ELSEVIER

Computers in Biology and Medicine 36 (2006) 748–767

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance

Corneliu Henegar[a], Cédric Bousquet[a,b], Agnès Lillo-Le Louët[b], Patrice Degoulet[a], Marie-Christine Jaulent[a,*]

[a] *INSERM, U729, F-75006 Paris, France*
[b] *Centre Régional de Pharmacovigilance, Hôpital Européen Georges Pompidou, Paris, France*

Received 5 April 2005; accepted 5 April 2005

Abstract

Automated signal generation in pharmacovigilance implements unsupervised statistical machine learning techniques in order to discover unknown adverse drug reactions (ADR) in spontaneous reporting systems. The impact of the terminology used for coding ADRs has not been addressed previously. The Medical Dictionary for Regulatory Activities (MedDRA) used worldwide in pharmacovigilance cases does not provide formal definitions of terms. We have built an ontology of ADRs to describe semantics of MedDRA terms. Ontological subsumption and approximate matching inferences allow a better grouping of medically related conditions. Signal generation performances are significantly improved but time consumption related to modelization remains very important.

Keywords: Adverse drug reaction reporting systems; Terminology; Automatic data processing; Knowledge representation (computer); Description logic; Ontological modeling

* Corresponding author. INSERM U729, Université Paris Descartes, Faculté de médecine, 15 rue de l'école de médecine 75006 Paris, France. Tel.: +33 1 42 34 69 86; fax: +33 1 53 10 92 01.
E-mail address: marie-christine.jaulent@spim.jussieu.fr (M. Jaulent).

0010-4825/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2005.04.009

1. Introduction

1.1. Medical context: signal generation in pharmacovigilance

The World Health Organization (WHO) defines a signal in pharmacovigilance as "any reported information on a possible causal relationship between a drug and an adverse drug reaction (ADR), the relationship being unknown or incompletely documented previously" [1]. A continuous review of all reported drug–ADR combinations is needed in order to detect serious or unexpected events, which constitute the main objective of any pharmacovigilance reporting system. Traditionally, analysis is carried out by a systematic manual expert review of spontaneous reports sent by physicians and registered in pharmacovigilance database systems.

Qualitative review by experts of all reported drug–ADR combinations is becoming increasingly difficult because of the constant raise in the number of cases stored in national and international databases. Moreover, the continuous development of new drugs requires an early detection of their unknown adverse effects.

Drawing expert's attention on relevant combinations [2] is becoming necessary. During the last 5 years, automated signal detection methods have been developed to supplement qualitative clinical methods [1,3–5]. While these automated methods cannot replace expert clinical reviewers, they can provide assistance with the difficult task of screening huge numbers of drug–ADR combinations in databases for potential signals. Commonly used methods search databases for significant occurrence disproportionalities. Dependencies between drug–ADR pairs are based on an underlying model of statistical association [6].

1.2. Statistical approaches for signal generation

Techniques based on Bayesian methods were proposed for the exploitation of huge pharmacovigilance databases in the WHO Uppsala Center for drug safety reports [7] and the Food and Drug Administration pharmacovigilance database [8]. Proportionate Reporting Ratios and χ^2 test are implemented in the Medicines Control Agency [9] and the Drug Safety Research Unit [10] in Great Britain. Reporting odds ratios (RORs) are associated with logistic regression to test for confounding factors and the identification of drug–drug interactions and drug syndromes in the Netherlands [11].

However, to date, these methods have not been prospectively evaluated, as there is no true gold standard for signal detection [6]. Moreover, the number of cases needed to trigger a signal is not well defined. It is not usual to accept a signal when the number of cases is lower than three because the reporting of cases may be the result of a random process. Recently, Van Puijenbroeck assessed the performances of most commonly used signal detection methods on Dutch data by choosing the Bayesian method as reference. He showed that the main limitation for these methods was the small number of occurrences of each drug–ADR combination in the database [1].

Our working hypothesis was that performances of these quantitative methods could be ameliorated by grouping similar cases based on semantic information existing in the controlled vocabularies used to code ADR in case reports. Such an approach is susceptible to increase the number of occurrences of similar drug–ADR combinations in databases and thus ameliorate statistical relevance. Although the issue of terminology was quoted by several authors, no study was realized to test the choice of a system of coding ADRs on the performances of signal detection systems [12,6].