[40]	Neutropenia	Antibacterials Betablockers	GPRD	Case–control	1/1/1987 to 12/31/1999	3 and older	M/F	1/1/1998 to 12/31/2004
21	Haerian et al.[22]	Hypoglycaemia	Gatifloxacin Ciprofloxacin	VA hospital	Case–control	1995–2002	all ages	M	1/1/1999 to 12/31/2003
22	Meropol et al.[41]	Pseudomembranous colitis	Ciprofloxacin Doxycycline	UnitedHealthcare	Retrospective cohort	1/1/1999 to 06/30/2001	18 and older	M/F	
23			Ciprofloxacin Amoxicillin	UnitedHealthcare	Retrospective cohort	1/1/1999 to 06/30/2001	18 and older	M/F	

number of outcomes) were considered for implementation. Selected publications were reproduced in claims and, if originally performed in either PM or GE, they were repeated in both databases.

2.5 Components of assessment of performance

Although publications which reported IRR and 95% CI were preferred, the proportional hazards ratio (HR), risk ratio (RR) or OR were alternatively substituted for comparison. To evaluate the concordance and the magnitude of discrepancy, the following pairwise analyses were performed: (1) the statistical difference between the IRRs was based on a standard normal z-test as follows:[42–44]

$$z = \frac{\ln(\mathrm{IRR}_{\mathrm{Publication}}) - \ln\left(\mathrm{IRR}_{\mathrm{Approach}}\right)}{\left\{\mathrm{var}[\ln(\mathrm{IRR}_{\mathrm{Publication}})] + \mathrm{var}\left[\ln\left(\mathrm{IRR}_{\mathrm{Approach}}\right)\right]\right\}^{1/2}}$$

An absolute z score > 1.96 suggested a statistical difference within each pair at 0.05 significance level; (2) concordance defined as point estimates on the same side of the null was calculated by observed agreement, Kappa coefficient and Spearman correlation coefficient;[42,45,46] (3) the pairwise ratio of IRR at least double or less than half;[42] and (4) the pairwise ln(IRR) at least 50% larger or smaller.[42] To assess reproducibility intra-class correlation coefficient ICC $= \sigma_B^2/\left(\sigma_B^2 + \sigma_W^2\right)$, where σ_B^2 – variance between different studies, σ_W^2 – variance within specific study pairs and $\sigma_B^2 + \sigma_W^2$ – total variance was applied.[47] To estimate precision, the pairwise confidence limit ratio (CLR = upper limit/lower limit)[48] was calculated. Publications not reporting confidence intervals were excluded from any statistical comparison requiring CI.

3 Results

3.1 Overall assessment of performance

Twenty-three drug-event associations from the 21 publications[22–41] were repeated using the claims database and 4 of the 23 were also duplicated in the EHR database for a total 27 pairs (identified as Topics, see Table 1). The study designs included 15 retrospective cohort, 5 case–control and 2 nested case–control. A variety of databases (Canadian, UK and US, including data from Medicaid (1), Medicare (1) or the VA system (4)) and ages, including paediatric patients (4) or ages $\geq$65 (4) were also represented in the publications. The medications evaluated included a breadth of drugs from different therapeutic classes, such as anti-psychotics, cardiac agents, antibiotics and anti-inflammatory agents. The medical conditions evaluated included acute and chronic conditions, as well as diseases treated with short- or long-term exposure to medications.

Study design elements of the publications were modified when necessary by: creation of an alternative to a non-user cohort (Topic 2), restriction of the population to males to mimic the VA database (Topic 21), substitution of medical diagnoses for laboratory values (Topics 1, 17 and 21), and extension of the exposure period to increase the number of exposed patients or increase the number of outcomes (Topics 3, 8, 9, 9EHR, 10, 15, 15EHR, 16, 20 and 21). The detailed set of all modifications is listed in Table 1. The number of patients in the cohorts, approximate person time, IRR, 95% CI, concordance measures and magnitudes of discrepancy are presented in Tables 2 and 3.

For all 27 pairs, an observed agreement between the published results and the results from the semi-automated approach was greater than 85% and the Kappa coefficient was 0.61, 95% CI: 0.19–1.00. The IRR varied less than 2-fold for 19/27 pairs, and ln(IRR) differed by less than 50% for 13/27 pairs. Spearman correlation coefficient and ICC were 0.64 and 0.35, respectively.

Table 2. Details of risk estimates obtained from literature studies and the semi-automated approach

				Publications				Approach					
Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	# Patients	# Outcomes	Person time (years)	IRR/HR[a]/ OR[b]/ RR[c] (95% CI)	# Patients in unmatched cohort	# Outcomes in unmatched cohort	~Person time (years)[d]	IRR (95% CI)	% of covariates included	P for Z
1	Haerian et al.[22]	Hyperglycaemia	Ciprofloxacin	100	8	N/A	0.26 (0.10–0.67)[b]	4592	824	7825	0.81 (0.70–0.94)	91	0.02
			Gatifloxacin	144	38	N/A	1.00	2794	362	2793	1.00		
2	Nordstrom et al.[23]	Pneumonia	Oseltamivir	11 632	149	N/A	0.72 (0.60–0.86)[a]	660	10	500	0.78 (0.23–2.62)	77	0.9
			Non-oseltamivir	60 427	1575	N/A	1.00	392	10	391	1.00		
3	McAFee et al.[24]	Renal dysfunction	Rosuvastatin	11 249	12		0.9 (0.47–1.73)[a]	2278	8	2667	0.75 (0.24–2.36)	92	0.78
			Other statins	37 282	42	33 333	1.00	95 958	415	103 750	1.00		
4	Wang et al.[25]	Hypertension	Celecoxib	17 148	222	4227	1.01 (0.86–1.19)[a]	23 134	1155	32 083	1.07 (0.97–1.17)	89	0.54
			Non-selective NSAID	34 296	446	8611	1.00	390 344	9018	267 596	1.00		
4EHR			Celecoxib					26 174	577	30 691	0.93 (0.82–1.05)	78	0.79
			Non-selective NSAID					155 017	2211	109 455	1.00		
5	Gill et al.[26]	Ischaemic strokes	Atypical anti-psychotics	14 865	284	11 137	1.01 (0.81–1.26)	2671	82	7455	0.94 (0.46–1.93)	95	0.85
			Typical anti-psychotics	17 845	227	10 179	1.00	940	11	940	1.00		
6	L'Allier et al.[27]	Atrial fibrillation	Calcium channel blockers	5591	N/A	N/A	≈1.09 (≈0.9–≈1.3)[a]	509	21	1810	1.16 (0.58–2.34)	100	0.87
			ACE inhibitors	12 608	N/A	N/A	1.00	341	17	3864	1.00		
7	Seeger et al.[28]	Tendon rupture	Fluoroquinolone	N/A	49	N/A	1.3 (0.90–1.80)[b] for 18–59 years old	9935	5	8333	1.29 (0.37–4.47)	100	0.99
			Non-fluoroquinolone	N/A	751	N/A	1.1 (0.50–2.30)[b] for 60+ years old	27 865	13	26 000	1.00		
8	Cheng et al.[29]	Acute pancreatitis	ACE inhibitors	174 824	231	256 190	1.35 (0.94–1.93)	195 590	1137	222 941	1.05 (0.90–1.22)	100	0.21
			Warfarin	40 057	34	44 841	1.00	62 689	310	63 265	1.00		

(continued)

Table 2. Continued

				Publications				Approach					
Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	# Patients	# Outcomes	Person time (years)	IRR/HR[a]/ OR[b]/ RR[c] (95% CI)	# Patients in unmatched cohort	# Outcomes in unmatched cohort	~Person time (years)[d]	IRR (95% CI)	% of covariates included	P for Z
9	Suissa et al.[30]	Interstitial lung	Leflunomide	N/A	16	N/A	1.36 (0.80–2.60)	280	8	189	2.96 (0.81–10.78)	100	0.28
			Methotrexate	N/A	41	N/A	1.00	3138	58	4056	1.00		
9EHR			Leflunomide					647	4	64	20.24 (1.41–290.41)	90	0.96
			Methotrexate					6307	40	12 903	1.00		
10	Ray et al.[31]	Hip fracture	Statins	12 506	49	33 189	1.4 (0.83–2.43)[a]	43 287	70	9722	1.50 (0.49–4.59)	100	0.91
			Other lipid-lowering agents	4798	17	8990	1.00	4333	5	4167	1.00		
11	Cadarette et al.[32]	Non-vertebral fractures	Calcitonin	8372	309	19 649	1.40 (1.20–1.63)[a]	4946	113	2229	2.09 (1.56–2.78)	93	0.02
			Alendronate	21 007	448	7667	1.00	26 176	453	18 642	1.00		
11EHR			Calcitonin					5609	33	4521	1.77 (1.04–3.03)[d]	100	0.44
			Alendronate					27 009	97	23 659	1.00		
12	Schade et al.[33]	Cardiac valve regurgitation	Pergolide	N/A	6	N/A	1.45 (0.47–4.55)	499	80	492	1.49 (0.98–2.24)	90	0.97
			Cabergoline	N/A	6	N/A	1.00	952	75	686	1.00		
13	Johannes et al.[34]	Allergic reactions	Levofloxacin	290 365	216	11 130	1.5 (NA)[c]	342 863	1157	361 563	0.72 (0.62–0.84)	100	N/A
			Moxifloxacin	252 579	121	9681	1.00	90 320	401	91 136	1.00		
14	Abraham et al.[35]	Myocardial infarction	Rofecoxib	N/A	74	6022	1.76 (1.34–2.68)	2763	52	4262	1.46 (0.89–2.40)	94	0.11
			Naproxen	N/A	240	34 437	1.00	2850	53	6310	1.00		
15	Smitten et al.[36]	Herpes zoster	Oral steroids	N/A	166	N/A	1.83 (1.50–2.23)[b]	14 036	268	7953	1.91 (1.51–2.42)	100	0.79
			Traditional DMARDs	N/A	306	N/A	1.00	7674	136	7727	1.00		
15EHR			Oral steroids					4822	65	2968	1.98 (1.31–2.99)	100	0.74
			Traditional DMARDs					4809	47	4234	1.00		
16		Rhabdomyolysis	Rosuvastatin	11 249	1	10 169	1.98 (0.18–21.90)	34 297	7	35 000	2.13 (0.62–7.30)	92	0.96

(continued)

Table 2. Continued

				Publications				Approach					
Topic no.	Source	Outcomes evaluated	Study drug *vs.* comparator	# Patients	# Outcomes	Person time (years)	IRR/HR[a]/ OR[b]/ RR[c] (95% CI)	# Patients in unmatched cohort	# Outcomes in unmatched cohort	~Person time (years)[d]	IRR (95% CI)	% of covariates included	P for Z
	McAFee et al.[24]												
			Other statins	37 282	2	33 333	1.00	642 012	126	1 260 000	1.00		
17	Duncan et al.[37]	Diabetes-level hyperglycaemia	Olanzapine	2013	41	826 940	2.14 (1.31–3.49)[b]	16 210	647	12 966	1.04 (0.93–1.16)	100	0.005
			Risperidone	2426	24	883 064	1.00	13 061	625	13 048	1.00		
18	Cole et al.[38]	Venous thromboembolism	Norgestimate patch	98 790	20	49 048	2.2 (1.30–3.80)	29 074	77	22 647	1.92 (1.33–2.76)	90	0.68
			Norgestimate OC	256 981	37	202 344	1.00	49 183	95	52 778	1.00		
19	Patterson et al.[39]	Peptic ulcer and bleeding	Naproxen	N/A	239	N/A	2.22 (1.72–2.86)[b]	197 721	112	31 111	3.13 (1.93–5.07)	92	0.22
			Celecoxib	N/A	152	N/A	1.00	65 122	75	62 500	1.00		
20	van Staa et al.[40]	Neutropenia	Antibacterials	N/A	603	N/A	2.70 (2.30–3.10)[b]	9 522 686	3927	1 033 421	2.34 (1.63–3.36)	95	0.47
			Betablockers	N/A	153	N/A	1.00	352 956	582	363 750	1.00		
21	Haerian et al.[22]	Hypoglycaemia	Gatifloxacin	100	6	N/A	5.00 (1.70–14.70)[b]	191 466	218	72 667	2.29 (0.87–6.01)	91	0.29
			Ciprofloxacin	144	21	N/A	1.00	39 783	52	40 000	1.00		
22	Meropol et al.[41]	Pseudomembranous colitis	Ciprofloxacin	63 261		2407	18.20 (NA)	36 047	20	33 333	5.91 (1.65–21.22)	100	N/A
			Doxycycline	46 334	30	3750	1.00	27 741	3	30 000	1.00		
23			Ciprofloxacin	63 261		2407	20.22 (NA)	35 749	19	31 667	1.57 (0.77–3.19)	100	N/A
			Amoxicillin	268 130	87	12 380	1.00	118 566	20	50 000	1.00		

[a]Hazards ratio; [b]Odds ratio; [c]Risk ratio; [d]Approximate person time in years.

Table 3. Measures of assessment of performance of semi-automated approach

Description	Using claims database ($n = 23$)	Using both claims and EHR databases ($n = 8$)
Statistical |z| score ≤ 1.96	17/20 (85.0%)	7/8 (87.5%)
Observed agreement[a]	21/23 (91.3%)	7/8 (87.5%)
Kappa coefficient, 95% CI (substantial: 0.61–0.80)[45]	0.71 (0.31–1.00)	N/A
Spearman correlation coefficient (moderate: 0.35–0.50; strong: >0.5)[52]	0.70	0.48
Intra-class correlation coefficient (fair-to-good reproducibility: 0.4–0.75; poor: <0.4)[53]	0.54	0.02
IRR varied less than 2-fold	16/23 (69.6%)	6/8 (75%)
Ln(IRR) differed less than 50%	11/23 (47.8%)	3/8 (37.5%)
CLR of the approach vs. publications (median)	5.04 vs. 2.03	2.03 vs. 1.44

[a]Point estimates in the same side of the null.

3.2 Publications versus approach using the healthcare claims database

The comparison of risk estimates and 95% CI between the semi-automated approach and previous literature report for each topic are shown in Figure 1. Results showed good observed agreement, a substantial Kappa and strong Spearman correlation coefficients; CI overlapped for 95% of the 20 pairs with CI (Figure 1); and the difference in IRR of 85% was not statistically significant by z-test. The variance between and within pairs (residual) were 0.21 and 0.26, respectively. When two pairs with extreme point estimates (Topics 22 and 23)[41] were excluded, the variance between and within pairs were 0.17 and 0.10, respectively, with ICC of 0.64. The ratio of CLR was greater than 1 for 60% (12/20) and 46% (6/13 pairs reporting both sample size and CI). A median of 92% (range 77–100%) of the covariates, identified by the publications as important to confounding were generated automatically for consideration.

3.3 Publications versus approach using the claims and EHR databases

Three publications used GE and one used PM. These studies were reproduced in their respective database of which the results are included in the above section. These four publications were then repeated in the alternative database available in the application (Figure 2). Eight pairs were examined using both claims and EHR (Topics 4, 4EHR; 9, 9EHR; 11, 11EHR; and 15, 15EHR). Good observed agreement and moderate Spearman correlation coefficient were seen (Table 3). Kappa coefficient determination was not appropriate to report due to the paradox of high observed agreement but low Kappa.[49] Confidence intervals overlapped for 100% of the pairs (Figure 2), and the difference of IRRs of 87% was not statistically significant by z-test. The variances between and within pairs were 0.01 and 0.5, respectively. Excluding the pair with an unstable point estimate (Topic 9), the variance between and within pairs was 0.06, respectively, with an ICC of 0.49. The ratio of CLR was greater than 1 for 75% of the pairs (6/8). A median of 89% (range 78–100%) of the covariates, identified by the authors of the manuscripts as important to confounding were generated automatically for consideration.

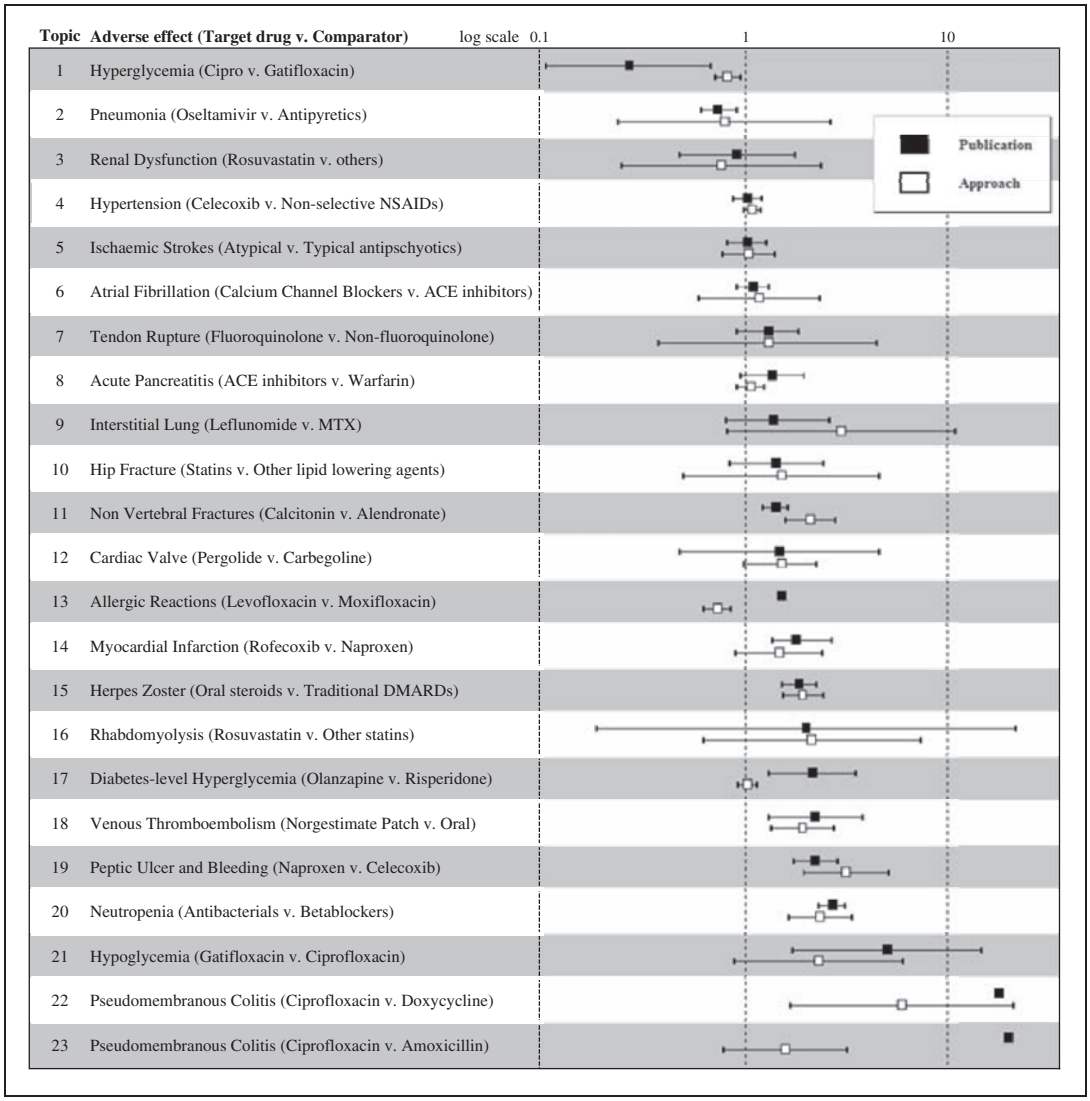

Figure 1. Comparison of risk estimates and 95% CI in publications vs. semi-automated approach using healthcare claims database.

4 Discussion

A semi-automated approach using a CDM provided results compatible with findings based on conventional epidemiology approaches for the evaluation of drug safety signals in observational data and has potential value for the assessment of potential drug-condition associations. A number of qualitative and quantitative analyses[43,44] were used to determine whether the output from this approach would be reliable and informative. Such analyses have been applied when comparing results of randomised controlled trials (RCT) and observational studies.[42–44] Concordance measures were consistent within pairs regardless of claims or EHR data source. If a more

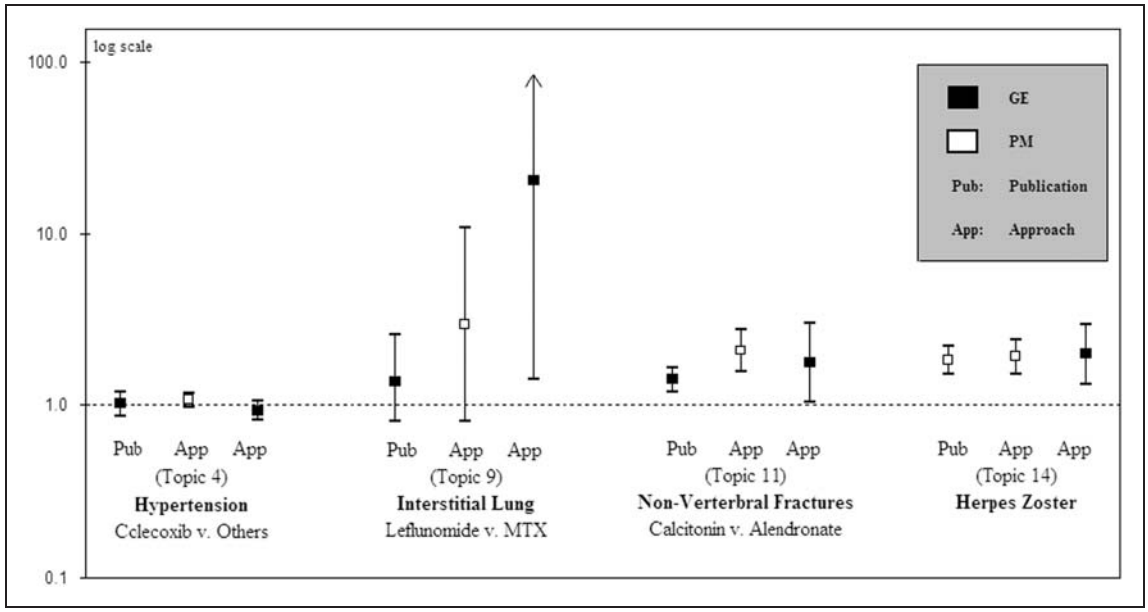

Figure 2. Comparison of risk estimates and 95% CI in publications based on either GE or PM vs. semi-automated approach using both databases.

conservative threshold for concordance was applied, i.e. greater than 1.2 or less than 0.83 (1/1.2), the results still produced an acceptable observed agreement of 85% and a Kappa coefficient of 0.57, 95% CI: 0.19–0.96 (data not shown).

Magnitudes of discrepancy differed by less than 2-fold for the majority of pairs, while half differed by less than 50%. Within pairs discordance and discrepancy of magnitude of effect measures might be due to differences in their databases, study design, analytic approaches or the assumptions built into the approach. Which parameter(s) specifically contributed to the discrepancies would only be speculative. Most CLRs were slightly wider than those presented in the publications; yet, they still reflected good precision and the ICC indicated satisfactory 'reproduction.' Since ICC coefficients are influenced by the variance of the sample/population[47] so that a 'low' ICC as obtained with the EHR database should be interpreted with caution. On the other hand, the ICC derived with the claim-based cases paired with the literature indicated satisfactory 'reproduction.'

Although the current approach incorporates methods designed to assess increased risk, it was important to examine protective or null associations as negative and null controls, respectively. For those publications with a risk less than 0.83, 2/2 of the duplicated studies were also less than 0.83; 4/5 agreed for those with a null effect (0.83–1.2). Despite the few examples with known risk less than 1.2, the results provide some assurance that the semi-automated approach adequately reflected a reduction in risk and would have provided reliable information for hypothesis strengthening. Few publications were reproduced in both claims and EHR because the commercially available EHR does not have sufficient healthcare information per the selection criteria described earlier. Relatively consistent results (ignoring ICC) suggest that there is value to analysing two disparate databases, when appropriate. A powerful strength of the approach relies on the automated generation of a comprehensive list of co-morbidities and co-medications which may not have

been pre-defined as covariates. The reduction of residual confounding based on another approach has been recently reported.[50]

The CDM, a potentially important component for standardising disparate longitudinal healthcare databases in the national sentinel initiative[10,12] (1) accommodates observational data elements including relevant to identifying drug exposures, condition occurrences and other clinical observations; (2) allows each datum to be standardised on a common vocabulary; and (3) prevents the use of protected health information except where necessary to protect the public health. More details of CDM are described elsewhere.[12,17] During this study, limitations of the current CDM were identified: (1) translation of medical conditions and drug exposures into a common vocabulary may not be applicable in all instances. Cross-mapping may be incomplete or result in unclassified or misclassified concepts. MedDRA allergic reactions contained ICD-9 codes for food and medication reactions which were not separable (Topic 13); (2) drugs were grouped irrespective of dose, combinations or formulations. Higher doses of naproxen were combined with low doses (Topic 19); (3) brand name drugs were grouped at the generic names; (4) medications, if missing days of supply, were imputed to 30 days which may not be appropriate for all situations; (5) creation of drug eras might result in miscalculation of real exposure time to a specific drug for patients in some situations. Despite the limitations identified, the results were close to those using traditional methods in the publication.

Together the automation and CDM facilitate efficient analyses, reduction in programming error, standardisation of variable definitions and utilisation of multiple healthcare databases. Such an approach could allow for the timely assessment of large numbers of potential signals. Our initial attempt was to explore the feasibility of such an approach using CDM and semi-automated methodology for the assessment of drug safety signals. If drug-condition associations found by conventional methodology and by this approach had been grossly different, then further examination was not warranted.

Automatic generation of covariates has been explored[50,51] and inclusion of additional variables into pre-defined list covariates for PS adjustment has been shown to reduce residual confounding in selected examples.[52] Our automation is accomplished by (1) generation of a list of co-morbidities and co-medications using information from the entire database relevant to the cohorts of interest; (2) automatic inclusion of all covariates meeting investigator's pre-specified thresholds for PS and outcome models is possible. However, for this study the authors used the option to review and select those satisfying medical rationale as confounders;[19] (3) pre-programmed calculation of person time at risk with retrospective cohort design of the longitudinal data; (4) PS balancing by matching with replacement within each stratum; (5) creation of IRR from multivariate Poisson regression; (6) a selection of trimming methods (untrimmed, one or both tailed trimming at different percentile levels of the common support region of the PS distribution) for PS matching within stratum.

Although RCTs might be considered to represent a gold standard for comparison,[43,52,53] observational data are commonly used when RCTs are not practical due to the large time and expense. Thus, the intent was to compare the methodology to situations in which it would most likely be used. It is encouraging that the semi-automated results evaluated in this exercise would have led, in many instances, to similar interpretations as found with published pharmacoepidemiology studies. However, it may be useful to perform comparisons of results to those found in RCTs in the future to increase the confidence of the results from observational data.

The selection of publications focused on important drug adverse effects previously studied in observational data and might be considered somewhat arbitrary. Once a potential study was identified using criteria described earlier, it was reviewed for feasibility of reproduction by two physician epidemiologists, a potential source of selection bias. In order to complete any risk estimation in observational data, it is necessary to have sufficient numbers of exposures and

outcomes, a well-defined outcome, the ability to select exposures/outcomes, etc. The decision process for conducting a risk assessment for a real pharmacoepidemiology study follows these same principles. Importantly, publications were not chosen with the pre-conceived notion that results of the approach would be similar.

One of the limitations of the semi-automated approach is that for an extremely rare outcome, the variance of the estimate may be larger than its mean, and negative binomial regression may be more appropriate than Poisson regression. More features for incident user study design will be available in future releases. In some instances, it was necessary for HR, OR or RR to serve as proxies for IRR based on the assumption that the outcomes were neither delayed nor frequent. Some patients were used more than once while using matching with replacement. The Wald confidence interval was calculated and corrected for the influence of sampling with replacement by variance inflation factor.[18] Further research on Generalised Estimating Equations to adjust for clustering has been carried out.[54] Target–comparator pairs were reversed to increase the number of publications with a positive association (Topics 6 and 19) or to examine a performance for a reduction of risk (Topic 1).

In conclusion, an approach utilising a CDM and semi-automated methodology reproduced relatively consistent results from published pharmacoepidemiologic studies utilising traditional approaches. The results provide more confidence in a semi-automated approach which may be valuable in evaluating safety signals from multiple disparate databases. Further evaluation of this approach for rapidly identifying potential drug-condition associations is useful to understand its roles, strengths and limitations.

Acknowledgements

Preparation of this article was funded in part by the Harry A. Guess-Merck Scholarship in Pharmacoepidemiology and a VEF fellowship to HVL.

References

1. Piazza-Hepp TD and Kennedy DL. Reporting of adverse events to MedWatch. *Am J Health Syst Pharm* 1995; **52**(13): 1436–1439.
2. Brewer T and Colditz GA. Postmarketing surveillance and adverse drug reactions: current perspectives and future needs. *JAMA* 1999; **281**(9): 824–829.
3. Wysowski DK and Swartz L. Adverse drug event surveillance and drug withdrawals in the United States, 1969-2002: the importance of reporting suspected reactions. *Arch Intern Med* 2005; **165**(12): 1363–1369.
4. Furberg CD, Levin AA, Gross PA, Shapiro RS and Strom BL. The FDA and drug safety. A proposal for sweeping changes. *Arch Intern Med* 2006; **166**: 1938–1942.
5. Chan KA and Hauben M. Signal detection in pharmacovigilance: empirical evaluation of data mining tools. *Pharmacoepidemiol Drug Saf* 2005; **14**(9): 597–599.
6. Platt R, Madre L, Reynolds R and Tilson H. Active drug safety surveillance: a tool to improve public health. *Pharmacoepidemiol Drug Saf* 2008; **17**(12): 1175–1182.
7. Mann RD, Wilton LV, Pearce GL, Mackay FJ and Dunn NR. Prescription-event monitoring (PEM) in 1996 - a method of non-interventional observational cohort pharmacovigilance. *Pharmacoepidemiol Drug Saf* 1997; **6**(S3): S5–S11.
8. Psaty BM and Korn D. Congress responds to the IOM drug safety report-in full. *JAMA* 2007; **298**(18): 2185–2187.
9. Food and Drug Administration Amendments Act of 2007, Public Law 110-85, § 905(a)(3)(B)(i)–(ii).
10. U.S. Food and Drug Administration. The Sentinel Initiative: National Strategy for Monitoring Medical Product Safety, http://www.fda.gov/oc/initiatives/advance/reports/report0508.html (2008, accessed 1 October 2009).
11. Bright AR and Nelson RC. Automated support for pharmacovigilance: a proposed system. *Pharmacoepidemiol Drug Saf* 2002; **11**(2): 121–125.
12. Common Data Model (CDM) Specification. Observational Medical Outcomes Partnership (OMOP) Research Lab; Working Document. Version 1.2. May 18, 2009 at http://omop.fnih.org/sites/default/files/OMOP%20CDM%20Specifications%2022may2009_Post.pdf (2009, accessed 25 September 2010).
13. Pesquita C, Faria D, Falcão AO, Lord P and Couto FM. Semantic similarity in biomedical ontologies. *PLoS Comput Biol* 2009; **5**: 7.
14. The Medical Dictionary for Regulatory Activities (MedDRA). http://www.meddramsso.com/index.asp (accessed 20 September 2010).
15. Systematized Nomenclature of Medicine-Clinical Terms (SNOMED). http://www.ihtsdo.org/snomed-ct/ (accessed 20 September 2010).
16. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**(1): 41–55.
17. Mera MR, Beach JK, Powell EG and Pattishall NE. Semi-automated risk estimation using large databases:

quinolones and *Clostridium difficile* associated diarrhea. *Pharmacoepidemiol Drug Saf* 2010; **19**: 610–617.
18. Hochberg A. 'Risk Estimation Analytical Methods', ProSanos Corporation at http://www.prosanos.com/uploaded/RiskEstimationAnalyticalMethods.pdf (accessed 22 September 2010).
19. Brookhart MA, Schneeweiss S, Rothman K, Glynn RJ, Avorn J and Sturmer T. Variable selection in propensity score models. *Am J Epidemiol* 2006; **163**: 1149–1156.
20. Peduzzi P, Concato J, Kemper E, Holford TR and Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; **49**(12): 1373–1379.
21. Vittinghoff E and McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007; **165**(6): 710–718.
22. Haerian H, McHugh P, Brown R, Somes G and Solomon SS. Gatifloxacin produces both hypoglycemia and hyperglycemia: a retrospective study. *Am J Med Sci* 2008; **335**(2): 95–98.
23. Nordstrom BL, Sung I, Suter P and Szneke P. Risk of pneumonia and other complications of influenza-like illness in patients treated with oseltamivir. *Curr Med Res Opin* 2005; **21**(5): 761–768.
24. McAfee AT, Ming EE, Seeger JD, et al. The comparative safety of rosuvastatin: a retrospective matched cohort study in over 48,000 initiators of statin therapy. *Pharmacoepidemiol Drug Saf* 2006; **15**(7): 444–453.
25. Wang J, Mullins CD, Mamdani M, Rublee DA and Shaya FT. New diagnosis of hypertension among celecoxib and nonselective NSAID users: a population-based cohort study. *Ann Pharmacother* 2007; **41**(6): 937–943.
26. Gill SS, Rochon PA, Herrmann N, et al. Atypical antipsychotic drugs and risk of ischaemic stroke: population based retrospective cohort study. *BMJ* 2005; **330**(7489): 445, Epub 2005 Jan 24.
27. L'Allier PL, Ducharme A, Keller PF, Yu H, Guertin MC and Tardif JC. Angiotensin-converting enzyme inhibition in hypertensive patients is associated with a reduction in the occurrence of atrial fibrillation. *J Am Coll Cardiol* 2004; **44**(1): 159–164.
28. Seeger JD, West WA, Fife D, Noel GJ, Johnson LN and Walker AM. Achilles tendon rupture and its association with fluoroquinolone antibiotics and other potential risk factors in a managed care population. *Pharmacoepidemiol Drug Saf* 2006; **15**(11): 784–792.
29. Cheng RM, Mamdani M, Jackevicius CA and Tu K. Association between ACE inhibitors and acute pancreatitis in the elderly. *Ann Pharmacother* 2003; **37**(7–8): 994–998.
30. Suissa S, Hudson M and Ernst P. Leflunomide use and the risk of interstitial lung disease in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**(5): 1435–1439.
31. Ray WA, Daugherty JR and Griffin MR. Lipid-lowering agents and the risk of hip fracture in a Medicaid population. *Inj Prev* 2002; **8**(4): 276–279.
32. Cadarette SM, Katz JN, Brookhart MA, Stürmer T, Stedman MR and Solomon DH. Relative effectiveness of osteoporosis drugs for preventing nonvertebral fracture. *Ann Intern Med* 2008; **148**(9): 637–646.
33. Schade R, Andersohn F, Suissa S, Haverkamp W and Garbe E. Dopamine agonists and the risk of cardiac-valve regurgitation. *N Engl J Med* 2007; **356**(1): 29–38.
34. Johannes CB, Ziyadeh N, Seeger JD, Tucker E, Reiter C and Faich G. Incidence of allergic reactions associated with antibacterial use in a large, managed care organization. *Drug Saf* 2007; **30**(8): 705–713.
35. Abraham NS, El-Serag HB, Hartman C, Richardson P and Deswal A. Cyclooxygenase-2 selectivity of non-steroidal anti-inflammatory drugs and the risk of *myocardial infarction and cerebrovascular accident*. Aliment Pharmacol Ther *2007; 25(8): 913–924*.
36. Smitten AL, Choi HK, Hochberg MC, et al. The risk of herpes zoster in patients with rheumatoid arthritis in the United States and the United Kingdom. *Arthritis Rheum* 2007; **57**(8): 1431–1438.
37. Duncan E, Dunlop BW, Boshoven W, Woolson SL, Hamer RM and Phillips LS. Relative risk of glucose elevation during antipsychotic exposure in a Veterans Administration population. *Int Clin Psychopharmacol* 2007; **22**(1): 1–11.
38. Cole JA, Norman H, Doherty M and Walker AM. Venous thromboembolism, myocardial infarction, and stroke among transdermal contraceptive system users. *Obstet Gynecol* 2007; **109**(2 Pt 1): 339–346. Erratum in: Obstet Gynecol 2008; 111(6): 1449.
39. Patterson MK, Castellsague J and Walker AM. Hospitalization for peptic ulcer and bleeding in users of selective COX-2 inhibitors and nonselective NSAIDs with special reference to celecoxib. *Pharmacoepidemiol Drug Saf* 2008; **17**(10): 982–988.
40. van Staa TP, Boulton F, Cooper C, Hagenbeek A, Inskip H and Leufkens HG. Neutropenia and agranulocytosis in England and Wales: incidence and risk factors. *Am J Hematol* 2003; **72**(4): 248–254.
41. Meropol SB, Chan KA, Chen Z, et al. Adverse events associated with prolonged antibiotic use. *Pharmacoepidemiol Drug Saf* 2008; **17**(5): 523–532.
42. Moses LE, Mosteller F and Buehler JH. Comparing results of large clinical trials to those of meta-analyses. *Stat Med* 2002; **21**: 793–800.
43. Ioannidis JPA, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001; **286**: 821–830.
44. Tannen RL, Weiner MG and Marcus SM. Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *J Clin Epidemiol* 2006; **59**(3): 254–264.
45. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
46. Landis JR and Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; **33**: 363–374.
47. Barrett P. Assessing the reliability of rating data. *Intraclass Correlation and Variance Component Methods* April 2 2007.
48. Poole C. Low P-values or narrow confidence intervals: which are more durable? *Epidemiology* 2001; **12**(3): 291–294.
49. Feinstein AR and Cicchetti DV. High agreement but low kappa: I. *The problems of two paradoxes*. J Clin Epidemiol 1990; **43**(6): 543–549.
50. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H and Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**(4): 512–522.
51. i3 Aperio at http://www.i3global.com/Solutions/DrugSafety/i3Aperio/ (accessed 20 September 2010).
52. Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley & Sons, 1986.
53. Xie F, Li SC, Lo NN, Yeo SJ, Yang KY, Yeo W, et al. Cross-cultural adaptation and validation of Singapore English and Chinese Versions of the Oxford Knee Score (OKS) in knee osteoarthritis patients undergoing total knee replacement. *Osteoarthritis Cartilage* 2007; **15**(9): 1019–24. Epub 2007 Apr 3.
54. Liang KY and Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**: 13–22.