PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2011; **20**: 1–11
Published online 8 November 2010 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.2053

ORIGINAL REPORT

Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project

Preciosa M. Coloma[1]*, Martijn J. Schuemie[1], Gianluca Trifirò[1], Rosa Gini[2], Ron Herings[3], Julia Hippisley-Cox[4], Giampiero Mazzaglia[5], Carlo Giaquinto[6], Giovanni Corrao[7], Lars Pedersen[8], Johan van der Lei[1] and Miriam Sturkenboom[1,9] on behalf of the EU-ADR consortium

[1]*Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[2]*Agenzia Regionale di Sanità, della Toscana Florence, Italy*
[3]*PHARMO Institute, Utrecht, The Netherlands*
[4]*Division of Primary Care–School of Community Health Sciences, University of Nottingham, Nottingham UK*
[5]*Società Italiana di Medicina Generale, Florence, Italy*
[6]*Pedianet-Società Servizi Telematici SRL, Padova, Italy*
[7]*Department of Statistics, Università di Milano-Bicocca, Milan, Italy*
[8]*Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark*
[9]*Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands*

ABSTRACT

Purpose In this proof-of-concept paper we describe the framework, process, and preliminary results of combining data from European electronic healthcare record (EHR) databases for large-scale monitoring of drug safety.

Methods Aggregated demographic, clinical, and prescription data from eight databases in four countries (Denmark, Italy, Netherlands, the UK) were pooled using a distributed network approach by generation of common input data followed by local aggregation through custom-built software, Jerboa©. Comparison of incidence rates of upper gastrointestinal bleeding (UGIB) and nonsteroidal anti-inflammatory drug (NSAID) utilization patterns were used to evaluate data harmonization and quality across databases. The known association of NSAIDs and UGIB was employed to demonstrate sensitivity of the system by comparing incidence rate ratios (IRRs) of UGIB during NSAID use to UGIB during all other person-time.

Results The study population for this analysis comprised 19 647 445 individuals corresponding to 59 929 690 person-years of follow-up. 39 967 incident cases of UGIB were identified during the study period. Crude incidence rates varied between 38.8 and 109.5/100 000 person-years, depending on country and type of database, while age-standardized rates ranged from 25.1 to 65.4/100 000 person-years. NSAID use patterns were similar for databases within the same country but heterogeneous among different countries. A statistically significant age- and gender-adjusted association between use of any NSAID and increased risk for UGIB was confirmed in all databases, IRR from 2.0 (95%CI:1.7–2.2) to 4.3 (95%CI: 4.1–4.5).

Conclusions Combining data from EHR databases of different countries to identify drug-adverse event associations is feasible and can set the stage for changing and enlarging the scale for drug safety monitoring. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS — drug safety; signal detection; upper gastrointestinal bleeding; nonsteroidal anti-inflammatory drugs; EU-ADR

Received 25 March 2010; Revised 3 July 2010; Accepted 18 August 2010

INTRODUCTION

Every year numerous drugs are introduced into the international healthcare market. In 2008 alone, a total of 66 medicinal products received a positive opinion from the European Medicines Agency (EMA)[1] while 99 drugs were approved by the US Food and Drug Administration (FDA).[2] These new drugs often represent important advances that improve the care or quality of life for many patients worldwide. While a drug's efficacy and safety must be demonstrated in randomized controlled clinical trials prior to approval,

* Correspondence to: P. M. Coloma, Ee 2116, Department of Medical Informatics, Erasmus Medical Center, PO Box 2040, 3000CA Rotterdam, The Netherlands. E-mail: p.coloma@erasmusmc.nl

these trials are rarely designed or powered to detect uncommon or unexpected adverse events.[3–6] Safety surveillance of drugs in the post-marketing phase has traditionally been performed by analyses of spontaneous case reports for signal generation and of observational healthcare databases for signal evaluation.[7–9] The sudden worldwide withdrawal of Rofecoxib (Vioxx) in 2004[10] stimulated an international review of how best to measure and monitor drug safety so that the balance of risks and benefits can be continually and rapidly evaluated.[11–14] As mandated by the FDA Amendments Act of 2007, the US is establishing the Sentinel System, a nationwide network of databases targeted to capture by the year 2012 data on more than 100 million subjects for prospective drug safety surveillance.[15] The European Commission (EC) has similarly issued initiatives to develop new methodologies for drug safety monitoring based on analysis of large databases. Under the auspices of the EC, the EU-ADR Project (''Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge'') (http://www.euadr-project.org) was launched in 2008. EU-ADR aims to exploit information from various electronic healthcare record (EHR) databases in Europe to produce a computerized integrated system for the early detection of drug safety signals. Once detected, the safety signals will be substantiated by semantic mining of literature and computational analysis of pharmacological and biological information on drugs, molecular targets, and pathways (Figure 1).[16]

In this proof-of-concept paper, we endeavor to set the stage for large-scale drug safety monitoring by describing the framework, process, and preliminary results of combining data from a federation of eight EHR databases in four countries, a resource of unprecedented size for monitoring of drug safety in Europe. We further illustrate the challenges encountered in the amalgamation of data from diverse locations into a uniform repository and the obstacles that had to be overcome in terms of heterogeneity in database structure, language, coding of drugs and diseases, and diversity in the organization of European healthcare systems.

METHODS

Study population and data sources

The EU-ADR database platform currently comprises anonymous healthcare data from eight established European databases located in four countries. Health Search/CSD Patient (HSD, Italy),[17,18] Integrated

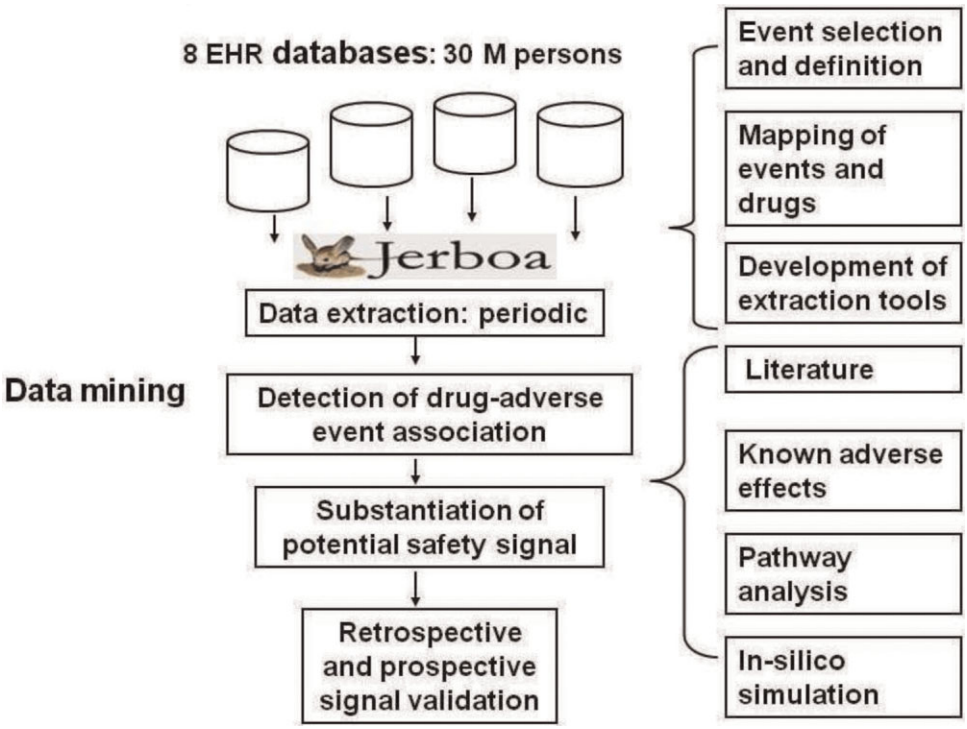

Figure 1. EU-ADR Project Schema

DOI: 10.1002/pds

Primary Care Information (IPCI, Netherlands),[19,20] Pedianet[21,22] (Italy), and QRESEARCH[23,24] (United Kingdom) are general practice (GP) databases, where both clinical information and drug prescriptions are recorded. The Aarhus University Hospital Database (Aarhus, Denmark),[25,26] the PHARMO Network[27,28] (Netherlands), and the regional Italian databases of Lombardy[29,30] and Tuscany[31,32] are all comprehensive record-linkage systems in which drug dispensing data of a well-defined population are linked to a registry of hospital discharge diagnoses and various other registries. Table 1 provides an overview of characteristics of each database. Most healthcare services, including pharmaceutical services, are provided for, or subsidized by, the state in Italy (ITA), Denmark (DK), and the United Kingdom (UK) and covered by obligatory health insurance in the Netherlands (NL) and turnover is low. In all of the countries with GP databases, GPs function as ''gatekeepers'' of the healthcare system.

Common data framework: distributed data processing

Founded on the basic governance principle that database owners should be involved in the elaboration of data as they best understand the context within which the data are recorded, we have chosen in the EU-ADR Project a distributed network approach that requires standardization of input files from the different databases (Figure 2). These common input files (patient, drug, and event files) are created locally and are subsequently managed by purpose-built software called Jerboa©, which has been tested against different scripts. Jerboa© uses flat text files as input and is written entirely in Java™ to ensure that it will run in a wide variety of computational environments. The software queries patient-level data in the different databases, which are later aggregated, de-identified and sent in encrypted format to a central repository for evaluation and further analysis. This repository is managed by the Department of Medical Informatics at Erasmus Medical Center in the Netherlands, the project's coordinating center.

Cohort definition and follow-up time

To harmonize follow-up definitions across databases, we defined three dates for each patient file: (1) the eligibility period begins on the date the patient is registered in the database; (2) the eligibility period ends on the date the patient transfers out of the system, with the last supply of data, or on the patient's death, whichever is earlier; and (3) the date of birth. From these dates and in combination with the drug prescription file, Jerboa© marks the beginning of follow-up, which is the start of eligibility date plus one year, or the date of birth (for subjects whose date of birth-start of eligibility is less than one year).

Drug exposure

Drug prescriptions and dispensings are locally coded using the national product codes, which differ among countries (see Table 1). Most countries, however, link these product codes to the Anatomical Therapeutic Chemical (ATC) classification system. The ATC code is used as the drug code in the EU-ADR input files. Each database owner estimates the duration covered by each prescription or dispensing according to the legend duration (if dosing regimen is available), or is otherwise based on the defined daily dose (DDD).[33]

Event definition

The EU-ADR Project started out by defining a selected number of events that are subsequently mapped to a common terminology system. This process has been described in more detail in separate publications.[16,34,35] Databases in EU-ADR use one of four nomenclature systems to describe the events: the International Classification of Diseases (ICD9-CM and ICD10), the International Classification of Primary Care (ICPC), and the READ Code (RCD) classification. These different terminologies are mapped using the Unified Medical Language System® (UMLS®). The UMLS is a biomedical terminology integration system handling more than 150 terminologies including the four used in the EU-ADR project.[36] Ascertainment of the event of interest from the databases follows an iterative process with seven stages: (1) event definition using clinical criteria established from literature; (2) identification of UMLS concepts corresponding to the event; (3) revision and validation of medical concepts by database owners and pharmacovigilance experts; (4) translation of the medical concepts into each database terminology; (5) extraction of data and computation of event rates; (6) comparison of query structure—to detect and harmonize eventually any major disagreement across databases; and (7) creation of event input files for Jerboa©.

Statistical analyses

Incidence rates. We calculated age- and gender-specific incidence rates of the event of interest within

DOI: 10.1002/pds

Table 1. EHR database characteristics

Characteristics	Pedianet (Italy)	HSD (Italy)	Lombardy Regional (Italy)	Tuscany Regional (Italy)	IPCI (Netherlands)	PHARMO (Netherlands)	QRESEARCH (UK)	Aarhus (Denmark)
Current source population	160 000 children	1 500 000	9 000 000	3 500 000	1 500 000	3 000 000	4 000 000	1 800 000
Years covered for this study	2003–2007	2003–2007	2003–2005	2003–2006	1996–2006	1998–2007	2000–2007	2001–2006
Type of database	General Practice pediatric database	General Practice database	Administrative	Administrative	General Practice database	Hybrid (administrative and medical record/registries)	General Practice database	Administrative
Age range	0–14	From 15 onwards	All ages	All ages	All ages	All ages	All ages	All ages
% Males	52.2	47.2	48.8	48.1	49.6	45.8	49.6	49.9
Demographic information available								
Date of registration	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of transferring out	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of birth	MM-YY	MM-YY	DD-MM-YY	DD-MM-YY	MM-YY	DD-MM-YY	YY	MM-YY
Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ethnicity/race	No	No	No	No	No	No	No	No
Drug information available								
Product coding	MINSAN	MINSAN	MINSAN	MINSAN	HPK	Z index	EMIS	Varenr
Active international principle coding system	ATC	ATC	ATC	ATC	ATC	ATC	BNF	ATC
Date of prescription/ dispensing	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dosing regimen	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Quantity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indication of use	Yes	Yes	No	No	Yes	Yes for in-hospital	No	Yes
Outcome information available								
Symptoms (Yes/No)	Yes, as free text/codes	Yes, as free text/codes	No	No	Yes, as free text/codes	Yes for some	Yes, as codes	No
Outpatient primary care diagnoses	Yes, as free text /codes	Yes Free text /codes	No	No	Yes, as free text /codes	No	Yes	No
Outpatient specialist care diagnoses	Yes, as free text/codes	Yes	No	No	Yes	No	Yes	No
Hospital discharge diagnoses	Yes, as free text /codes	Yes, as free text /codes	Yes	Yes	Yes, as free text /codes	Yes	Yes	Yes
Diagnosis coding scheme	ICD-9CM	ICD-9CM	ICD-9CM	ICD-9CM	ICPC	ICD-9CM	RCD	ICD-10
Diagnostic procedures	Yes	Yes	Yes	Yes	No	Yes for in-hospital interventions	Yes	Yes, in-hospital only
Laboratory tests	Yes	Yes	No	No	Yes	Yes subset	Yes	Yes, in-hospital only

ICPC, International Classification of Primary Care; ICD9-CM, International Classification of Diseases–9th revision Clinical Modification; RCD, READ CODE Classification; ICD-10, International Classification of Diseases–10th revision; MINSAN, Italian Ministry of Health; HPK, Handelsproductcode; EMIS, Egton Medical Information Systems; Varenr, Danish Ministry of Health coding.

DOI: 10.1002/pds

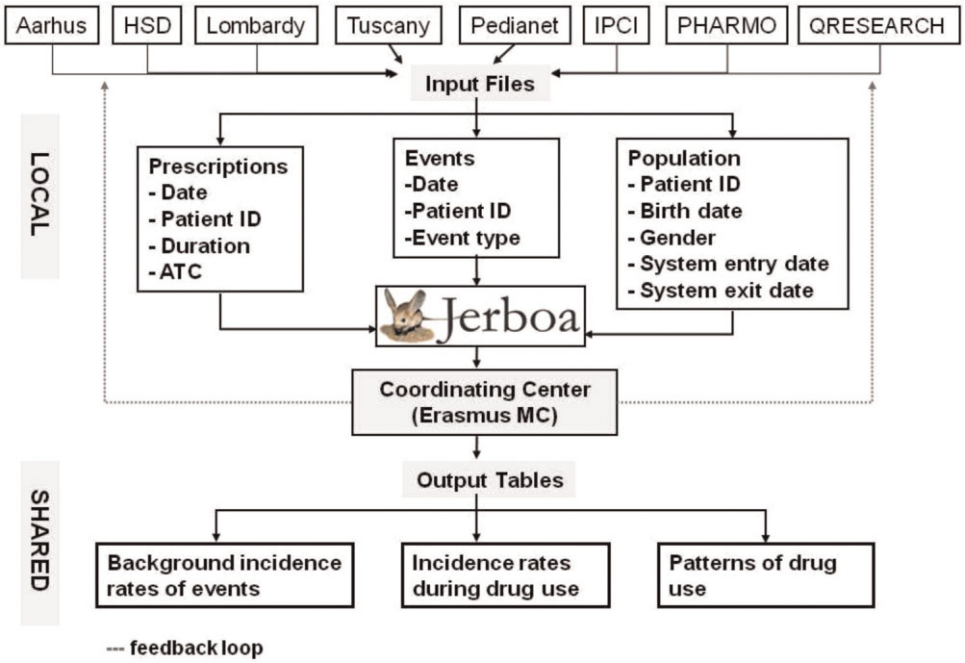

Figure 2. Common data framework: distributed network

each database as a means to compare and benchmark the data extraction. We performed direct standardization using the WHO World Standard Population as reference to account for age differences when comparing the overall rates.[37] Jerboa© manages and aggregates the data over all patients locally, producing as output the amount of person-time of follow-up and the number of events per gender and age group. To calculate the incidence rates, we only considered the first recorded occurrence of the event of interest after a one year run-in period.

Associations. Drug prescription and/or dispensing data were used to estimate drug utilization and event rates during drug exposure. Overlapping treatment episodes with the same drug (same ATC code) are combined into a single episode of drug use that starts when the first prescription begins and stops when the last prescription ends. When a patient uses more than one drug at a time, the corresponding person-time is labeled accordingly. Using individual data on start date and end date of prescription or dispensing, Jerboa© determines and marks as unexposed those periods during which an individual is included in the study but is not using any drug. Events are then assigned to the episodes (drug use/non-use) in which they occurred.

The validity of the methodology previously described has to be evaluated in terms of the system's objective of drug safety monitoring. The system's ability to identify a drug-adverse event pair for which the association is established (*i.e.*, true positive signal) will provide an indication of its sensitivity. We illustrate this methodology on actual data, first focusing on the known association between use of NSAIDs and increased risk of upper gastrointestinal bleeding (UGIB). For the initial estimation of the association between event and drug use, we summarized exposure to each drug in tables, stratified by age, gender, and calendar year. We performed the analyses at three different ATC levels: pharmacological subgroup (*i.e.*, M01A); chemical subgroup (*i.e.*, M01AB, M01AC, etc.); and chemical substance (*i.e.*, M01AC01, etc.). Using the Mantel–Haenszel method, we calculated age- and gender-adjusted rate ratios, with all remaining person-time (*i.e.*, non-use and use of all other drugs) as reference.

RESULTS

Study population

The total study population comprised 19 647 445 individuals corresponding to 59 929 690 person-years (PYs) of follow-up. The databases contributed varying follow-up time to the study period which covered the years 1996–2007. Comparison of follow-up time across databases revealed similar age and gender distributions, except for Pedianet which, by design,

DOI: 10.1002/pds

only includes patients aged less than 14 years and HSD which covers only patients older than 14 years old (see Figure 3).

Incidence rates of upper gastrointestinal bleeding

We identified a total of 39 967 incident cases of UGIB in the study population. Analysis of crude overall incidence rates demonstrated heterogeneity with non-standardized rates of UGIB ranging from 38.8 (PHARMO) to 61.1 (IPCI) per 100 000 PYs in the Netherlands to 87.7 (Aarhus) and 109.5 (HSD) per 100 000 PYs in Denmark and Italy, respectively (Table 2). Incidence rates for the UK database QRESEARCH (84.3) and the regional Italian databases of Tuscany (70.7) and Lombardy (52.5) were somewhere in between. Direct age standardization attenuated the variation in overall incidence rates between databases. Table 2 also shows the incidence rates of UGIB separately for children (14 years old and below) and those older than 15 years.

Across all databases, incidence rates increased with advancing age (Figure 4), the risk for UGIB in individuals 70 years and older was 5.2 times greater compared to those aged 50–59 years old (95% CI: 5.1–5.4). There were also consistently higher incidence rates in males overall compared to females (incidence rate ratio (IRR) 1.24 (95% CI: 1.22–1.27)).

Patterns of NSAID use

Patterns of use of NSAID classes varied among different countries but were similar among different databases in the same country (Figure 5). The Dutch databases IPCI and PHARMO showed identical utilization profiles, with the acetic acid derivatives (*e.g.*, diclofenac) representing about 40% of all NSAID exposure time. The UK database QRESEARCH had a profile similar to NL, while the Danish database Aarhus showed more use of propionic acid derivatives (*e.g.*, naproxen) compared to the acetic acid derivatives (47% vs. 30%). Except for the pediatric database Pedianet, the other three Italian databases had similar drug use patterns with relatively high use of the ATC M01AX class of NSAIDs (other NSAIDs). This group of drugs, which includes nimesulide, represented the lowest percentage of NSAID use in the other databases (except Aarhus). Heterogeneity in exposure among countries became more apparent when we explored the use of individual drugs (data not shown).

Association of NSAID use with upper gastrointestinal bleeding

Overall, we detected 4 934 incident cases of UGIB during NSAID use. Data from Pedianet were not taken into account in this analysis due to low number of events in children. The incidence rates were consist-

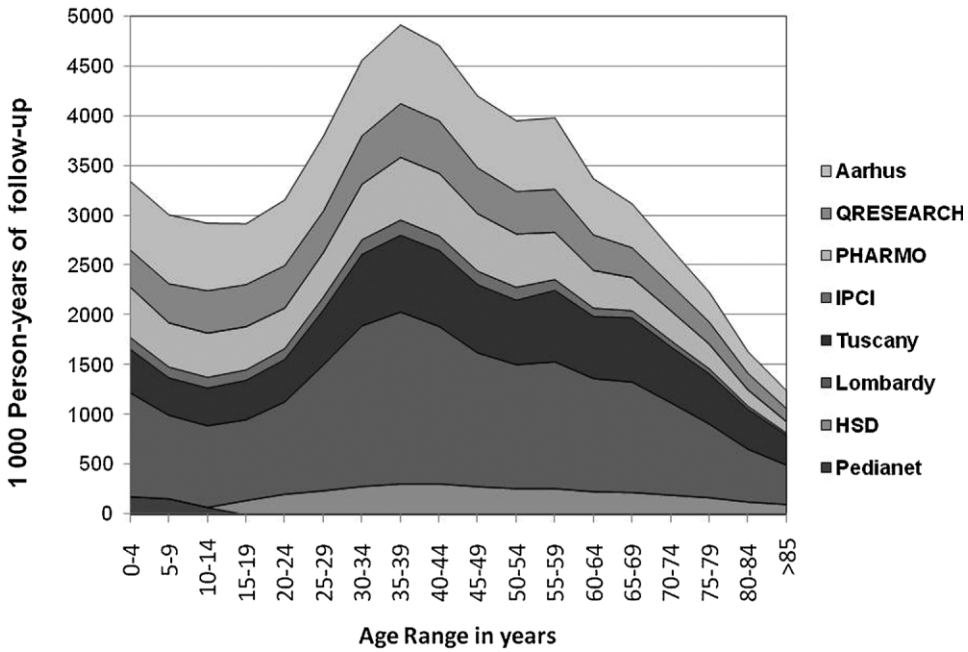

Figure 3. Distribution of follow-up time per age group across databases

DOI: 10.1002/pds

Table 2. Incidence rates of UGIB (per 100 000 PYs)

Country	Database	No. of events	Population-time	Overall incidence rate (*age-standardized*)	Incidence rate in <15 years (*age-standardized*)	Incidence rate in ≥15 years (*age-standardized*)
ITALY	Pedianet	65	407 150	16.0 (*14.5*)	16.0 (14.5)	-
	HSD	3639	3 324 089	109.5 (*65.4*)	-	109.7 (*71.6*)
	Lombardy	10 127	19 282 776	52.5 (*29.2*)	9.8 (9.3)	59.5 (*36.2*)
	Tuscany	7024	9 940 626	70.7 (*31.6*)	13.3 (13.0)	78.5 (*38.2*)
NETHERLANDS	IPCI	1139	1 862 932	61.1 (*44.3*)	7.6 (7.5)	73.2 (*57.4*)
	PHARMO	2944	7 591 284	38.8 (*25.1*)	5.0 (4.9)	46.3 (*32.3*)
UNITED KINGDOM	QRESEARCH*	5721	6 788 121	84.3 (*60.1*)	15.4 (16.0)	99.0 (*75.8*)
DENMARK	Aarhus	9308	10 611 047	87.7 (*54.6*)	3.7 (3.8)	108.0 (*72.6*)
Total		39 967	59 807 984	66.8	9.1	77.4

*Represent data from 30% of database population.

ently around 3–4 per 1000 PYs of NSAID exposure, except for PHARMO which had lower incidence rates (1.9/1000 PYs) and Aarhus which had higher incidence rates (6.5/1000 PYs). The rates of UGIB were significantly increased during use of NSAIDs as compared to all other follow-up time in each of the individual databases (Table 3), with IRRs ranging from 2.0 (95% CI 1.7 to 2.2) to 4.3 (95% CI 4.1 to 4.5).

DISCUSSION

We have developed and tested a methodology that enables combining data from EHR databases of various countries and origins (medical records, administrative registries, record-linkage databases). Revisiting the known association of UGIB and nonsteroidal anti-inflammatory drug (NSAID) use, we have shown that data sharing can take place within this distributed networking system to provide consistent incidence rates and detect a known drug-adverse event association. This network system yields other interesting side products such as age- and gender-specific disease rates and drug utilization patterns across a wide variety of settings. Together this opens a new avenue for conducting observational research on a wider, more global perspective.

In combining the data of various databases it was crucial to take into account ethical issues regarding the processing of anonymized healthcare data. The databases involved in EU-ADR have considerable

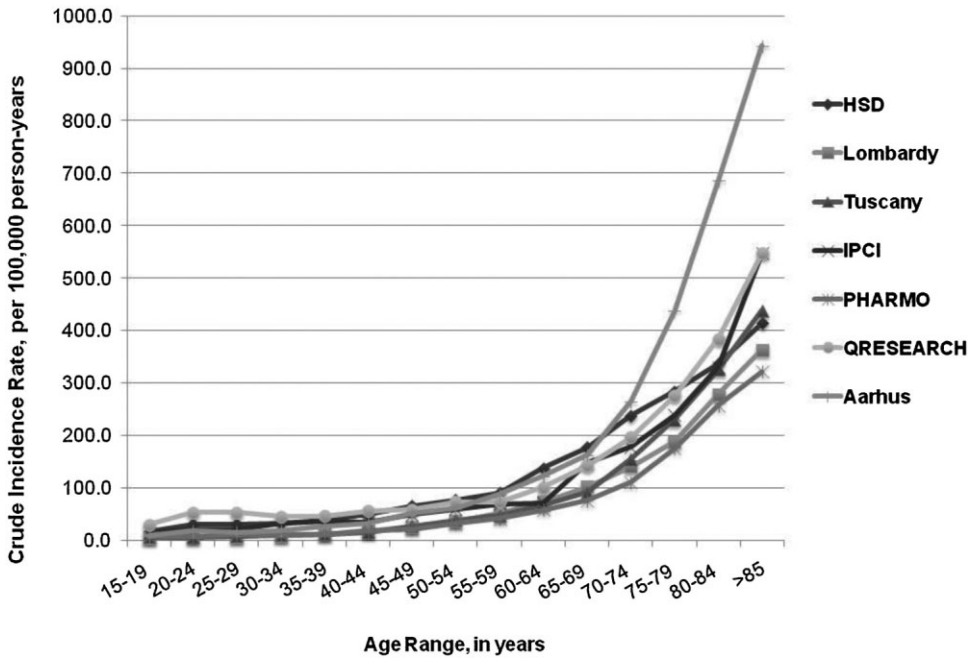

Figure 4. Incidence rates of UG1B across databases, age 15 years and above

DOI: 10.1002/pds

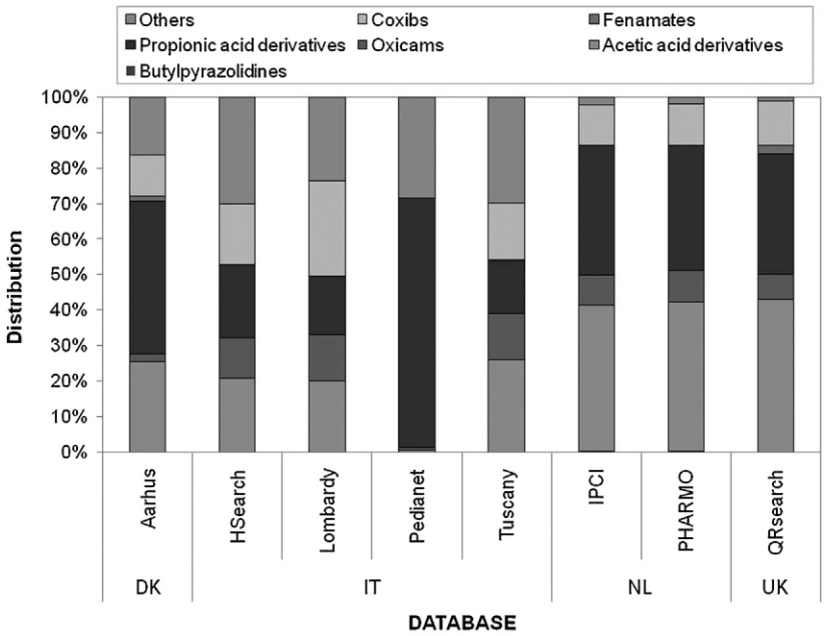

Figure 5. Comparison of use of specific NSAID classes across databases

experience in using patient data for research purposes, and have well-developed safeguard mechanisms ensuring compliance with the European directives and national regulations, as well as database governance rules. Since no new data are collected, other than those made available by the participating databases, the cornerstone of ensuring proper ethical and legal conduct is set down in the rules and regulations that govern each database. Rather than imposing a one-size-fits-all approach and compel people to change their data, we leveraged on the diversity of the databases to use local expertise and to maximize extraction of relevant information, hence effectively dealing with all methodological, cultural, ethical, governance, and political issues of sharing data across borders. Databases retain ownership of their respective data, extraction being done locally, and only the aggregated, de-identified data are shared with the rest of the network.

Table 3. IRRs of UGIB during NSAID use

Country	Database	No. of events	Exposure*	Incidence rate†	Rate ratio‡ (95%CI)
ITA	HSD	250	81 734	3.1	2.0 (1.7–2.2)
	Lombardy	991	314 852	3.1	2.9 (2.7–3.1)
	Tuscany	698	205 012	3.4	2.4 (2.3–2.6)
NL	IPCI	116	26 780	4.3	4.0 (3.3–4.9)
	PHARMO	342	177 698	1.9	2.8 (2.5–3.2)
UK	QRESEARCH	467	158 783	2.9	2.4 (2.2–2.6)
DK	Aarhus	2070	316 348	6.5	4.3 (4.1–4.5)
Total		4934	1 281 207	3.9	

*In person-years.
†Per 1000 PYs.
‡Age and gender-adjusted; non-NSAID use as comparator; p value $<<<$ 0.01.

A similar concept of distributed processing of healthcare data has been employed by other research collaborations, albeit with different research objectives. This model has previously been described in bioterrorism and syndromic surveillance as well as vaccine safety surveillance.[38–44] The ongoing Sentinel Initiative of the FDA is also adopting a distributed data architecture for combining healthcare databases to improve drug safety monitoring.[45–47] While the scale may be comparable, there are issues in combining data that are unique to Europe. Challenges stem from the fact that different countries have distinct natural languages, aside from having different drug and disease coding systems. The diversity of healthcare systems throughout Europe makes merging data from databases a more complex task that requires striking a balance between international cooperation and adequate protection of patient confidentiality.[48]

DOI: 10.1002/pds

Incidence rates

Standardized incidence rates were generally higher than the crude incidence rates, the European EU-ADR network population being older than the reference WHO World Standard Population. Comparison of incidence rates was used for benchmarking of the data extraction process and the heterogeneity in incidence rates of UGIB reflects interesting peculiarities in the populations themselves, in the type of database, and in the respective healthcare systems. There may be inter-country differences as to causes of bleeding, risk factors, and variations in the extent of diagnostic evaluation. Different sources of data are apt to capture UGIB of varying severity, the events identified from hospitalization records likely to be more severe than the events ascertained from general practitioner (GP) visits. This may explain the lower incidence rates observed in administrative databases compared to general practice databases in the same country (*e.g.*, HSD vs. Tuscany or IPCI vs. PHARMO).

Although there is not yet a gold standard to validate our results, the observed incidence rates were in line with previously published literature. Most studies on the epidemiology of UGIB are hospital-based studies that cite incidence varying from 48/100 000 adults in the Netherlands to 103/100 000 adults in the UK.[49–52] Other reports estimate the incidence to be 95/100 000 adults (USA) and 172 (Scotland).[53,54] EU-ADR provides data specifically for children. To date, there are no published population-based studies that include children below 15 years of age, which underlines the importance of this type of information. The increase in risk of UGIB with advancing age and with men compared to women, which we observed in all of the databases, are consistent with the literature.[55] Exact comparisons between studies are difficult because of discrepancies in event and drug exposure definitions, selection criteria of study populations, and methodologies. The advantage of EU-ADR is that the variations in event definition are documented in a standard way and the methodological differences in rate estimations have been eliminated which allows for easier comparison across countries and databases.

Sensitivity of the system

In general, the relative risk of UGIB during NSAID use is around four.[56] This is similar to what we obtained in this study even if the rate ratios were adjusted only for age and gender and were compared to all other person-time (*i.e.*, exposed to other drugs or unexposed). The consistency of the degree of association between NSAID use and UGIB within the databases studied reiterates the huge burden of drug-related disease. Its concordance with published literature increases the face validity of the methodology and reinforces the potential of drug safety monitoring and risk quantification being performed on a very different scale, without limitations as to the size or to the number of databases. Key steps for success in this process were the chosen distributed approach that dealt with different governance issues, the definition of common data input files to deal with database heterogeneity, the availability of common customized software that allowed for local elaboration of data, and terminology mapping using the Unified Medical Language System. Improving the sensitivity of the database network system will require further development of scalable methods for better control of bias and confounding to allow further inferences regarding causality.

While the motivation for merging disparate data sources primarily comes from the need to investigate drug safety in larger populations, it is a well-acknowledged limitation that EHR databases are only able to describe exposures and outcomes of interest to the extent that they are documented within the database systems.[57–59] The databases capture information on outpatient (pharmacy-based) drug use. Most databases do not capture all vaccinations (*e.g.*, those provided in the childhood vaccination programs) but do register, for example influenza, vaccination. Specific drug groups of interest, such as biologicals, are captured if provided through routine dispensing system. The type of drugs that can be captured in the system is being addressed in a separate paper. Furthermore, the issue of accuracy and completeness of information concerning diagnoses and actual drug use is common to all types of databases and some degree of misclassification is unavoidable. Although review of every individual medical record in each database would incur prohibitive costs, some sensitivity analyses should be done to determine the validity of the extracted data.

CONCLUSION

In this proof-of-concept paper, we have demonstrated the feasibility of combining diverse and differently structured data in an effective way to detect comparative risks of potential adverse drug events and pave the way for large-scale drug safety monitoring. The common data framework described takes advantage of multiple, routinely collected, aggregated healthcare data while minimizing sharing of confidential patient-level information. Although still a work-in-progress, this system

DOI: 10.1002/pds

KEY POINTS

- Large-scale monitoring of drug safety is currently hampered by the diversity and isolation of healthcare databases which stem from differences in language, coding, structure of healthcare systems, as well as issues of confidentiality and data protection.
- The EU-ADR Project demonstrates the feasibility of combining EHR databases in Europe using a distributed network approach that allows combining databases by generation of common input data and local aggregation through custom-built software to pave the way for large-scale drug safety monitoring.
- Age-adjusted incidence rates of upper gastrointestinal bleeding (UGIB), drug utilization patterns, and confirmation of increased risk of UGIB with nonsteroidal anti-inflammatory drug (NSAID) use across all databases demonstrated the viability of data harmonization using a common data model.

can facilitate the planning of permanent, global-scale, EHR-based structures for the early detection of drug safety signals.

CONFLICT OF INTEREST

There is no specific product being studied. M. S. is running a research group that occasionally performs studies for pharmaceutical industries according to unconditional grants. The companies include AstraZeneca, Pfizer, Lilly, Boehringer. MS has been consultant to Lundbeck and Pfizer in the past.

G. C. received educational grants from Abbott, GSK, GSK-Bio, Gilead, BMS, Tibotec, and SP-MSD.

J. H. C. is director of QRESEARCH which is a partnership between the University of Nottingham and EMIS (commercial supplier of 56% of GP clinical computer systems in the UK).

ACKNOWLEDGEMENTS

This research has been funded by the European Commission Seventh Framework Programme (FP7/2007–2013) under grant no. 215847–The EU-ADR Project.

REFERENCES

1. http://www.emea.europa.eu/pdfs/general/direct/emeaar/AnnualReport2008.pdf [Accessed 21 October 2009].
2. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Reports.ReportsMenu [Accessed 01 October 2009].
3. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312**: 1215–1218.
4. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006; **174**: 635–641. DOI: 10.1503/cmaj.050873.
5. Stricker BHCh, Psaty BM. Detection, verification, and quantification of adverse drug reactions. *BMJ* 2004; **329**: 44–47.
6. Ray Wayne A. Population-based studies of adverse drug effects. *New Engl J Med* 2003; **349**: 1592–1594.
7. Curri CJ, Macdonald TM. Use of routine healthcare data in safe and cost-effective drug use. *Drug Saf* 2000; **22**(2): 97–102.
8. Rodriguez EM, Staffa JA, Graham DJ. The role of databases in drug postmarketing surveillance. *Pharmacoepidemiol Drug Saf* 2001; **10**: 407–410. DOI: 10.1002/pds.615.
9. Härmark L, van Grootheest AC. Pharmacovigilance: methods, recent developments and future perspectives. *Eur J Clin Pharmacol* 2008; **64**: 743–752. DOI: 10.1007/s00228-008-0475-9.
10. Singh D. Merck withdraws arthritis drug worldwide. *BMJ* 2004; **329**(7470): 816. DOI: 10.1136/bmj.329.7470.816-a.
11. Avorn J. Evaluating drug effects in the post-Vioxx world: there must be a better way. *Circulation* 2006; **113**: 2173–2176. DOI: 10.1161/CIRCULATIONAHA.106.625749.
12. Wadman M. Experts call for active surveillance of drug safety. *Nature* 2007; **446**(7134): 358–359. DOI: 10.1038/446358b; Published online: 21 March 2007.
13. EMEA. European Risk Management Strategy: 2008–2009 work programme adopted. London, 2 January, 2008. http://www.hma.eu/uploads/media/ERMS_2008-2009_work_programme_adopted.pdf [Accessed 03 February 2009].
14. FDA. The Future of Drug Safety-Promoting and Protecting the Health of the Public, January 2007. http://www.fda.gov/oc/reports/iom013007.pdf [Accessed 03 February 2009].
15. http://www.fda.gov/oc/initiatives/advance/reports/report0508.pdf [Accessed 23 February 2009].
16. Trifirò G, Pariente A, Coloma PM, *et al*. on behalf of the EU-ADR consortium. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* 2009; **18**(12): 1176–1184. DOI: 10.1002/pds.1836; Published online: 15 September 2009.
17. Sacchetti E, Trifirò G, Caputi A, *et al*. Risk of stroke with typical and atypical anti-psychotics: a retrospective cohort study including unexposed subjects. *J Psychopharmacol* 2008; **22**(1): 39–46. DOI: 10.1177/0269881107080792.
18. Mazzaglia G, Mantovani LG, Sturkenboom MC, *et al*. Patterns of persistence with antihypertensive medications in newly diagnosed hypertensive patients in Italy: a retrospective cohort study in primary care. *J Hypertens* 2005; **23**(11): 2093–2100.
19. Vlug AE, van der Lei J, Mosseveld BM, *et al*. Postmarketing surveillance based on electronic patient records: the IPCI project. *Methods Inf Med* 1999; **38**: 339–344.
20. Verhamme K, Mosis G, Dieleman J, Stricker B, Sturkenboom M. Spironolactone and risk of upper gastrointestinal events: population based case-control study. *BMJ* 2006; **333**(7563): 330. DOI: 10.1136/bmj.38883.479549.2F; Published online: 13 July 2006.
21. Sturkenboom MCJM, Nicolosi A, Cantarutti L, *et al*. for the NSAIDs Paediatric Research Group. Incidence of mucocutaneous reactions in children treated with niflumic acid, other non steroidal anti-inflammatory drugs and analgesics. *Pediatrics* 2005; **116**(1): e26–e33. DOI: 10.1542/peds.2004-0040; Published online: June 1 2005.
22. Giaquinto C, Van Damme P, Huet F, *et al*. Clinical consequences of rotavirus acute gastroenteritis in Europe, 2004–2005: The REVEAL Study. *J Infect Dis* 2007; **195**, S26–S35. DOI: 10.1086/516717.
23. Hippisley-Cox J, Coupland CAC, Logan R. Risk of adverse gastro-intestinal outcomes in patients taking cyclooxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *BMJ* 2005; **331**: 1310–1316. DOI: 10.1136/bmj.331.7528.1310.
24. Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. *Inform Prim Care* 2004; **12**(1): 49–50.
25. Christensen S, Riis A, Nørgaard M, Thomsen W, Sørensen HT. Introduction of newer selective cyclooxygenase-2 inhibitors and rates of hospitalization with bleeding and perforated peptic ulcer. *Aliment Pharmacol Ther* 2007; **25**: 907–912. DOI: 10.1111/j.1365-2036.2007.03274.x.
26. Dalton SO, Johansen C, Poulsen AH, *et al*. Cancer risk among users of neuroleptic medication: a population-based cohort study. *Br J Cancer* 2006; **95**: 934–939. DOI: 10.1038/sj.bjc.6603259; Published online: 22 August 2006.
27. Penning-van Beest FJ, Termorshuizen F, Goettsch WG, Klungel OH, Kastelein JJ, Herings RM. Adherence to evidence-based statin guidelines reduces the risk of hospitalizations for acute myocardial infarction by 40%: a cohort study. *Eur Heart J* 2007; **28**(2): 154–159. DOI: 10.1093/eurheartj/ehl391; Published online: 7 December 2006.

28. Goettsch WG, de Jong RB, Kramarz P, Herings RM. Developments of the incidence of osteoporosis in The Netherlands: a PHARMO study. *Pharmacoepidemiol Drug Saf* 2007; **16**(2): 166–172. DOI: 10.1002/pds.1245.
29. Corrao G, Zambon A, Conti V, *et al*. Menopause hormone replacement therapy and cancer risk: an Italian record linkage investigation. *Ann Oncol* 2007; **19**: 150–155. DOI: 10.1093/annonc/mdm404; Published online: 4 September 2007.
30. Corrao G, Botteri E, Bertù L, Zambon A, Favilli S. Exploring the effect of transient exposure on the risk of acute events by means of time-window designs: an application to fluoroquinolone antibacterials and arrhythmia. *Pharmacoepidemiol Drug Saf* 2006; **15**: 31–37. DOI: 10.1002/pds.1125.
31. Barchielli A, Balzi D, Pasqua A, Buiatti E. Incidence of acute myocardial infarction in Tuscany, 1997–2002: data from the Acute Myocardial Infarction Registry of Tuscany (Tosc-AMI). *Epidemiol Prev* 2006; **30**(3): 161–168.
32. Monami M, Balzi D, Lamanna C, *et al*. Are sulphonylureas all the same? A cohort study on cardiovascular and cancer-related mortality. *Diabetes Metab Res Rev* 2007; **23**(6): 479–484. DOI: 10.1002/dmrr.736; Published Online: 23 March 2007.
33. http://www.whocc.no/atcddd/ [Accessed 14 February 2009].
34. Trifiro G, Fourrier-Reglat A, Sturkenboom MC, Díaz Acedo C, van der Lei J. The EU-ADR project: preliminary results and perspective. *Stud Health Technol Inform.* 2009; **148**: 43–49. DOI: 10.3233/978-1-60750-043-8-43.
35. Avillach P, Mougin F, Joubert M, *et al*. A semantic approach for the homogeneous identification of events in eight patient databases: a contribution to the European eu-ADR project. *Stud Health Technol Inform* 2009; **150**: 190–194. DOI: 10.3233/978-1-60750-044-5-190.
36. http://www.nlm.nih.gov/research/umls/about_umls.html [Accessed 14 February 2009].
37. http://www.who.int/healthinfo/paper31.pdf [Accessed 06 January 2009].
38. Platt R, Davis R, Finkelstein J, *et al*. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf* 2001; **10**: 373–377. DOI: 10.1002/pds.607; Published Online: 14 December 2001.
39. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. *BMC Public Health* 2006; **6**: 235. DOI: 10.1186/1471-2458-6-235.
40. Lieu TA, Kulldorff M, Davis RL, *et al.* Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 2007; **45**(10 Supl 2): S89–S95. DOI: 10.1097/MLR.0b013e3180616c0a.
41. Brown JS, Kulldorff M, Chan KA, *et al*. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf* 2007; **16**: 1275–1284. DOI: 10.1002/pds.1509; Published Online: 22 October 2007.
42. McMurry AJ, Gilbert CA, Reis BY, *et al*. A self-scaling, distributed information architecture for public health, research, and clinical care. *J Am Med Inform Assoc* 2007; **14**(4): 527–533. DOI: 10.1197/jamia.M2371.
43. Velentgas P, Bohn RL, Brown JS, *et al*. A distributed research network model for post-marketing safety studies: the meningococcal vaccine study. *Pharmacoepidemiol Drug Saf* 2008; **17**: 1226–1234. DOI: 10.1002/pds.1675; Published Online: 27 October 2008.
44. Brown JS, Kulldorff M, Petronis KR, *et al*. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf* 2009; **18**(3): 226–234. DOI: 10.1002/pds.1706; Published Online: 15 January 2009.
45. http://www.fda.gov/oc/initiatives/criticalpath/behrman_interview.html [Accessed 23 February 2009].
46. http://omop.fnih.org/?q=node/34 [Accessed 23 February 2009].
47. Brown J, Lane K, Moore K, Platt R. Defining and Evaluating Possible Database Models to implement the FDA Sentinel Initiative. http://www.regulations.gov/fdmspublic/component/main?main=DocumentDetail&d=FDA-2009-N-0192-0005 [Accessed 12 June 2009].
48. Rynning E. Public trust and privacy in shared electronic health records. *Eur J Health Law* 2007; **14**(2): 105–112.
49. van Leerdam ME, Vreeburg EM, Rauws EA, *et al*. Acute upper GI bleeding: did anything change? Time trend analysis of incidence and outcome of acute upper GI bleeding between 1993/1994 and 2000. *Am J Gastroenterol* 2003; **98**(7): 1494–1499. DOI: 10.1111/j.1572-0241.2003.07517.x.
50. Rockall TA, Logan RF, Devlin HB, Northfield TC. Incidence of and mortality from acute upper gastrointestinal haemorrhage in the United Kingdom. Steering committee and members of the National Audit of Acute Upper Gastrointestinal Haemorrhage. *BMJ* 1995; **311**(6999): 222–226.
51. Laporte JR, Vidal X, Carne X. Non-steroidal anti-inflammatory drugs in elderly people. Gastrointestinal bleeding is common. *BMJ* 1995; **311**: 391–392.
52. Loperfido S. Incidence of gastrointestinal bleeding in Italy is similar to that in Britain. *BMJ* 1995; **311**(7009): 874.
53. Longstreth GF. Epidemiology of hospitalization for acute upper gastrointestinal hemorrhage: a population-based study. *Am J Gastroenterol* 1995; **90**(2): 206–210.
54. Blatchford O, Davidson LA, Murray WR, Blatchford M, Pell J. Acute upper gastrointestinal haemorrhage in west of Scotland: case ascertainment study. *BMJ* 1997; **315**: 510–514.
55. van Leerdam ME. Epidemiology of acute upper gastrointestinal bleeding. *Best Pract Res Clin Gastroenterol* 2008; **22**(2): 209–224. DOI: 10.1016/j.bpg.2007.10.011.
56. Hernandez Diaz S, Garcia Rodriguez LA. Association of nonsteroidal anti-inflammatory drugs and upper gastrointestinal bleeding/perforation. *Arch Intern Med* 2000; **160**: 2093–2099.
57. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**: 323–337. DOI: 10.1016/j.jclinepi.2004.10.012.
58. Strom BL, Kimmel SE (eds.). *Textbook of Pharmacoepidemiology.* John Wiley & Sons: Chichester, 2006; pp. 167–170.
59. Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects–advantages and disadvantages. *Nat Clin Pract Rheumatol* 2007; **3**(12): 725–732. DOI: 10.1038/ncprheum0652.