

# JADE: A tool for medical researchers to explore adverse drug events using health claims data

D. Edlinger<sup>1</sup>; S.K. Sauter<sup>1</sup>; C. Rinner<sup>1</sup>; L.M. Neuhofer<sup>1</sup>; M. Wolzt<sup>2</sup>; W. Grossmann<sup>3</sup>; G. Endel<sup>4</sup>; W. Gall<sup>1</sup>

<sup>1</sup>Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria; <sup>2</sup>Department of Clinical Pharmacology, Medical University of Vienna, Austria; <sup>3</sup>Research Group Scientific Computing, University of Vienna, Austria; <sup>4</sup>Main Association of Austrian Social Security Organizations, Vienna, Austria

## Keywords

Patient safety, drug-drug interactions, adverse drug events, adverse drug reaction reporting systems, medical informatics applications

## Summary

**Objective:** The objective of our project was to create a tool for physicians to explore health claims data with regard to adverse drug reactions. The Java Adverse Drug Event (JADE) tool should enable the analysis of prescribed drugs in connection with diagnoses from hospital stays.

**Methods:** We calculated the number of days drugs were taken by using the defined daily doses and estimated possible interactions between dispensed drugs using the Austria Codex, a database including drug-drug interactions. The JADE tool was implemented using Java, R and a PostgreSQL database.

**Results:** Beside an overview of the study cohort which includes selection of gender and age groups, selected statistical methods like association rule learning, logistic regression model and the number needed to harm have been implemented.

**Conclusion:** The JADE tool can support physicians during their planning of clinical trials by showing the occurrences of adverse drug events with population based information.

## Correspondence to:

Doron Edlinger  
Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna  
Spitalgasse 23  
A-1090 Vienna  
Austria  
E-mail: de@doron-edlinger.de

## Appl Clin Inform 2014; 5: 621–629

<http://dx.doi.org/10.4338/ACI-2014-04-RA-0036>

received: April 4, 2014

accepted: May 25, 2014

published: July 9, 2014

**Citation:** Edlinger D, Sauter SK, Rinner C, Neuhofer LM, Wolzt M, Grossmann W, Endel G, Gall W. JADE: A tool for medical researchers to explore adverse drug events using health claims data. *Appl Clin Inf* 2014; 5: 621–629

<http://dx.doi.org/10.4338/ACI-2014-04-RA-0036>

## Introduction

Older patients are particularly affected by polypharmacy. A survey taken in the United States [1] showed that about 20% of participants aged 65 years and older took at least five drugs per week. Polypharmacy increases the risk of drug-drug interactions and adverse drug events (ADEs) which can cause hospitalisation. A prospective study [2] linked 6.5% of hospital admissions to ADEs. 16.6% of these admissions were linked to drug-drug interactions listed in literature. 2.3% of the patients admitted to hospital due to an ADE died. Retrospective studies based on discharge diagnoses linked 1.83% of non-planned hospital admissions in the Netherlands with a median stay of 5 days [3] and 0.92% of inpatient admissions in Germany with a median stay of 3 days [4] to ADEs.

Following market authorisation, knowledge about ADEs is mostly gained through spontaneous reporting systems in the context of pharmacovigilance and clinical trials. Reporting systems are subjected to underreporting, a review of 37 articles [5] showed estimated underreporting rates of 6% to 100% with a median of 94%. In Austria, only 823 reports were filed in 2012 [6]. A survey among doctors [7] listed several reasons for not reporting adverse drug reactions. In order to identify rare ADEs, clinical trials would need a high number of cases, which is often not achievable.

In our view, exploring health claims data is a contribution to the generation of hypotheses about ADEs and can help physicians in their daily work, especially while planning clinical trials. The Main Association of the Austrian Social Insurance Institutions provides the research database GAP-DRG [8] which contains dispensed medications, hospital stays and related diagnoses as well as socio-demographic attributes.

In this work we describe JADE, a software tool that facilitates exploration of health claims data with regard to ADEs, which, in this project, are defined as any adverse event related to drugs. We used the experience of a previous project related to adverse drug events [9]. After selecting a set of patients, predefined statistical procedures can be executed to analyse their medication and hospital data.

## Methods

The JADE tool uses health claims data from the GAP-DRG database. The GAP-DRG database contains health claims data of all public health insurance companies of the years 2006 and 2007, and therefore of all Austrians who received healthcare services in 2006 and 2007 (including inpatient and outpatient care like services of hospitals, general practitioners or prescriptions). Names, social security numbers and precise dates of birth are not contained in the GAP-DRG database. Since the temporal proximity of drug intake is important for the estimation of hospitalisations due to ADEs as well as for the interaction between drugs, we selected only patients insured by health insurance companies which provide the actual date on which a medication was dispensed. In order to ensure a sufficient data quality we only considered patients whose year of birth and gender are documented in the database. To exclude children and patients whose year of birth was probably listed incorrectly we excluded patients younger than 20 or older than 99 years, respectively those born after 1987 or before 1908. To be able to analyse the hospitalisations due to ADEs we used a preliminary lead (February 14<sup>th</sup> 2006 to June 30<sup>th</sup> 2006) and one year time period (July 1<sup>st</sup> 2006 to June 30<sup>th</sup> 2007). Only patients having at least one prescription in this time period were considered. The study cohorts, including gender, year of birth, dispensed prescriptions and information about hospital stays were combined in a separate database schema. For each hospital stay, exactly one main diagnosis and several additional diagnoses are documented as ICD-10 codes in the GAP-DRG database. Both main diagnoses and additional diagnoses were added to the hospital stays in the schema and are not distinguished in later analysis. To identify hospital diagnoses which are associated with ADEs (and to highlight them in the tool) we used the ICD-10 codes listed by Stausberg and Hasford [10] and adapted the codes to the Austrian coding conditions.

The Austria Codex [11], a register containing all drugs which are authorised in Austria, was used to identify potential drug-drug interactions between two prescriptions of a patient, which were taken at the same time according to our estimation. The Austrian pharmaceutical registration number (in German called “Pharmazentralnummer”) was used as an identifier for drugs which

were mapped to Anatomical Therapeutic Chemical Classification System (ATC) codes for the later identification of a drug in the JADE tool.

Drugs were considered to be relevant for a hospitalisation if prescribed less than 2 months prior to the date of hospital admission and if the theoretical duration of intake overlapped with the date of hospitalisation. For each prescription we used the defined daily dose (DDD) and the packages to calculate the theoretical number of days the drug was taken. The DDD was assigned to the drug using the ATC-DDD classification [12].

The application of the JADE tool is to provide its users an overview of hospitalisations due to ADEs and drug-drug interactions of an entire year in Austria and help to answer questions like: “Does the prescription of a drug combination that is a risky combination according to the Austria Codex, depend on socio-demographic parameters like age, gender and days the patient spent in hospital?”. The tool supports two basic evaluation steps: (i) Selection of a study cohort by specifying socio-demographic parameters and medical conditions. (ii) Execution of predefined statistical procedures using one or two ATC codes and a hospital diagnosis (ICD-10 Code) as input parameter. Three statistical procedures were implemented: Association rule learning, logistic regression analysis and calculation of the “number needed to harm”.

The JADE tool was developed using open source technologies. The database is hosted on a PostgreSQL server. The JADE tool itself is programmed with Java. The graphic user interface is realized using Swing and JGoodies libraries. JFreeChart is used to display interactive diagrams. To allow executing advanced statistical procedures, R is integrated and called from the Java application using the JRI library. We used the R package RPostgreSQL to access the database directly from R to avoid the loading process of data in the Java application and then pass it to R.

## Results

The GAP-DRG database contains data of about 8 million patients. Excluding patients insured by health insurance companies that do not provide the actual dispensing date of the prescription, this number was reduced to about 3 million patients. Only 1,279,197 of these patients had a dispensed prescription during the time we focused on. After applying the exclusion criteria, data of 1,032,596 patients remained for use. For this patient cohort, about 26.4 million dispensed prescriptions and 0.4 million hospital stays are documented and about 11.7 million interaction warnings were calculated.

The sequence diagram in ► Figure 1 describes the processes and the involved actors in the JADE tool. After logging in, a personalised study cohort can be selected and visualised using the JADE tool. Using this cohort, three statistical procedures can be selected and the calculations are propagated to R. R loads the needed data from the GAP-DRG database, calculates the results and sends back the data to the JADE tool where the results are presented to the user. The software architecture was designed for simple extensibility: each statistical procedure is implemented in a separate class which share database access, encapsulated access to R and a generic user interface, which allows showing different types of results including numerical values, tables, interactive diagrams and diagrams generated using R.

## Selection of a study cohort

In the initial screen of the JADE tool the user can specify a study cohort by selecting gender and predefined age groups (20 to 29 years, 30 to 39 years and so on). To define interesting subgroups, patients' prescriptions were used to create a study cohort of diabetes mellitus patients and patients that received painkilling drugs as a permanent medication for more than 6 months. 73,914 diabetes patients and 58,026 patients with permanent treatment with pain medication were identified. After confirming the selected parameters and the selection of a study cohort, the user can start its analysis. The number of patients in the user's study cohort is given, as well as a histogram which depicts the distribution of age groups and gender. Currently the following three statistical procedures are implemented in the JADE tool.

## Medication given prior to hospital stays

The first procedure aims to analyse the patients' medication given prior to their hospitalisation with one specific hospital diagnosis as depicted in ► Figure 2. More specifically, after selecting an ICD-10 code, hospital stays related to that diagnosis are identified and the support of drugs and combinations of drugs given to patients prior to these stays is shown. The support indicates the fraction of hospital stays before which a specific drug or drug combination was prescribed to the patient. A hospital stay is related to a diagnosis when documented as its main diagnosis or one of its additional diagnoses. The support is also shown for hospital stays not related to the selected diagnosis to allow distinguishing between commonly prescribed drugs of the selected cohort and those correlated with the selected diagnosis.

If, for instance, published studies indicate a certain drug increases the risk of heart attacks, the JADE tool can list prescribed medication prior to hospital stays with the diagnosis heart attack.

To assist the user selection of an interesting diagnosis, all diagnoses that are related to hospitalisations due to adverse drug events and drug-drug interactions are highlighted in red in the searchable list.

## Prediction of risky prescriptions using socio-demographic parameters

The second use case allows the user to select ATC codes or combination of ATC codes which cause an interaction according to the Austria Codex and then analyse the patient cohort who received the selected drug or combination of drugs. This involves use of a logistic regression model which explains if an ATC code was prescribed or not to a patient. Since socio-demographic parameters are used in the model, the entire study cohort of 1,032,596 patients needs to be analysed. The used input parameters are the patient's age, gender and total duration of hospital stays (stratified into groups).

The result of the logistic regression analysis shows which parameters are significantly correlated to the prescription and how strong their influence is. The performance is illustrated as ROC curve (Receiver operating characteristic curve). Its Area under the curve is a rough indicator that states how well the model fits; a value above 0.7 signifies a good fit. An example of an ROC curve generated by JADE is depicted in ► Figure 3.

## Number needed to harm

In general, the number needed to harm (NNH) indicates how many patients need to be exposed to a risk to harm one additional person. In this use case, the harm represents hospitalisation due to medication. The user selects a hospital diagnosis (ICD-10 code) and a drug or combination of drugs that causes an interaction (ATC codes). The calculated NNH is the number of patients that need to receive the selected drug or combination of drugs to cause one additional hospitalisation related to the selected diagnosis. In the calculation, a hospitalisation was considered to be related to the selected diagnosis when it was the main diagnosis or one of the additional diagnoses documented for that hospitalisation. A negative NNH indicates that there is no correlation between the selected drug or drug combination and the selected diagnosis.

## Discussion

Health claims data are a valuable resource in addition to clinical trials. These data are collected routinely and thereby allow exploring a large number of individuals in studies at low cost. This increases the chance of detecting rare events. Health claims data have been used increasingly for epidemiologic research: Tricco et al. reviewed 325 studies based on two Canadian health claims databases [13], Hoffmann identified 70 studies based on German health claims data published from 1998 to 2007 with more than half of the studies being published from 2006 to 2007 [14].

However, limitations arise when health claims data are reused for clinical research. As to the JADE tool, the patients' prescriptions are not fully covered: drugs administered or dispensed in hospitals and drugs which are not reimbursed by the insurance (e.g. over-the-counter drugs sold di-

rectly to a consumer without prescription) are not documented. The results achieved by the JADE tool cannot be applied directly to the whole of Austria because we exclusively selected patients whose health insurance companies report the actual dispensing date for a drug. Therefore, the selection of patients was not random. Furthermore the prescribed dose and time period of the drug intake are not documented. In general, the quality of diagnoses documented for hospital stays needs to be questioned. For example, in a study conducted in Australia Sansom et al. suggest that only for a fraction of hospital stays related to adverse drug reactions a drug-related diagnosis has been documented [15]. However, in the Austrian payment system for some medical procedures validation checks are implemented which yield a warning if a procedure is being billed without documenting a consistent diagnosis.

Although users need to consider these characteristics and limitations, the JADE tool allows analysing these data without technical expertise and can therefore be used by medical research scientists or physicians. They can gain valuable information on prescriptions, drug interactions and adverse drug events which can be used for planning clinical trials and formulation of hypotheses. Additionally it can be used to determine which groups of the population have received a particular drug which is useful when adverse reactions to this drug are reported.

The JADE tool was intended to be a prototype to show the potential of such tools and the opportunities of analysing health claims data with regard to adverse drug events and drug-drug interactions that cause hospitalisation. The three use-cases were developed in collaboration with physicians. Further procedures can be added with little effort. As a next step, the tool can be used and evaluated by physicians in different fields of research.

#### **Acknowledgments**

Financial support for this project was provided by the Main Association of Austrian Social Security Institutions.

#### **Clinical Relevance**

The JADE tool allows physicians to explore adverse drug events by means of health claims data. Results obtained from the tool can be used for generating hypotheses and planning clinical trials.

#### **Conflict of Interest**

G. Endel is an employee of the Main Association of Austrian Social Security Institutions. No other conflict of interest is declared.

#### **Human Subjects Protections**

The procedures used have been reviewed in compliance with ethical standards of the responsible committee on human experimentation

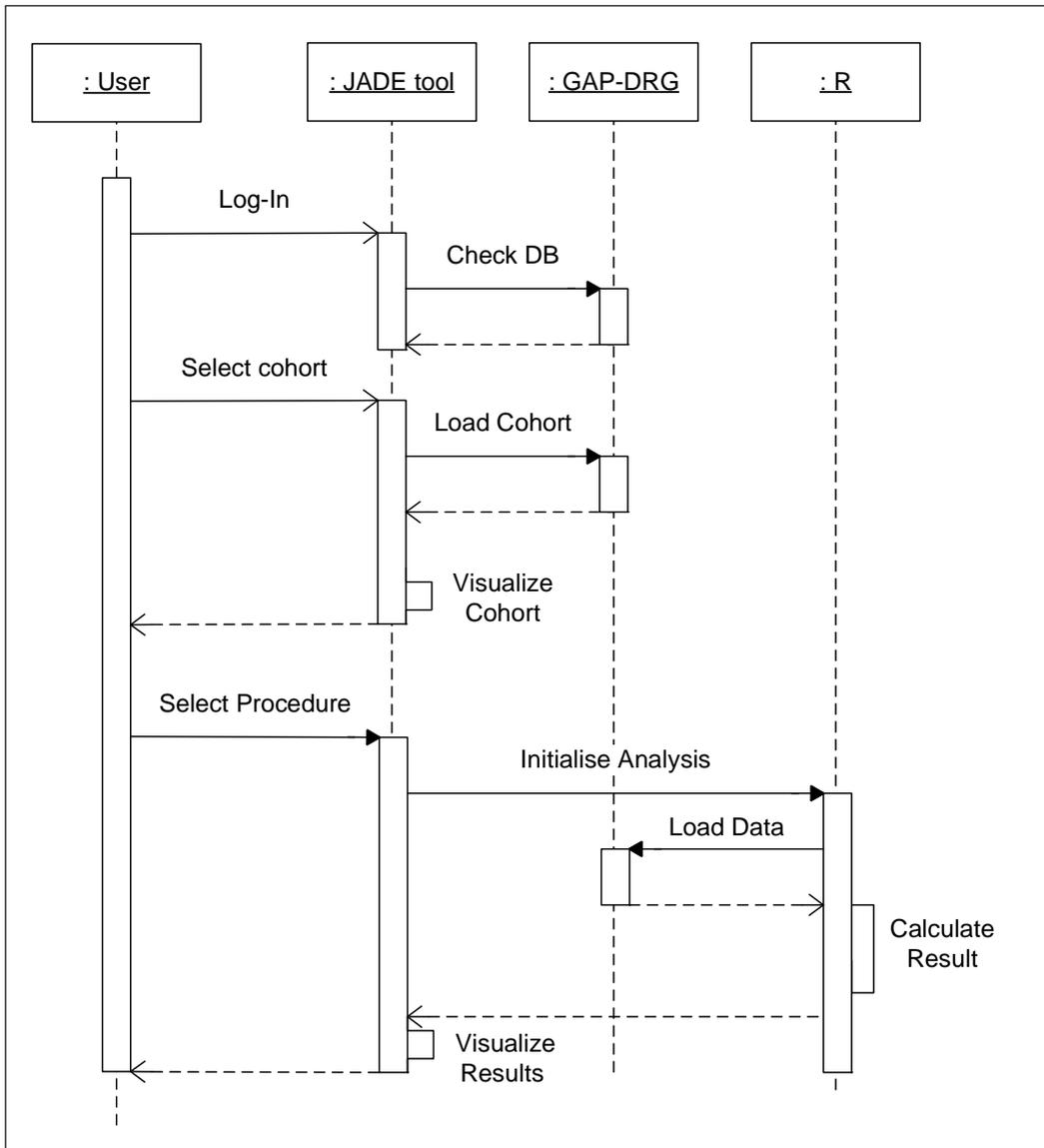
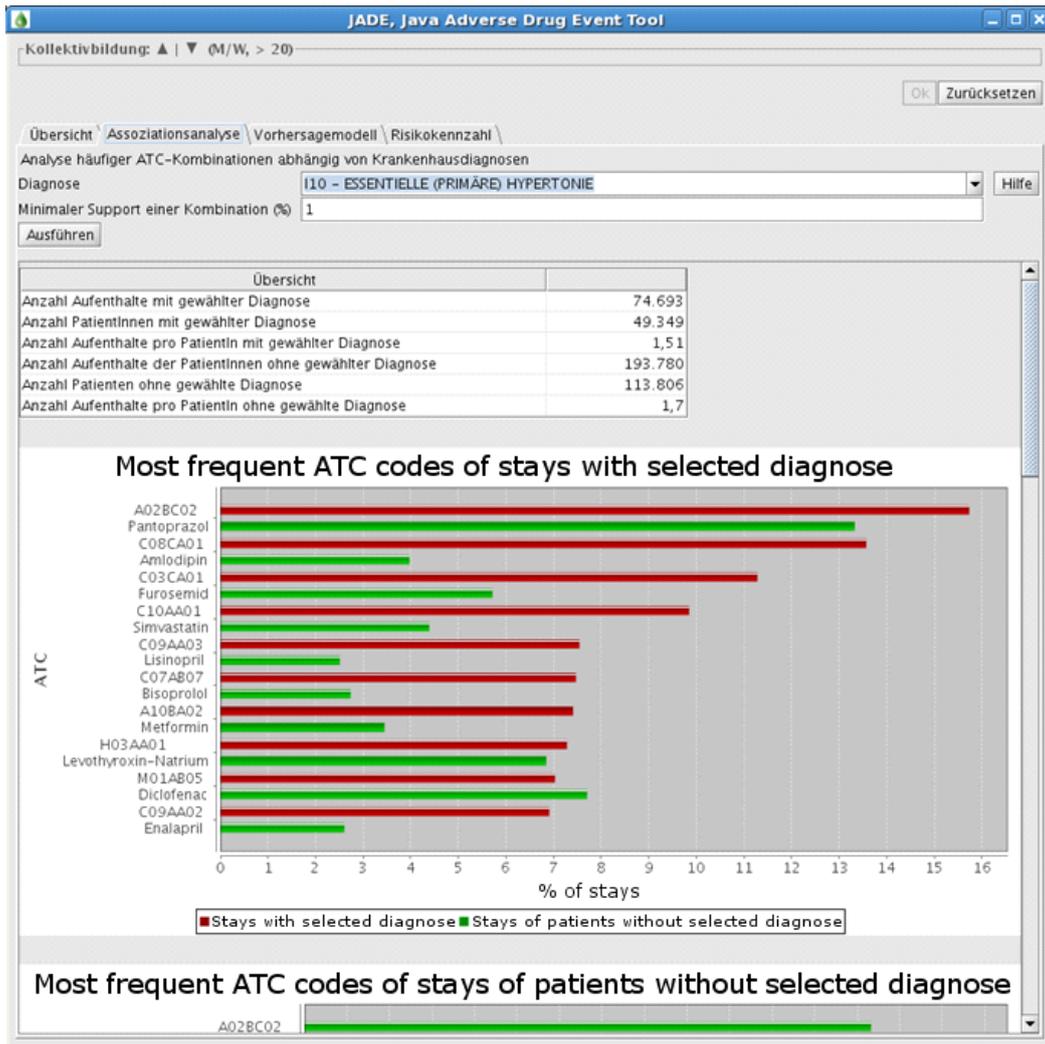
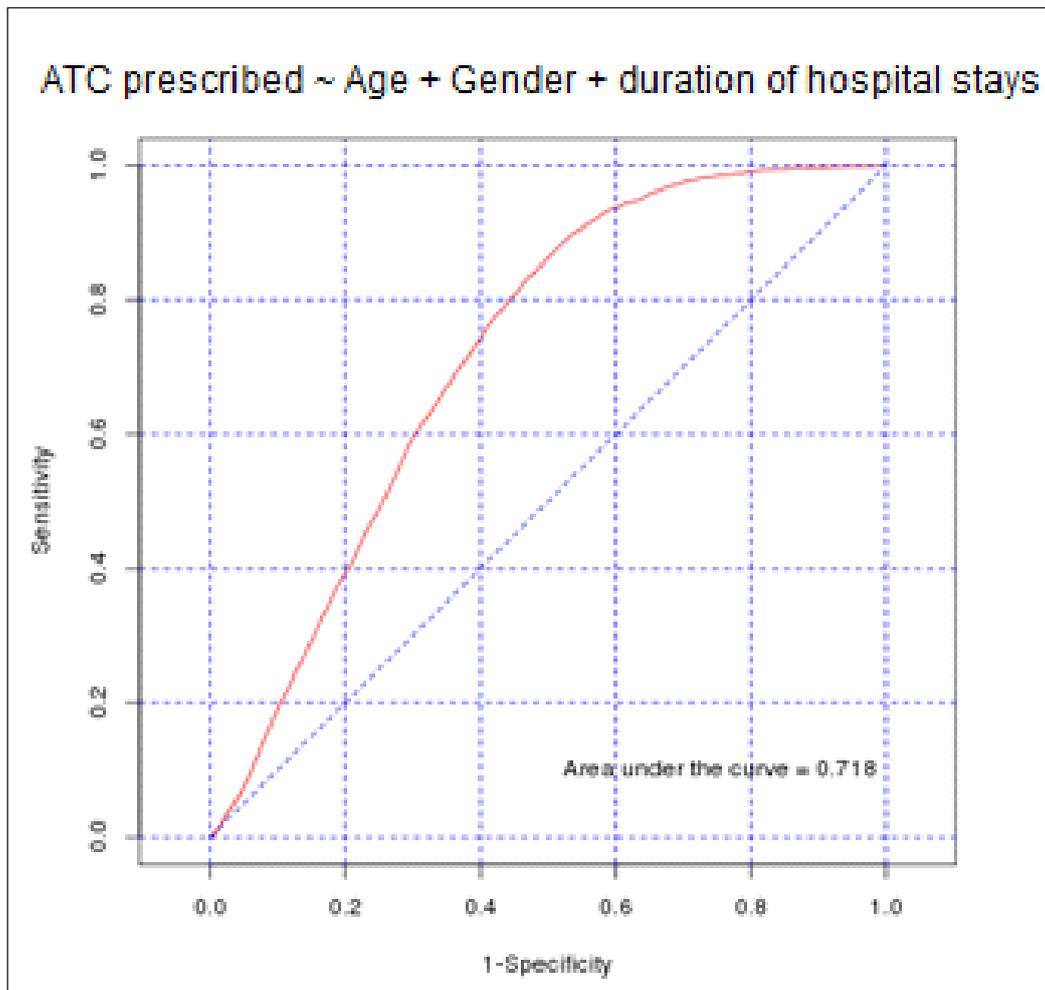


Fig. 1 Sequence diagram describing the processes and the involved actors in the JADE tool.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.



**Fig. 2** JADE user interface. Drugs given prior to hospital stays with the diagnosis of essential (primary) hypertension. Red bars indicate the fraction of hospital stays related to the selected diagnosis before which the ATC code shown on the left was prescribed. Green bars indicate the fraction of hospital stays not related to the selected diagnosis before which the ATC code shown on the left was prescribed.



**Fig. 3** ROC curve of the linear regression model used in JADE to predict the prescription of “ACE inhibitors and calcium channel blockers” (ATC code C09BB) based on age, gender and total duration of hospital stays. The area under the curve is 0.718.

## References

1. Kaufman D, Kelly J, Rosenberg L, Anderson T, Mitchell AA. Recent patterns of medication use in the ambulatory adult population of the united states: The slone survey. *JAMA* 2002; 287(3): 337-344.
2. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 2004; 329(7456): 15-19.
3. van der Hooft CS, Sturkenboom MC, van Grootheest K, Kingma HJ, Stricker BHC. Adverse drug reaction-related hospitalisations. *Drug Safety* 2006; 29(2): 161-168.
4. Amann C, Hasford J, Stausberg J. Stationäre Aufnahmen wegen unerwünschter Arzneimittelereignisse (UAE): Analyse der DRG-Statistik 2006. *Das Gesundheitswesen* 2012; 74(10): 639-644. German.
5. Hazell L, Shakir SA. Under-reporting of adverse drug reactions. *Drug Safety* 2006; 29(5): 385-396.
6. AGES Medizinmarktaufsicht [Internet]. Erstmeldungen von Angehörigen der Gesundheitsberufe. Vienna (AT): Bundesamt für Sicherheit im Gesundheitswesen; [revised 2014 Jan 24; cited 2014 Jan 26]. Available from <http://www.basg.gv.at/news%20-%20center/statistiken/anzweimittelsicherheit>. German.
7. Williams D, Feely J. Underreporting of adverse drug reactions: attitudes of Irish doctors. *Irish Journal of Medical Science* 1999; 168(4): 257-261.
8. Endel G [Internet]. Gesundheitssystemforschung in Österreich. Vienna (AT): Hauptverband der österreichischen Sozialversicherungsträger; 2011 Oct [cited 2014 Jan 21]. Available from [http://www.hauptverband.at/portal27/portal/hvbportal/channel\\_content/cmsWindow?action=2&p\\_me-nuid=71299&p\\_tabid=2&p\\_pubid=650904](http://www.hauptverband.at/portal27/portal/hvbportal/channel_content/cmsWindow?action=2&p_me-nuid=71299&p_tabid=2&p_pubid=650904). German.
9. Gall W, Dorda W, Durfts Schmid G, Endel G, Hronsky M, Neuhofer L, et al. Krankenhausaufenthalte infolge unerwünschter Arzneimittelereignisse. In: Ammenwerth E, Hörbst A, Hayn D, Schreier G, editors. Proceedings of eHealth2013; 2013 May 23–24, Vienna, Austria. Vienna: OCG; 2013. p. 31-36. German.
10. Stausberg J, Hasford J. Identification of adverse drug events: the use of ICD-10 coded diagnoses in routine hospital data. *Deutsches Arzteblatt International* 2010; 107(3): 23-29.
11. Österreichische Apotheker-Verlagsgesellschaft m.b.H [Internet]. Austria Codex 2006; 2006 [cited 2014 Jan 19]. Available from <http://www3.apoverlag.at/dynasite.cfm?dsmid=106234>. German.
12. WHO Collaborating Center for Drug Statistics Methodology [Internet]. ATC/DDD Index 2014; [cited 2014 Jan 19]. Available from [http://www.whocc.no/atc\\_ddd\\_index/](http://www.whocc.no/atc_ddd_index/).
13. Tricco AC, Pham B, Rawson NS. Manitoba and Saskatchewan administrative health care utilization databases are used differently to answer epidemiologic research questions. *Journal of clinical epidemiology* 2008; 61(2): 192-197.
14. Hoffmann F. Review on use of German health insurance medication claims data for epidemiological research. *Pharmacoepidemiology and drug safety* 2009; 18(5): 349-356.
15. Sansom L, Roughead E, Gilbert A, Primrose J. Coding drug-related admissions in medical records: is it adequate for monitoring the quality of medication use?. *Australian Journal of Hospital Pharmacy* 1998; 28(1): 7–12.