Drug Saf (2013) 36:573–582
DOI 10.1007/s40264-013-0064-4

ORIGINAL RESEARCH ARTICLE

Application of Information Retrieval Approaches to Case Classification in the Vaccine Adverse Event Reporting System

Taxiarchis Botsis · Emily Jane Woo · Robert Ball

Published online: 24 May 2013

Abstract
Background Automating the classification of adverse event reports is an important step to improve the efficiency of vaccine safety surveillance. Previously we showed it was possible to classify reports using features extracted from the text of the reports.
Objective The aim of this study was to use the information encoded in the Medical Dictionary for Regulatory Activities (MedDRA®) in the US Vaccine Adverse Event Reporting System (VAERS) to support and evaluate two classification approaches: a multiple information retrieval strategy and a rule-based approach. To evaluate the performance of these approaches, we selected the conditions of anaphylaxis and Guillain–Barré syndrome (GBS).
Methods We used MedDRA® Preferred Terms stored in the VAERS, and two standardized medical terminologies: the Brighton Collaboration (BC) case definitions and Standardized MedDRA® Queries (SMQ) to classify two sets of reports for GBS and anaphylaxis. Two approaches were used: (i) the rule-based instruments that are available by the two terminologies (the Automatic Brighton Classification [ABC] tool and the SMQ algorithms); and (ii) the vector space model.
Results We found that the rule-based instruments, particularly the SMQ algorithms, achieved a high degree of specificity; however, there was a cost in terms of sensitivity in all but the narrow GBS SMQ algorithm that outperformed the remaining approaches (sensitivity in the testing set was equal to 99.06 % for this algorithm vs. 93.40 % for the vector space model). In the case of anaphylaxis, the vector space model achieved higher sensitivity compared with the best values of both the ABC tool and the SMQ algorithms in the testing set (86.44 % vs. 64.11 % and 52.54 %, respectively).
Conclusions Our results showed the superiority of the vector space model over the existing rule-based approaches irrespective of the standardized medical knowledge represented by either the SMQ or the BC case definition. The vector space model might make automation of case definitions for spontaneous report review more efficient than current rule-based approaches, allowing more time for critical assessment and decision making by pharmacovigilance experts.

1 Introduction

The Vaccine Adverse Event Reporting System (VAERS) is a spontaneous reporting system (SRS) that collects reports of adverse events following immunization (AEFI) with vaccines licensed for use in the US [1]. There are two main safety surveillance purposes supported by VAERS: first, to identify any new and unexpected AEFIs; and second, to further characterize the safety profile of a vaccine by reviewing previously described AEFIs in the context of comorbidities and other medical products among the general population. While the latter requires a thorough cross-

Electronic supplementary material The online version of this article (doi:10.1007/s40264-013-0064-4) contains supplementary material, which is available to authorized users.

T. Botsis (✉) · E. J. Woo · R. Ball
Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research (CBER), US FDA, Woodmont Office Complex 1, Rm 306N, 1401 Rockville Pike, Rockville, MD 20852, USA
e-mail: Taxiarchis.Botsis@fda.hhs.gov

T. Botsis
Department of Computer Science, University of Tromsø, Tromsø, Norway

examination of various sources (e.g. medical records, epidemiological data), the former is highly related to the timely review of many incoming reports. The classification of reports according to established case definitions of AEFIs is highly prioritized in both cases, and automation of this process is important to improve timeliness, efficiency and consistency of safety surveillance.

The classification of cases based on the information extracted from within a large collection of documents lies in the field of information retrieval (IR) and deals mostly with unstructured (free text) information [2]. While research to apply natural language processing (NLP) techniques is underway to extract the appropriate information from the free text and support the text classification, i.e. the categorization of the topic or theme of a document [3], currently in VAERS and most other SRS databases, the encoding of the textual and other information in the reports is undertaken manually. This process is supported by medical terminology, namely the Medical Dictionary for Regulatory Activities (MedDRA®)[1] [4]. Current practice uses MedDRA® Preferred Terms (PTs) and Standardized MedDRA® Queries (SMQs) to identify cases of interest, followed by the application of a case definition, such as the Brighton Collaboration (BC) case definitions [5], in preparation for a case-series evaluation.

Several IR approaches have been developed in biomedicine. The common framework in IR is the use of medical terminologies and ontologies that have a general focus, such as the Unified Medical Language System [6], or a more specific one, such as the BioThesaurus [7] and the recently developed BioLexicon [8]. Subsequently, a number of steps are applied to extract the information of interest, which can be used for various purposes, including text classification [2, 3].

We have previously investigated the classification of VAERS reports by combining the BC case definition for anaphylaxis with medical expert knowledge to build a rule-based text mining algorithm [9]. Despite its high performance, the generalization of this strategy to the broad spectrum of AEFIs requires considerable effort. We therefore sought a more generalizable approach from the IR literature using a measure of semantic similarity [10]. An information theoretic similarity measure [11] has been used to find similar cases in electronic medical records within a case-based reasoning framework [12, 13]. We approached the problem of case classification as one of document retrieval in which the document query vector is constructed from key words in the case definition and applied the commonly used cosine similarity measure to find likely matches [2]. We demonstrated that it is possible to classify reports for anaphylaxis by applying this approach to key features extracted from VAERS texts and the BC case definition for anaphylaxis [14]. However, how applying the vector space model with the extracted text (using NLP) would compare to the current, widely available standard for encoding information (i.e. MedDRA®) in terms of case classification has yet to be examined. We also wanted to extend this approach to another case definition. Thus, we explored alternative solutions by using the MedDRA® PTs in VAERS to support two classification approaches: the vector space model and a rule-based approach. To evaluate the performance of these approaches, we selected the conditions of anaphylaxis with which we previously worked and Guillain–Barré syndrome (GBS) to represent a clinical condition with different characteristics than anaphylaxis and a more difficult case classification problem.

1.1 MedDRA® and Standardized MedDRA® Queries (SMQs)

VAERS reports contain both structured (e.g. demographics) and unstructured data (e.g. symptom text). Data-entry personnel apply MedDRA® PTs to the text in the reports according to coding conventions and algorithms [15]. The PTs are not considered to be medically confirmed diagnoses (i.e. physicians have not verified the diagnoses based on a review of medical records and elimination of alternative explanations) but are used in the review process, primarily as search terms for the selection of reports of potential interest and the identification of data aberrations using statistical data mining procedures, all in support of potential safety signal identification. Moreover, MedDRA® groups the PTs into SMQs to aid in case identification [16] and improve the sensitivity and specificity of search strategies; thus, SMQs represent a type of case classification algorithm. In this study we used the information represented by MedDRA® PTs to explore the performance of the BC rule-based automated classification, the Automatic Brighton Classification (ABC) tool (see Sect. 1.2), and the IR-based strategy for the case definitions of anaphylaxis and GBS [17, 18], comparing them with the anaphylaxis and GBS SMQs as the current standards for case identification in SRSs.

1.2 Brighton Collaboration (BC) Case Definitions and Automatic Brighton Classification (ABC) Tool

The BC (https://brightoncollaboration.org) develops standardized and globally accepted case definitions for a number of AEFIs [5]. It also offers to its members an online tool (namely the 'ABC' tool), which allows the confirmation of AEFIs based on user input for the specific criteria of the corresponding case definitions. For instance,

[1] MedDRA® terminology is the international medical terminology developed under the auspices of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH).

for anaphylaxis and GBS the user is asked to respond to 34 and 10 questions, respectively, by selecting one of three potential answers: 'Yes', 'No' and 'Don't Know'. Subsequently, the ABC tool processes user input and, unless there is insufficient evidence (in this case the ABC tool marks the report as having 'Insufficient evidence'), classifies a spontaneous report as 'Not a case' (when the criteria are not met) or into certain diagnostic levels (Level 1, 2 or 3) as defined by the BC case definitions. When applied to data, BC definitions may be operationalized as MedDRA® PTs or other adverse reaction terms; in some cases, BC definitions are developed based on prior knowledge of existing MedDRA® PTs and/or SMQs.

It should be noted that in the case of anaphylaxis two criteria must be initially satisfied to allow the consideration of the remaining criteria: (i) the 'Rapid progression of signs and symptoms'; and (ii) the 'Sudden onset of signs and symptoms'—if the user selects 'Don't know' the ABC tool will return a notification for 'insufficient evidence'. An analogous condition must be satisfied in the case of GBS. Specifically, any report will be classified as 'Not a case' if either of the following two criteria is not met: (i) the 'monophasic illness pattern'; and (ii) the 'interval between onset and nadir of weakness between 12 hours and 28 days followed by subsequent clinical plateau'; both are necessary for all levels of diagnostic certainty. However, if the user selects 'Don't know' the ABC tool will conditionally classify the report to one or more diagnostic levels based on the remaining criteria.

2 Materials and Methods

In this study we used the MedDRA® PTs (version 14.1) for each VAERS report to evaluate the contribution of both the BC case definition and the SMQ to the classification of spontaneous reports for the two AEFIs (anaphylaxis and GBS). Three directions were followed using two sets of reports, one for anaphylaxis and one for GBS (Fig. 1): (i) we measured the performance of the ABC tool; (ii) we used the SMQ algorithms that have been proposed by the MedDRA® working groups for the identification of reports [19]; and (iii) we applied an IR-based strategy to the same sets using the medical terms included in either the BC case definitions or the SMQs. The first two directions represent our rule-based approach (Fig. 1).

2.1 Sample

2.1.1 Anaphylaxis Sample

In the case of anaphylaxis we used the same (as in our previous study) training and testing sets that were randomly formed from 6034 VAERS reports for the H1N1 vaccine from 22 November 2009 to 31 January 2010 [9]. Based on the BC case definition for anaphylaxis (see Appendix 1 [Online Resource 1]), these reports had been previously classified by the FDA medical experts as potentially positive or negative for anaphylaxis.

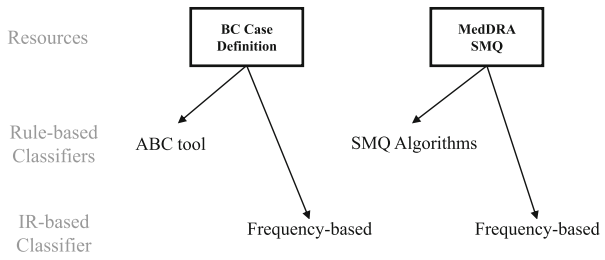

Fig. 1 Resources and classification approaches. *ABC* Automatic Brighton Classification, *BC* Brighton Collaboration case definition, *IR* information retrieval, *MedDRA*® Medical Dictionary for Regulatory Activities, *SMQ* Standardized MedDRA® Query

2.1.2 Guillain–Barré Syndrome Sample

In the case of GBS we searched all reports that were submitted to the VAERS from 17 January 2007 (the date that MedDRA® coding officially began in the VAERS) to 26 March 2012 (the date we started our analysis) for three types of cases: (i) those that were likely to be GBS (so we would have sufficient numbers to evaluate our procedures); (ii) reports with neurological signs and symptoms (to test the ability to distinguish GBS from other neurological conditions); and (iii) reports without neurological signs and symptoms. The search strategies for the three subgroups were as follows:

- GBS and the synonym PTs: 'Acute polyneuropathy', 'Chronic inflammatory demyelinating polyradiculoneuropathy', 'Guillain–Barré syndrome' and 'Miller Fisher syndrome' (group of GBS events, N = 1291; 500 reports randomly selected);
- any PT from the System Organ Class (SOC) of *Nervous System Disorders* (other than the four GBS terms mentioned above) [group of non-GBS neurological events, N = 55,363; 250 reports randomly selected];
- no PTs relating to GBS or other nervous system disorders (group of non-neurological events, N = 112,886; 250 reports randomly selected).

Henceforth, an experienced medical officer (EJW, one of the co-authors) reviewed and classified the 1000 reports into suspicious versus not likely to be GBS by reading the full VAERS report using the BC case definition as a guide with the caveats described in the next paragraph [18]; both subsets were randomly split into a training and a testing set following the 75–25 % rule as in the anaphylaxis case [9].

Of note, the BC case definition of GBS includes strict clinical and laboratory criteria (see Appendix 2 [Online Resource 1]). In our experience, many initial VAERS reports do not include information about the nadir, reflexes, electromyography and other laboratory details required for a BC level 1–3 classification, but based on follow-up information are later found—based on expert review by neurologists and other clinicians—to be GBS. Therefore, the reports in the aforementioned three groups that (i) included only some of the clinical and/or laboratory findings (particularly if they described classic and characteristic symptoms, such as rapidly progressive ascending weakness and tingling), or (ii) described neurological symptoms and treatment with intravenous immunoglobulins, plasmapheresis and/or mechanical ventilation, or (iii) had an MD's diagnosis of GBS (i.e. medical records contained documentation of a diagnosis of GBS by an attending neurologist or other physician treating the patient, even with no description of symptom), were flagged as being 'suspicious' for GBS and earmarked for further evaluation. This is an important distinction in the use of the BC definition for day-to-day vaccine safety surveillance and formal epidemiological evaluation. We made this adaptation to the BC case definition because the ability to automatically classify reports as 'suspicious' adds a substantial benefit to the efficiency of surveillance operations, while minimizing the number of reports likely to be GBS that would not be selected for further review.

2.2 Information Retrieval-Based Classification

2.2.1 Adverse Events Following Immunization Queries

The actual step to recognize and verify a pattern in an SRS such as the VAERS is the case-series analysis. During case-series analysis, an expert reviews the text describing the adverse event and looks for unusual patterns by analysing a set of features, such as demographics, time to onset, symptoms and diagnosis, and alternative explanations [14]. The primary task in case-series analysis is the extraction of the appropriate information by querying the spontaneous reports. Either the criteria of a BC case definition or the PTs of an SMQ could form a query to support the automated extraction of the required information and the subsequent classification of reports. Based on this assumption, we created two query versions for each of the AEFIs under study, i.e. anaphylaxis and GBS.

To create the first version of the queries, we used the complete subset of the PTs in MedDRA® terminology and built two lists with all the potential synonyms for the medical terms that were included in the BC criteria for anaphylaxis and GBS; both lists were enriched by the synonyms for the 'anaphylaxis' and 'GBS' terms, respectively. It should be noted that it was not possible to find PTs for all the BC criteria in the case definition of (i) anaphylaxis ('capillary refill time >3', 'mast cell tryptase elevation >upper normal limit', 'rapid progression of signs and symptoms', 'recessions' and 'sudden onset of signs and symptoms'), and (ii) GBS ('illness pattern, monophasic' and 'interval between onset and nadir of weakness between 12 hours and 28 days followed by subsequent clinical plateau'). Subsequently, an experienced medical expert (RB, one of the co-authors) reviewed and finalized the lists that are included in Appendix 3 (Online Resource 1). Some of the choices are subjective and might be different for another expert. Further work might be fruitfully undertaken to optimize these lists.

In the case of anaphylaxis (see Appendix 3 [Online Resource 1]), the medical expert excluded items that suggested an alternative cause, e.g. 'milk allergy'; neutral statements, e.g. 'blood pressure'; specific references to vaccine injection site findings, 'injection site erythema'; terms with specific references to treatments, e.g. 'blood pressure inadequately controlled'; conditions that specifically referenced the genitalia or perineum; items that had clear references to other body parts than the one in question, e.g. 'breast swelling' is not a synonym for 'upper airway swelling'.

In the case of GBS, the BC case definition is structured in such a way that the main clinical features of weakness/paralysis and diminished/absent reflexes account for the main terms of interest; there are MedDRA® PTs for these conditions. The laboratory (cerebrospinal protein and white blood cell count) and electromyography /nerve conduction studies likewise have specific, although not exhaustive, MedDRA® PTs. Because Fisher syndrome involves cranial nerve findings, all MedDRA® PTs involving cranial nerve abnormalities were also included. Other neurological disorders might also involve these clinical or laboratory features, so the diagnosis of GBS relies on the particular pattern with which these conditions occur. Some of these aspects (e.g. nadir) cannot be coded with MedDRA® PTs. All other MedDRA® PTs representing neurological signs and symptoms were considered not to contribute to GBS.

The second version of the queries was based on the SMQs and included all the corresponding MedDRA® PTs per AEFI. The SMQ was not explicitly compared with the list developed to represent the ABC tool, although there is substantial overlap.

2.2.2 Vector Space Model

According to the IR theory, it is possible to represent both the reports and the above queries as vectors with components made of the corresponding medical terms. Then, a score can be assigned to each report by calculating the

cosine similarity of the report with the query vector [2]; the mathematical details of the calculations are included in Appendix 4 (Online Resource 1). After calculating the scores for all reports in the anaphylaxis and GBS set, we used the scores of the training set as the predictors and the medical experts' classification as the gold standard to build the receiver operating characteristic (ROC) curve; specificity, sensitivity and area under the curve (AUC) were also calculated for the best cutoff point. The scores of the testing set were used to evaluate the performance of this approach.

2.3 ABC Tool Classification

The first version of the queries that were described above (see Sect. 2.2.1) should be considered comparable, however not equivalent, to the ABC tool, considering not only the aforementioned inexistence of PTs to satisfy the specific criteria of the BC case definitions that were included in the ABC tool but also the existence of specific BC medical terms that were missing from the ABC tool. Particularly, four ('respiratory distress', 'reduced peripheral circulation', 'uncompensated shock' and 'anaphylaxis') and one ('GBS related terms') terms were missing from the anaphylaxis and GBS ABC tool, respectively. Using the same list of MedDRA® PTs that was created for the IR-based approach (see Appendix 3 [Online Resource 1]), we collected all the appropriate information from the anaphylaxis (N = 6034) and GBS (N = 1000) datasets in two databases that were further processed by the ABC tool.

To treat the inexistence of MedDRA® PTs for the two key criteria in the BC case definition for anaphylaxis ('rapid progression of signs and symptoms' and 'sudden onset of signs and symptoms') as well as for the corresponding ABC tool questions, we hypothesized that the response to these questions was positive (i.e. 'Yes'); otherwise all 6034 reports would be classified as having 'insufficient evidence'. A report was considered positive when it was classified as 'Level 1', 'Level 2' and 'Level 3', and negative when it was classified as 'not a case' or having 'insufficient evidence'.

The inexistence of the aforementioned key GBS criteria (i.e. 'illness pattern, monophasic' and 'interval between onset and nadir of weakness between 12 hours and 28 days followed by subsequent clinical plateau') was treated differently. To allow the conditional classification of reports that did not include those criteria, we did not specify them as 'negative' and responded to the corresponding question of the ABC tool as 'Don't Know'.

In both anaphylaxis and GBS, ABC tool classification was compared with the gold standard to calculate sensitivity, specificity and AUC.

2.4 SMQ Algorithms

The MedDRA® working groups have developed algorithms to combine a number of symptoms to increase the specificity per AEFI [19]. In the case of anaphylaxis (three SMQ algorithms), a case must include either:

- a narrow term or a term from Category A (e.g. 'anaphylactic reaction') OR
- a term from Category B (e.g. 'asthma') AND a term from Category C (e.g. 'angioedema') OR
- a term from Category D (e.g. 'hypotension') AND [a term from Category B (e.g. 'stridor') OR a term from Category C (e.g. 'cyanosis')].

In GBS (four SMQ algorithms), cases to be selected for further review should include any cases meeting any one of the criteria listed below:

- at least one of the PTs listed for Category A (narrow scope—e.g. 'Guillain–Barré syndrome') OR
- any case reporting at least two PTs from Category B (e.g. 'cranial neuropathy' and 'neuropathy') OR
- any case reporting at least one PT from Category B (e.g. 'cranial neuropathy') and at least one PT from Category C (e.g. 'areflexia') OR
- any case reporting at least one PT each from Categories B (e.g. 'cranial neuropathy'), C (e.g. 'paralysis') and D (e.g. 'abasia').

Using the SMQ algorithms, we classified the reports as positive or negative for anaphylaxis and GBS. As in the ABC tool case, the classification per SMQ algorithm was compared with the gold standard to calculate sensitivity, specificity and AUC. We also evaluated the classification of reports combining the three and four SMQ algorithms in anaphylaxis and GBS, respectively. In other words, a case was classified as positive when either of the algorithms was satisfied, e.g. 1A or (1B + 1C) or (1D + 1B) or (1D + 1C) in anaphylaxis.

In this study, R was used for plotting the ROC curves and calculating the statistics.

3 Results

Table 1 presents the categorization of reports based on the encoded information meeting either the BC case definition or the SMQ criteria for the two conditions. Those reports that did not include any PTs meeting the corresponding BC criteria were marked as having 'no evidence' by the ABC tool in the case of anaphylaxis (N = 2373; Table 2). The remaining anaphylaxis reports were classified as positive (N = 329: N_{level_1} = 101, N_{level_2} = 221 and N_{level_3} = 7; Table 2) and negative (N = 3332: $N_{Not\_a\_case}$ = 2844 and

Table 1 The distribution of reports with and without MedDRA® PTs for the two standardized case definitions

	Anaphylaxis		Guillain–Barré syndrome	
	PTs present	PTs absent	PTs present	PTs absent
BC	3661	2373	531	469
SMQ	2532	3502	585	415

BC Brighton Collaboration case definition, *IR* information retrieval, *MedDRA®* Medical Dictionary for Regulatory Activities, *PTs* Preferred Terms, *SMQ* Standardized MedDRA® Query

$N_{Insufficient\_evidence}$ = 488; Table 2) based on the categorization that was described above (see Sect. 2.3). Interestingly, the ABC tool classified all the GBS reports as 'not a case' (N = 963) or having 'insufficient evidence' (N = 37; Table 2). It should be mentioned that the 37 GBS reports with 'insufficient evidence' were conditionally classified to the three diagnostic levels indicating the additional criteria that would be required to fully meet the respective level, i.e. the 'monophasic illness pattern' and the 'interval between onset and nadir of weakness between 12 h and 28 days followed by subsequent clinical plateau' that were initially marked as 'Don't know' in the corresponding ABC tool questions. Subsequently, the performance of the ABC tool to identify the positive and negative reports was evaluated both in the training and testing set per AEFI versus the gold standard (Tables 3, 4). In the case of anaphylaxis, the ABC tool's performance was excellent in terms of specificity (96.76 %), but only fair in terms of sensitivity (64.11 %). Given the conditional classification of reports into the diagnostic levels of the BC case definition for GBS, we did not evaluate the corresponding performance of the ABC tool.

The classification of reports in the anaphylaxis and GBS training sets using the frequency-based cosine similarity is illustrated in Figs. 2 and 3. At the best cutoff point of the ROC curves (marked with a dot, Figs. 2, 3) we calculated sensitivity and specificity for both the training and testing sets in anaphylaxis (Table 3) and GBS (Table 4). Thus, in the anaphylaxis and GBS testing sets sensitivity and specificity were equal to 84.75 % and 89.62 %, and 85.65 % and 95.14 %, respectively.

As expected, the SMQ algorithms did achieve high specificity for anaphylaxis and GBS. Interestingly, the narrow algorithm for GBS (at least one of the PTs falling in category A, as described in Sect. 2) classified the reports with high specificity (97.49 % and 99.06 % in the training and testing sets, respectively). These excellent figures should be attributed to the original synthesis of our sample that included 500 reports having one of the terms in category A; it was almost identical to the set of reports that were classified by the medical expert as possible GBS cases. In terms of sensitivity, the combined SMQ algorithm performed slightly better (54.24 % vs. 52.54 %) or equally (99.06 % vs. 99.06 %) to the best-performing SMQ category in anaphylaxis (1B + 1C) and GBS (1A), respectively; on the other hand, specificity was lower in all cases (Tables 3, 4).

4 Discussion

To the best of our knowledge, this is the first study to evaluate the use of MedDRA® PTs as input into formal case definitions using both two rule-based approaches (ABC tool and SMQ algorithms) and an IR strategy (vector space model) for the purposes of case classification in an SRS. In this study, we used the encoded information in VAERS as well as standardized medical knowledge to categorize a large set of reports for anaphylaxis and GBS. Three paths were followed to combine these sources: an existing classifier (ABC tool), an IR-based approach and a number of SMQ categorization algorithms. In the case of anaphylaxis, we demonstrated that the IR-based approach performed significantly better compared with both the ABC tool and the SMQ algorithms in terms of sensitivity (IR-based sensitivity equal to 86.44 % vs. 64.11 % and 52.54 %, respectively); the lower specificity could be corrected by selecting a different cutoff point other than the one marked in Fig. 2. In the case of GBS, sensitivity was still higher in all but one case, which, as aforementioned (see Sect. 2.1.2), should be attributed to the selection process in which 500 reports included at least one GBS or GBS-related term. These results showed the superiority of

Table 2 Classification using the ABC tool

	Positive			Negative		
	Level 1	Level 2	Level 3	Not a case	Insufficient evidence	No evidence[a]
Anaphylaxis	101	221	7	2844	488	2373
Guillain–Barré syndrome	(1)[b]	(13)[b]	(37)[b]	963	37	0

ABC Automatic Brighton Classification, *BC* Brighton Collaboration case definition

[a] No BC criteria were identified in these reports and, thus, were not processed by the ABC tool

[b] These reports were conditionally classified to this diagnostic level and were marked as having 'insufficient evidence'

Table 3 Anaphylaxis classification

		Sensitivity (95 % CI)	Specificity (95 % CI)	AUC (95 % CI)
Training set (N = 4526)				
BC				
IR approach	F-b	0.8427 (0.7820–0.8889)	0.7861 (0.7737–0.980)	0.8144 (0.8123–0.8165)
ABC tool		0.5618 (0.4884–0.6326)	0.9669 (0.9611–0.9718)	0.7643 (0.7609–0.7678)
SMQ				
IR approach	F-b	0.8315 (0.7696–0.8793)	0.8528 (0.8420–0.8630)	0.8421 (0.8404–0.8439)
SMQ categories	1A	0.0562 (0.0308–0.1003)	0.9989 (0.9973–0.9995)	0.5275 (0.5201–0.5349)
	1B + 1C	0.6236 (0.5505–0.6915)	0.9738 (0.9686–0.9781)	0.7987 (0.7957–0.8016)
	1D + 1B	0.0225 (0.0088–0.0563)	0.9975 (0.9955–0.9986)	0.5100 (0.5023–0.5176)
	1D + 1C	0.0169 (0.0057–0.0484)	0.9995 (0.9983–0.9999)	0.5082 (0.5005–0.5159)
	Combined	0.6404 (0.5677–0.7073)	0.9703 (0.9649–0.9750)	0.8054 (0.8026–0.8082)
Testing set (N = 1508)				
BC				
IR approach	F-b	0.8644 (0.7546–0.9297)	0.7861 (0.7642–0.8064)	NA
ABC tool		0.6411 (0.5166–0.7540)	0.9676 (0.9571–0.9755)	NA
SMQ				
IR approach	F-b	0.8475 (0.7348–0.9176)	0.8565 (0.8375–0.8736)	NA
SMQ categories	1A	0.0339 (0.0093–0.1154)	0.9979 (0.9939–0.9993)	NA
	1B + 1C	0.5254 (0.4004–0.6473)	0.9793 (0.9706–0.9855)	NA
	1D + 1B	0 (0–0.0611)	0.9945 (0.9891–0.9972)	NA
	1D + 1C	0 (0–0.0611)	0.9979 (0.9939–0.9993)	NA
	Combined	0.5424 (0.4166–0.6630)	0.9724 (0.9626–0.9797)	NA

F-b Frequency-based, *ABC* Automatic Brighton Classification, *AUC* area under the concentration–time curve, *BC* Brighton Collaboration case definition, *IR* information retrieval, *MedDRA®* Medical Dictionary for Regulatory Activities, *NA* not applicable, *SMQ* Standardized MedDRA® Query

the vector space model over the existing rule-based approaches irrespective of the standardized medical knowledge represented by either the SMQ or the BC case definition.

The main limitation of our study is the use of the encoded information, and not the textual information, in VAERS. Using the encoded information only, it has not been possible to identify MedDRA® PTs for all the criteria in the BC case definition. Therefore, one could argue that the inability of MedDRA® to fulfill all criteria, such as 'recessions' for anaphylaxis and 'illness pattern, monophasic' for GBS, might have resulted in considerable information loss. On the other hand, other criteria ('capillary refill time >3' and 'mast cell tryptase elevation >upper normal limit' for anaphylaxis; 'interval between onset and nadir of weakness between 12 h and 28 days followed by subsequent clinical plateau' for GBS) never or less frequently appear ('rapid progression of signs and symptoms' and 'sudden onset of signs and symptoms' for anaphylaxis; 'electrophysiologic findings' for GBS) in spontaneous reports. Even though the assumption behind text mining is the acquisition of more granular information, it appears that our classification approaches performed surprisingly well based on the encoded information (i.e. MedDRA® PTs). As a side benefit, identification of potentially important omissions in MedDRA® could lead to enhancements of that system. MedDRA® is continuously updated based on suggestions from subscribers and review by the MedDRA® Management Board [20].

A second limitation is related to the use of MedDRA® PTs in GBS that may be listed under more than one SOC. In our medical judgement, some of the PTs with a secondary listing under 'Nervous System' have only marginal neurological relevance (e.g. *astrocytoma malignant,* which is more usefully classified under 'Neoplasms Benign, Malignant, and Unspecified'; and *carotid artery dissection,* which is more germanely categorized under 'Vascular Disorders'). However, we decided by consensus to accept the potential limitations resulting from the overlapping categorization, i.e. we allowed the use of both primary and secondary listings and did not modify SOCs to make them mutually exclusive.

The two standardized medical terminologies used for classification purposes in safety surveillance, i.e. the BC case definition and the SMQ, appeared to have several important differences. First, the key BC criteria could not

Table 4 Guillain–Barré syndrome classification

		Sensitivity (95 % CI)	Specificity (95 % CI)	AUC (95 % CI)
Training set (N = 750)				
BC				
IR approach	F-b	0.8777 (0.8372–0.9093)	0.9443 (0.9185–0.9623)	0.9110 (0.9103–0.9118)
ABC tool		NA	NA	NA
SMQ				
IR approach	F-b	0.9624 (0.9354–0.9784)	0.8144 (0.7750–0.8483)	0.8884 (0.8874–0.8894)
SMQ categories	1A	0.9749 (0.9513–0.9872)	0.8770 (0.8427–0.9047)	0.9260 (0.9523–0.9267)
	2B	0.4608 (0.4069–0.5157)	0.9652 (0.9434–0.9788)	0.7130 (0.7101–0.7159)
	1B + 1C	0.6458 (0.5918–0.6962)	0.9582 (0.9350–0.9734)	0.8020 (0.8001–0.8039)
	1B + 1C + 1D	0.5737 (0.5188–0.6267)	0.9652 (0.9434–0.9788)	0.7694 (0.7671–0.7717)
	Combined	0.9969 (0.9825–0.9994)	0.8485 (0.8073–0.8757)	0.9207 (0.9199–0.9215)
Testing set (N = 250)				
BC				
IR approach	F-b	0.8962 (0.8237–0.9411)	0.9514 (0.9031–0.9763)	NA
ABC tool		NA	NA	NA
SMQ				
IR approach	F-b	0.9340 (0.8699–0.9676)	0.8056 (0.7333–0.8619)	NA
SMQ categories	1A	0.9906 (0.9485–0.9983)	0.9028 (0.8434–0.9412)	NA
	2B	0.3585 (0.2736–0.4533)	0.9583 (0.9121–0.9808)	NA
	1B + 1C	0.6698 (0.5757–0.7520)	0.9861 (0.9508–0.9962)	NA
	1B + 1C + 1D	0.6321 (0.5372–0.7178)	0.9861 (0.9508–0.9962)	NA
	Combined	0.9906 (0.9485–0.9983)	0.8819 (0.8191–0.9250)	NA

F-b Frequency-based, *ABC* Automatic Brighton Classification, *AUC* area under the concentration-time curve, *BC* Brighton Collaboration case definition, *IR* information retrieval; *MedDRA®* Medical Dictionary for Regulatory Activities, *NA* not applicable, *SMQ* Standardized MedDRA® Query

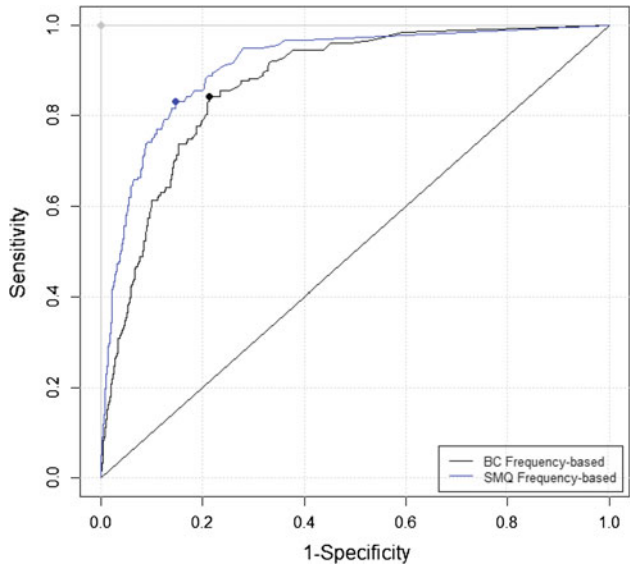

Fig. 2 Receiver operating characteristic curve in the anaphylaxis training set. *BC* Brighton Collaboration case definition, *MedDRA®* Medical Dictionary for Regulatory Activities, *SMQ* Standardized MedDRA® Query

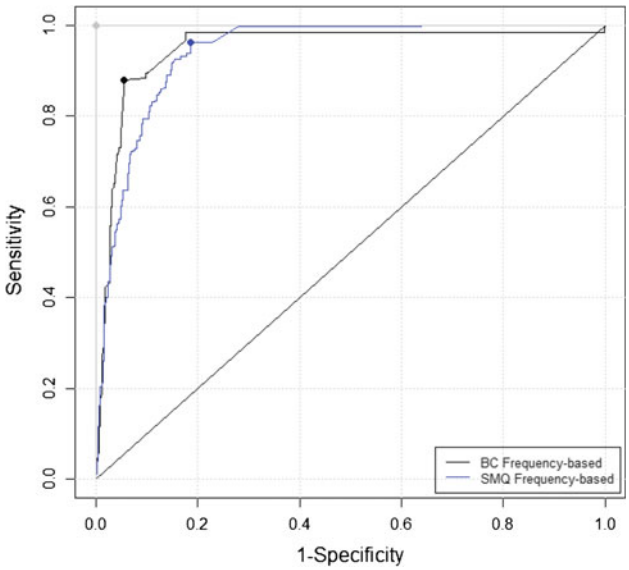

Fig. 3 Receiver operating characteristic curve in the Guillain–Barré syndrome training set. *BC* Brighton Collaboration case definition, *MedDRA®* Medical Dictionary for Regulatory Activities, *SMQ* Standardized MedDRA® Query

be mapped to MedDRA® PTs and, consequently, to an SMQ. Second, certain SMQ terms did not appear in the BC case definition, such as 'asthma' and 'nasal obstruction' in anaphylaxis. Third, in contrast to the SMQs, BC definitions did not include the 'anaphylaxis' and 'GBS' terms or their synonyms. Despite these differences, both instruments supported equally the classification of reports when coupled with the IR-based approach.

So what is the value of automated case classification for pharmacovigilance practice and what are the implications of our findings about the performance of the IR approach using MedDRA® terms? While the uptake of statistical data mining methods over the past decade has undoubtedly improved the efficiency of spontaneous report evaluation, the case-series evaluation remains the cornerstone of pharmacovigilance practice and is still largely dependent on human expert review of reports [21]. The first step in this often laborious process is the organization of clinical data and, when possible, the application of a formal case definition. Previously we showed it was possible to automate the rule-based BC anaphylaxis definition using features extracted from the narrative text [9]. But text-mining tools are not widely available and rule-based approaches require complex computer programming. Spontaneous reports usually have less information than a complete medical record but because our initial goal is simply to reduce the number of reports requiring expert review in a screening step, a less stringent approach can still provide important efficiency gains. In this study we demonstrate that we can obtain similar, if not superior, performance in identifying possible anaphylaxis and GBS cases using widely available MedDRA® codes and a generic IR algorithm.

Of course further validation with other conditions and comparisons with text-mining approaches are necessary, but this work shows that, in principle, this same approach might be applied to other conditions so that the automation of case classification for spontaneous report review might be a less complex process than anticipated. Such tools could improve the efficiency of report classification, making more time available for critical assessment and decision making by pharmacovigilance experts.

5 Conclusions

SMQs and BC case definitions are the standardized medical terminologies that are typically used in the case-series analysis. Using the MedDRA® PTs in the VAERS we demonstrated that the vector space model might support report classification better than the existing rule-based approaches (either the ABC tool or the SMQ categories). Depending on the availability of the appropriate terminologies, the same IR strategies can be applied to postmarketing surveillance of drugs and other medical products. We believe that our analysis may stimulate further research. Specifically, it raises two main questions: (i) what is the most efficient way to translate the existing medical terminologies into machine-readable algorithms and bridge them to either the MedDRA® PTs or textual information in spontaneous reports; (ii) how should we generate formal case definitions to consistently and efficiently perform case-series analyses for the universe of adverse events? These questions relate to our long-term goal of automating case classification and enhancing safety surveillance of medical products. Many drugs and biologics have novel characteristics and unique safety profiles, and may be used in populations with risk factors for specific AEFIs. We hope that further refinement of IR methods will improve our ability to review and prioritize safety topics for both old and new products.

Acknowledgments We would like to thank Hector Izurieta at Office of Biostatistics and Epidemiology for the fruitful discussions, and Jan Bonhoeffer and Benus Becker at the BC for their guidance on how to use the ABC tool.

MedDRA® is a registered trademark owned by the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA) on behalf of the International Conference on Harmonization (ICH).

This project was supported in part by the appointment of Taxiarchis Botsis to the Research Participation Program at the Center for Biologics Evaluation and Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US FDA.

Author contributions Taxiarchis Botsis conceived the idea, performed the analysis and authored the paper; Emily Jane Woo read adverse event reports and determined the diagnostic level of certainty regarding GBS and edited the manuscript; Robert Ball led the overall effort in applying medical informatics to adverse event evaluation, defined the outcomes for the study, selected terms to represent the case definitions and edited the manuscript.

Competing interests Taxiarchis Botsis, Emily Jane Woo and Robert Ball have no conflicts of interest to declare that are directly relevant to the content of this study.

References

1. Varricchio F, Iskander J, Destefano F, Ball R, Pless R, Braun MM, et al. Understanding vaccine safety information from the vaccine adverse event reporting system. Pediatr Infect Dis J. 2004;23(4):287–94.
2. Manning CD, Raghavan P, Schutze H. Introduction to information retrieval. 1st ed. Cambridge: Cambridge University Press; 2008.
3. Manning CD, Schutze H. Foundations of statistical natural language processing. 1st ed. Cambridge: MIT Press; 1999.
4. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999;20(2):109–17.

5. Bonhoeffer J, Kohl K, Chen R, Duclos P, Heijbel H, Heininger U, et al. The Brighton Collaboration: addressing the need for standardized case definitions of adverse events following immunization (AEFI). Vaccine. 2002;21(3–4):298–302.
6. Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The unified medical language system. J Am Med Inform Assoc. 1998;5(1):1–11.
7. Liu H, Hu ZZ, Zhang J, Wu C. BioThesaurus: a web-based thesaurus of protein and gene names. Bioinformatics. 2006;22(1):103–5.
8. Thompson P, McNaught J, Montemagni S, Calzolari N, Del Gratta R, Lee V, et al. The BioLexicon: a large-scale terminological resource for biomedical text mining. BMC Bioinformatics. 2011;12(1):397.
9. Botsis T, Nguyen MD, Woo EJ, Markatou M, Ball R. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. J Am Med Inform Assoc. 2011;18(5):631–8.
10. Pedersen T, Pakhomov SV, Patwardhan S, Chute CG. Measures of semantic similarity and relatedness in the biomedical domain. J Biomed Inform. 2007;40(3):288–99.
11. Lin D. An information-theoretic definition of similarity. In: Proceedings of 15th international conference on machine learning. San Francisco: Morgan Kaufmann Publishers Inc.; 1998. p. 296–304.
12. Cao H, Melton GB, Markatou M, Hripcsak G. Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases. J Biomed Inform. 2008;41(6):882–8.
13. Markatou M, Kuruppumullage-Don P, Hu J, Wang F, Sun J, Sorrentino R, et al. Case-based reasoning in comparative effectiveness research. IBM J Res Dev. 2012;56:5.
14. Botsis T, Buttolph T, Nguyen MD, Winiecki S, Woo EJ, Ball R. Vaccine Adverse Event Text Mining (VaeTM) system for extracting features from vaccine safety reports. J Am Med Inform Assoc. 2012;19(6):1011–8.
15. Brown EG. Using MedDRA: implications for risk management. Drug Saf. 2004;27(8):591–602.
16. Mozzicato P. Standardised MedDRA queries: their role in signal detection. Drug Saf. 2007;30(7):617–9.
17. Ruggeberg JU, Gold MS, Bayas JM, Blum MD, Bonhoeffer J, Friedlander S, et al. Anaphylaxis: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine. 2007;25(31):5675–84.
18. Sejvar JJ, Kohl KS, Gidudu J, Amato A, Bakshi N, Baxter R, et al. Guillain–Barré syndrome and Fisher syndrome: case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2011;29(3):599–612.
19. MedDRA Maintenance and Support Services Organization. Introductory guide for standardised MedDRA queries (SMQs) Version 14.1. Chantily: MedDRA; 2011.
20. Medical Dictionary for Regulatory Activities: Maintenance and Support Services Organization. http://www.medramsso.com. Accessed 4 Apr 2013.
21. Food and Drug Administration. Guidance for industry: good pharmacovigilance practices and pharmacoepidemiologic assessment. US Department of Health and Human Services. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf. Accessed 4 Apr 2013.