Eur J Clin Pharmacol (2009) 65:729–741
DOI 10.1007/s00228-009-0638-3

PHARMACOEPIDEMIOLOGY AND PRESCRIPTION

A web-based quantitative signal detection system on adverse drug reaction in China

Chanjuan Li · Jielai Xia · Jianxiong Deng · Wenge Chen · Suzhen Wang · Jing Jiang · Guanquan Chen

Received: 13 August 2008 / Accepted: 11 February 2009 / Published online: 5 March 2009

Abstract
Objective To establish a web-based quantitative signal detection system for adverse drug reactions (ADRs) based on spontaneous reporting to the Guangdong province drug-monitoring database in China.
Methods Using Microsoft Visual Basic and Active Server Pages programming languages and SQL Server 2000, a web-based system with three software modules was programmed to perform data preparation and association detection, and to generate reports. Information component (IC), the internationally recognized measure of disproportionality for quantitative signal detection, was integrated into the system, and its capacity for signal detection was tested with ADR reports collected from 1 January 2002 to 30 June 2007 in Guangdong.
Results A total of 2,496 associations including known signals were mined from the test database. Signals (e.g., cefradine-induced hematuria) were found early by using the IC analysis. In addition, 291 drug-ADR associations were alerted for the first time in the second quarter of 2007.
Conclusions The system can be used for the detection of significant associations from the Guangdong drug-monitoring database and could be an extremely useful adjunct to the expert assessment of very large numbers of spontaneously reported ADRs for the first time in China.

Keywords Adverse drug reactions · Signal detection · Data mining · Spontaneous reporting system · Disproportionality · Pharmacovigilance

C. Li · J. Xia (✉)
Department of Health Statistics,
Fourth Military Medical University,
No 17, Changle West Road,
Xian 710032, China
e-mail: jielaixia@yahoo.com

J. Deng (✉) · J. Jiang
Adverse Drug Reaction Monitoring
Centre of Guangdong Province,
Guangzhou, China
e-mail: jianxiong_deng@163.com

W. Chen · G. Chen
Computer Centre, Guangdong University of Technology,
Guangzhou, China

S. Wang
Department of Health Statistics, Weifang Medical College,
Weifang, China

Introduction

Limited information concerning possible adverse drug reactions (ADRs) is available at the time of marketing. In recent decades, spontaneous reporting systems (SRSs), which are used to monitor the safety of drugs after marketing, have earned a reputation for offering a fast and reasonably efficient way to detect ADRs [1]. Manual review of all reports from SRSs is time consuming and difficult when attempting to establish more complex associations among patient characteristics, reported ADRs, and suspected drugs. In the last 10 years, several approaches have been developed and implemented to provide additional information concerning a possible relationship between a suspected ADR and a drug, such as proportional ADR reporting ratio (PRR) used by the Medicines and Healthcare Products Regulatory Agency, reporting odds ratio (ROR) used by the Netherlands Pharmacovigilance Foundation Lareb, Bayesian confidence propagation neural network (BCPNN) used by the WHO Collaborating Centre for International Drug Monitoring (the Uppsala Monitoring Centre), multi-item gamma-

Poisson shrinker, Yule's Q test, Poisson probability, and chi-square tests [2-9].

Approaches for signal detection from SRS have been developed and implemented quickly in other countries. The Guangdong ADR spontaneous reporting system was set up in 2001. The database of ADRs set up by the Guangdong Monitoring Centre covers the 21 cities in Guangdong province with a population over 82.9 million and contains 41,596 qualified reports from the beginning of 2002 through 30 June 2007. The database includes many events of multiple drugs and multiple reactions. It is a large database now, and about 3,000-5,000 new reports are added quarterly [10]. The task of trying to find new drug-ADR signals has been carried out by a committee of experts, but with such a large volume of material, the task is daunting. Automated quantitative signal detection systems, which can easily detect complex associations between the reported ADRs and suspected drugs, is urgently needed in Guangdong.

BCPNN Method

Although the various approaches differ from each other in their principles and methodology, they all share the same characteristic that they search the databases for disproportionality [9]. The statistical measures of disproportionality all express the extent to which the reported ADR is associated with the suspected drug compared with the other drugs in the database. The occurrence of ADRs related to other drugs in the database is used as a proxy for the background incidence of ADRs. Table 1 shows the 2×2 contingency table for calculating disproportionality [1]. N is the total number of records in the database, A is the number of combinations between a specific drug (*i*) and a suspected ADR (*j*), A+B is the total number of records on the suspected drug (*i*) in the database, and A+C is the total number of records on the suspected ADR (*j*) in the database.

BCPNN is a statistical neural network where the nodes correspond to different events and the weights between nodes are proportional to the strength of association between different events. For the purpose of dependency derivation, only the weights between nodes in the network (referred to as information components or IC values) are of interest [11]. These can be estimated directly from data, so for transparency we shall refer to the use of the BCPNN for Bayesian dependency derivation as IC analysis throughout this article.

For quantitative signal detection, no true gold standard is available; IC analysis was selected for use with the Guangdong database for the following reasons. Firstly, this approach yielded a positive predictive value of 44% and a negative predictive value of 85% in the detection of signals as compared with reference literature sources when used for analyzing the WHO database [9, 12, 13]. Secondly, in contrast to other measures, such as ROR and PRR, both the point estimate and its confidence interval can be calculated under all circumstances [9, 12, 13]. Furthermore, we had examined the concordance of various measures with the IC measure based on the database of Guangdong province [10]. Finally, we had constructed a spontaneous reporting system model considering several factors, such as frequency of exposure, time to market, the background incidence of adverse events, the severity of ADR as well as the reporting probability to compare various signal detection methods based on a simulation database by using SAS software. The results showed that the sensitivity and specificity of the IC measure were acceptable, the area under the receiver-operating-characteristic curve (AUC) was 0.87, and the evaluation indexes increased when the number of reports per combination increased.

IC and IC−2SD (i.e., IC minus two standard deviations) were calculated in the IC analysis for all drug-ADR combinations in the Guangdong database [9, 14]. A negative association in the Guangdong quantitative signal detection system was defined as IC-2SD≤0 [5]. A weak association was defined as 0<IC-2SD≤1.5, a medium association as 1.5<IC-2SD≤3.0, and a strong association as IC-2SD>3.0 [15]. Association refers to drug-ADR combinations with positive IC-2SD values, based on the calculation of the measure of disproportionality. "Signal detection" relates to the entire process of detecting a signal including manual clinical review.

Table 1 The 2×2 contingency table for calculation of disproportionality

	Records with the suspected ADR	Records without the suspected ADR	Total
Records with the suspected drug	A	B	A+B
All other records	C	D	C+D
Total	A+C	B+D	N=A+B+C+D

ADR Adverse drug reaction

System overview

The Guangdong quantitative signal detection system (GDQSDS), a web-based system comprising three software modules that prepare data, detect associations, and generate reports, was developed based on the Guangdong ADR monitoring platform. GDQSDS was coded with Microsoft Visual Basic and Active Server Pages programming

languages and programmed to carry out the IC analysis. All datasets were managed with a Microsoft SQL Server database.

GDQSDS runs on two separate servers using the Windows 2000 Server operating system. One is the World Wide Web (WWW) server running Internet Information Services. WWW users with different privileges, such as regulatory agencies, boards of experts, provincial users, manufacturers, and medical and health institutions, log on through the Secure Socket Layer and use the three modules of GDQSDS. The other server is the database server running SQL Server 2000, which can not be accessed from the Internet [16].

GDQSDS has large computational power to consider all possible links in the database and can automatically process the association detection procedures and output significant pharmacovigilance associations quarterly based on all drug-ADR combinations in the database. A total of 71 different fields, such as patient characteristics, generic drug name, ADR, drug category, and ADR category are included in the Guangdong ADR database. Searches can be performed for a specific condition, for a specific drug or ADR, or for subdatasets.

Figure 1 is the flowchart describing the signal detection system of Guangdong. All computer programs for data preparation (statistical description, data cleaning and coding, and data disintegration), association detection (data extraction and calculation), and report generation are integrated into the platform.

Statistical description

A total of 20,962 reports in males and 20,652 reports in females were collected in Guangdong (in 5 cases the gender variable was missing). The 19-29 and 30-39 age groups were most commonly found. The reports included 36,283 general ADR reports (87.18%), 3,818 (9.17%) new ADRs, 1,230 (2.96%) serious ADRs, and 288 (0.69%) new serious ADRs. The frequency of ADRs suffered after intravenous injection was 1.47 times, which is higher than the average of 1.36 times. Half of the 10 drugs most frequently reported (i.e., ceftriaxone sodium, glucose, cefoperazone sodium, sulbactam sodium, azithromycin, ceftazidime, penicillin G, cefuroxime sodium, levofloxacin hydrochloride, cefotaxime sodium, fleroxacin) were cephalosporins. A total of 2,835 ADRs were related to ceftriaxone sodium. Rash, anaphylactoid reaction, and rigors were the most frequently reported adverse reactions in the database. Report characteristics showed no clinical significant difference between doctor and pharmacist report groups.

The generic drug name or ADR name was incorrectly reported and could not be determined from the case description in 23 cases, so these cases were excluded from signal detection. Ultimately, 41,596 reports were collected during 1 January 2002 and 30 June 2007 in Guangdong Monitoring Centre.

Data cleaning and coding

In the Guangdong ADR monitoring database, some unusual, colloquial, and misspelled drug names and ADR names were found. For example, cefradine was recorded as cefradine, cefradine capsule, cefradine injection, and 53 other Chinese names. Leucopenia was recorded as leucopenia, white blood cell decrease, WBC decrease, white C decrease, and 22 other Chinese names. Data cleaning and coding of the drugs and ADRs should be done at the very beginning.

Generic names for drug records were standardized based on a combination of *New Pharmacology* (15th ed, Chinese) [17] and *MCDEX Clinical Drug Reference* (2006, Chinese) [18] and the Anatomical Therapeutic Chemical (ATC) classification was used as a reference for Western medicine. Generic drug names were used for the "drug of interest" in

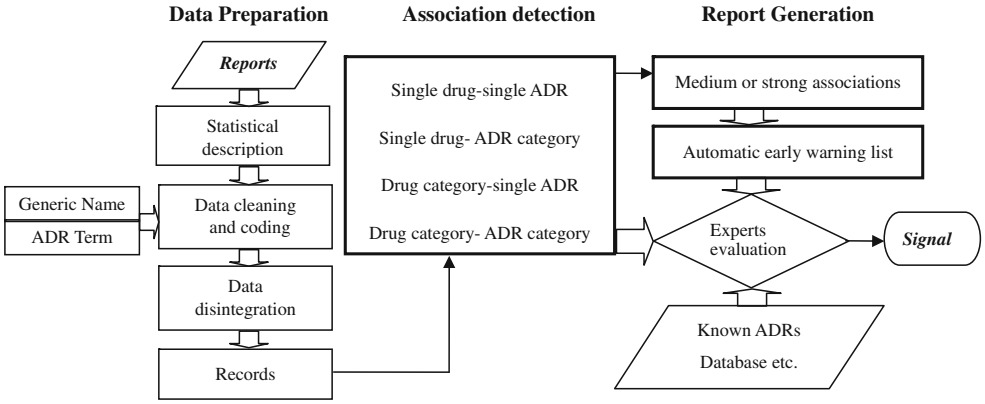

Fig. 1 Flowchart of signal detection system in Guangdong (*ADR* adverse drug reaction)

GDQSDS. Brand names were also considered in our generalized system. If one brand of a drug is linked to several reports in a short time, experts will pay more attention to this brand drug. ADR names were normalized using a combination of the WHO Adverse Reaction Terminology (Chinese) [19] and International Classification of Diseases, 10th revision (Chinese) [20].

The specific process for drug name standardization was as follows. (The process for ADR standardization was the same.) Firstly, we searched using a particular nonstandard drug name. Secondly, we renamed this drug based on standardization terminologies and added it to corresponding drug category. Thirdly, in the future, the same nonstandard drug names will be renamed too by automatic computer matching. The nonstandard drug name database, nonstandard ADR name database, standard drug name database, and standard ADR term (PT level) database were stored in the system. Through 30 June 2007, 5,407 rules for standardizing generic drug names and 4,031 standard drug names were created. A total of 8,715 rules for standardizing ADR names and 2,332 standard names of ADR were established. The rule databases were capable of completing 80% of the standardization work for the new cases. Only 20% of the reports needed to be standardized manually. In the future, we will train the reporters to report ADR using standardization databases.

Data disintegration

Reports sent to the SRS contain information about one or more (suspected) drugs and one or more suspected ADRs. For example, case A may have taken two suspected drugs and suffered three adverse reactions such as hematuria, leucopenia, and nausea. In that situation, this case will be disintegrated and the one report will be transformed into six records. In this way, 62,196 records were obtained from 41,596 reports. The reports database and disintegrated records database were both stored in the system. The disintegrated records can be linked back to the reports with a key field, the ID number. All calculations for signal detection in this article were performed based on 62,196 records.

Association detection

Subdatabases for analysis can be created by searching drug-related or ADR-related items such as generic name, ADR term, and date-received fields. Because the IC algorithm was programmed into the computer system, the A, B, C, D, N, and IC values can be calculated automatically by running SQL languages, and these results will be shown in tabular format. For example, the search for "cefradine-induced hematuria" collected from 2002 to 2004 would be conducted as follows. Firstly, the records database can be searched by entering 2002 to 2004 in the "date-received field." N is the total number of records in the selected database. Secondly, by searching the "generic name field" for cefradine and the "ADR term field" for hematuria, we can obtain the value for A. Calculations of B, C, and D occur similarly to those of N and A. Finally the IC values of the combination can be produced.

In terms of the hierarchies in the dictionaries used, four types of signal detection including detection for "single drug-single ADR," "single drug-ADR category," "drug category-single ADR," and "drug category-ADR category" combinations were constructed in GDQSDS. This yields four types of signal detection, for example, "cefradine-induced hematuria," "cefradine-induced urinary-system disorders," "cephalosporin-induced hematuria," and "cephalosporin-induced urinary-system disorders." Single ADR refers to the preferred term (PT) level of WHO-ART, and the ADR category refers approximately to the higher level group term of WHO-ART. "Single drug-single ADR" signal detection is carried out at the level of specific drug and preferred term, and it is the most important. At the same time, other types of signal detection can also be an extremely useful.

Report generation

All results of association detection will be shown in tabular format on the user interface. For medium or strong associations, an additional table will be provided, and a corresponding graph of dynamic IC values will be shown. If a particular combination always presents a medium or strong association over the course of 1 year, the user will be alerted by a pop-up dialog box. The automatic warning list is cumulative. On the user interface of the associations warning list, warned associations can be browsed and these warned associations are stored in the system by quarter.

Online assessments based on the known ADR database can be performed by experts, and the final signals should be evaluated by follow-up study. The database of known ADRs is compiled in Guangdong and includes known ADRs announced by regulatory agencies, listed in drug instructions, described in the literature, etc. Files containing information about known ADRs such as DOC or PDF files can be uploaded by users and need to be confirmed by experts and can then be stored in the system. Thus far, 8,055 drug instructions have been stored in the database, and the database is continuously expanding.

User interfaces for logging in, association detection, dynamic IC values, and quarterly automatic warning have been developed for the quantitative signal detection system and are given in the "Appendix."

Several different target audiences with different levels of authority can browse different interfaces. For regulatory agencies (State or Guangdong Food and Drug Administration), the board of experts and the provincial users have the highest authority to browse all the signal detection system including all reports; all positive associations (IC-2SD>0); the weak, medium or strong associations lists; and the automatic early warning list. Only the board of experts can evaluate the detected associations according to IC values, case-specific information, clinical experience, and the literatures. Regulatory agencies can make some decisions. Reporters such as manufacturers, medical and health institutions, individuals, and family-planning agencies have limited authority. They can browse and edit some data reported by them before it is confirmed, and they can browse the automatic warning list, experts' evaluations, and regulatory agencies' decisions. All users can download different user manuals from the system. User categories and the associated privileges are shown in Table 2.

Results

The capacity for signal detection of GDQSDS using the IC analysis was tested with the 62,196 ADR records disintegrated from 41,596 reports collected from the beginning of 2002 to 30 June 2007.

Results of association detection

Association detection results of the Guangdong monitoring database are shown in Table 3. A total of 9,046 drug-ADR combinations with a frequency greater than two were extracted from the test database. Of these, 2,183, 257, and 56 showed weak, medium, and strong associations respectively. In other words, there were 2,496 associations (total of weak, medium, and strong associations), which accounts for 27.59% of the total combinations (2,496/9,046).

Of the 3,662 combinations for "single drug-single ADR," 1,062 combinations were associations, and 28 were strong associations including "cefradine-induced hematuria", "captopril-induced coughing", and other known signals. For the "cefradine-induced hematuria" combination, the changes in association scores from 2004 to 2007 are shown in Table 4 and Fig. 2. The IC is listed or plotted at quarterly intervals with confidence limits shown automatically from GDQSDS. As shown in Table 4 and Fig. 2, the signal for cefradine-induced hematuria was found in the first quarter of 2004. Zero records for the cefradine-hematuria combination had been collected before 2004. The number of records for the combination collected from the beginning of 2004 to the second quarter of 2007 is up to 85. The IC value has increased steadily from 2.07 to 4.63 from 2004 to 2007. The confidence interval (4.28, 4.98) for the second quarter of 2007 is narrower than that for the first quarter of 2004 (0.63, 3.51).

Results of automatic warning

In the automatic warning module, all the warning associations quarterly from 2004 to 2007 can be calculated and browsed. In the second quarter of 2007, there were 110 "single drug-single ADR" associations, 61 "single drug-ADR category" associations, 89 "drug category-single ADR" associations, and 31 "drug category-ADR category" associations on the warning list. Table 5 shows the automatic warning module's top 10 associations for single drug-single ADR pairs based on the test database [10]. The complete list of warned associations for the second quarter of 2007 for "single drug-single ADR" combinations includes both associations that are common to other countries and others that are unique to Chinese medicine.

Table 2 User categories and the associated privileges

Users	Privileges
Reporters	Report ADR. Browse the ADRs reported by them. Edit the ADRs reported by them before they are confirmed. Browse the automatic warning list by quarter. Browse experts' evaluations and regulatory agencies' decisions.
Provincial users	Check or confirm the ADRs reported by the reporters. Browse all reports; all positive associations; weak, medium, or strong associations list; and automatic early warning list. Manually detect association. Browse experts' evaluations and regulatory agencies' decisions.
Board of experts	Evaluate the detected associations. Browse all reports; all positive associations; weak, medium, or strong associations list; and automatic early warning list. Manually detect association. Browse regulatory agencies' decisions.
Regulatory agency	Make some decisions. Browse all reports; all positive associations; weak, medium, or strong associations list; and automatic early warning list. Manually detect associations. Browse experts' evaluations.

Table 3 Association detection results of the Guangdong monitoring database collected from 1 January 2002 to 30 June 2007

Type of signal detection	Number of combinations	Negative associations	Weak associations	Medium associations	Strong associations
Single drug-single ADR	3,662	2,600	940	94	28
Single drug-ADR category	2,836	2,119	649	61	7
Drug category-single ADR	1,804	1,263	451	70	20
Drug category-ADR category	744	568	143	32	1
Total	9,046	6,550	2,183	257	56

ADR Adverse drug reaction

Results of signal detection

The board of experts evaluated all 2,496 detected associations, especially the 291 drug-ADR associations that were first alerted in the second quarter of 2007. Sixteen associations were recommended for further study. Three safety-assessment cohort studies based on signal tests about Qingkailing, Danshen and Xiangdan injections were undertaken voluntarily by the pharmaceutical companies and carried out in Guangdong starting in 2007.

Traditional and herbal drugs are popular in China. ADRs induced by these drugs are more often found in Chinese databases. A total of 7,055 ADR reports (7,055/41,596=16.96%) related to Chinese traditional and herbal drugs were collected from the beginning of 2002 to 30 June 2007 in Guangdong. Thus signal detection and statistical descriptive analysis in China could be used as a reference for traditional Chinese medicine in other countries.

Discussion

The IC analysis with some changes was applied in the GDQSDS according to the characteristics of the Chinese ADR database. For example, data disintegration, data coding and some criteria were established based on

Table 4 Changes in association scores of "cefradine-induced hematuria" combination from 2004 to 2007

Quarter	A	B	C	D	Serious association	IC_LL	IC	IC_UL
1st quarter 2004	4	36	15	4,226	0	0.63	2.07	3.51
2nd quarter 2004	5	49	18	5,171	1	0.96	2.26	3.56
3rd quarter 2004	13	67	21	6,400	1	2.39	3.29	4.18
4th quarter 2004	19	84	25	8,759	1	2.95	3.71	4.47
1st quarter 2005	19	89	25	9,466	1	2.97	3.73	4.48
2nd quarter 2005	21	109	29	11,076	1	3.08	3.79	4.50
3rd quarter 2005	24	126	32	13,207	2	3.26	3.93	4.59
4th quarter 2005	31	161	40	16,895	3	3.56	4.14	4.73
1st quarter 2006	33	179	47	19,324	3	3.61	4.18	4.74
2nd quarter 2006	43	235	57	25,336	4	3.89	4.39	4.89
3rd quarter 2006	57	320	88	34,449	4	4.06	4.49	4.92
4th quarter 2006	72	455	113	46,181	4	4.18	4.56	4.94
1st quarter 2007	77	499	123	51,582	5	4.23	4.60	4.97
2nd quarter 2007	85	609	135	61,367	6	4.28	4.63	4.98

IC Information component, *IC_LL* lower limit of confidence interval of IC (IC−2SD), *IC_UL* upper limit of confidence interval of IC (IC+2SD). *A* represents the number of records in which the combination cefradine + hematuria was mentioned. *B* represents the number of records concerning cefradine in combination with other possible adverse drug reactions (ADRs). *C* is the number of records concerning hematuria combined with other drugs. *D* reflects the number of records concerning other drugs combined with other ADRs

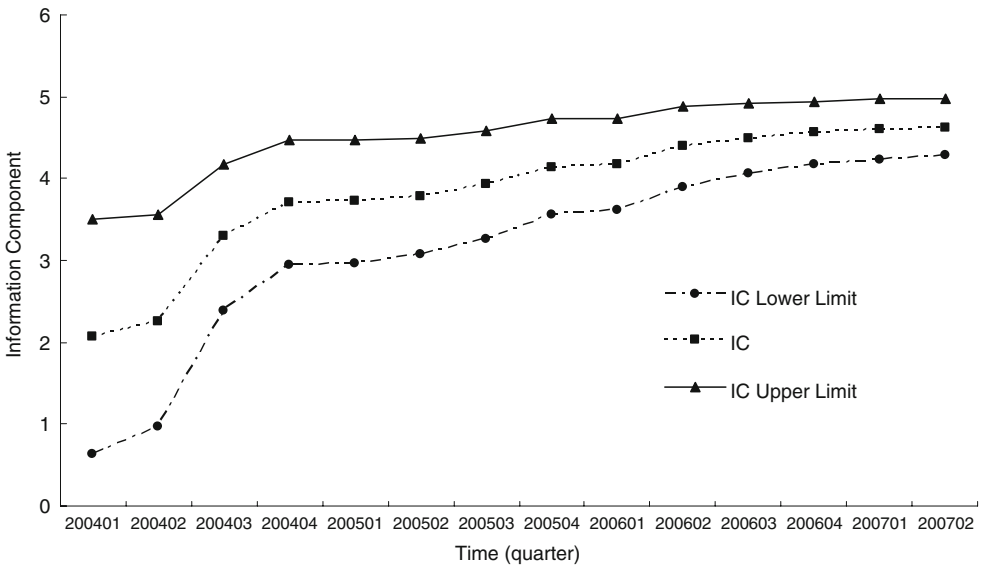

Fig. 2 Dynamic IC values for "cefradine-induced hematuria" combination

Chinese characteristics. Because signal detection is the first step in Guangdong database, only simple drug-ADR combinations were analyzed. However, if a certain ADR occurs more often than expected when two drugs are used concomitantly, this may indicate the existence of a drug-drug interaction. In the next step of our research, quantitative signal generation can be used to study more complex relationships, such as drug-drug interactions and drug-related syndromes. Techniques for the detection of drug interactions and syndromes offer a new challenge for pharmacovigilance in Guangdong.

Due to the "spontaneous" character of the reporting, the SRS method has some limitations. The most noticeable problem is underreporting. Since not all ADRs are reported, the data set of an SRS does not necessarily constitute a valid representation of the ADRs occurring in daily practice [1]. Although "intensive monitoring" (e.g., intensive monitoring for gatifloxacin) is available for the detection of ADRs of marketed drugs in Guangdong, SRS is still an important method for post-marketing surveillance. Guangdong SRS has its own unique characteristics [10], so the adaptability of the IC method needs to be confirmed.

Half of the top-reported drugs were cephalosporins. A greater number of reports doesn't necessarily mean a greater ADR incidence. The ADR incidence is the number of

Table 5 Top 10 warned associations based on IC_LL for the second quarter of 2007 (single drug-single ADR associations)

Drug	ADR	A	IC_LL	IC	IC_UL
Haloperidol	Dystonia	58	5.10	5.58	6.06
Haloperidol	Extrapyramidal disorder	84	4.89	5.32	5.76
Risperidone	Extrapyramidal disorder	81	4.72	5.14	5.57
Rifampicin	Abnormal hepatic function	123	4.53	4.86	5.19
Isoniazid	Abnormal hepatic function	81	4.46	4.88	5.30
Clozapine	Constipation	41	4.43	4.94	5.46
Rifampicin	Hepatitis	60	4.42	4.85	5.27
Captopril	Coughing	37	4.29	4.94	5.60
Cefradine	Hematuria	85	4.28	4.63	4.98
Pyrazinamide	Hepatitis	42	4.23	4.76	5.28

ADR Adverse drug reaction, *IC* information component, *IC_LL* lower limit of confidence interval of IC (IC−2SD), *IC_UL* upper limit of confidence interval of IC (IC+2SD). *A* represents the number of records in which the specific combination of the drug and suspected ADR was mentioned

subjects experiencing the ADR divided by the total number of subjects at risk. Cephalosporins may have induced more ADR reports for three reasons. Firstly, cephalosporin drugs are prone to induce rash and other anaphylactic reactions. Secondly, antibiotics abuse is a big problem in China. Higher levels of cephalosporin administration (i.e., higher frequency of cephalosporin drug exposure) resulted in higher levels of reactions reported. Thirdly, a higher probability of ADR reporting (or lower underreporting coefficient) for cephalosporins is a factor too. More and more attention has been paid to the problem of antibiotics abuse. Mathematical tools based on disproportionality will generate many associations from the large spontaneous ADR database, so the sensitivity and specificity of the IC method based on Guangdong database need to be confirmed.

There is currently no gold standard that establishes universal thresholds for statistical signals. The various measures that are being applied in quantitative signal detection in various national centers are comparable when four or more reports constitute the drug-ADR combination [9]. Because one or two cases have no significance in signal detection, the IC was applied to detect signals in the Guangdong database from drug-ADR combinations in which the number of records per combination is more than two. Two is a temporary threshold of "interest" but not a clinical one. Further practical experience and formal validation studies are necessary to assess what thresholds should be applied routinely in the quantitative signal detection system.

In GDQSDS, a negative association was defined as IC-2SD$\leq$0, a weak association as 0<IC-2SD$\leq$1.5, a medium association as 1.5<IC-2SD$\leq$3.0, and a strong association as IC-2SD>3.0. If one combination always presents a medium or strong association over 1 year, an alert will be generated. The risk of using these criteria is delayed detection of emerging issues. So we considered two ways to reduce the risk. First, if several ADR reports are collected from the same medical institution or the same district in a very short time, we will alert it, and call it an intensive event. Second, these criteria will be updated according to the numbers of associations and based on future research. The time of discovery is not only dependent on the measure used but also on the criteria chosen for alerting, and the time frame of interest is really when the IC starts to increase. For these reasons, we will advise the board of experts to evaluate reports with the IC analysis tool in a timely manner and consider revising our criteria for general users, for example, by changing the criterion A to $\geq$3 or automatically computing the IC each month instead of each quarter.

Several different target audiences with different levels of authority can be logged into the signal detection system. How great the level of authority should be for the different users needs to be confirmed through use.

Bayes's theorem expresses the relationship between the probability of a proposition before (prior) and after (posterior) the acquisition of additional data [1]. The results of the quantitative approach should be considered as additional information and be interpreted in combination with clinical information, findings reported in the literature, and other sources. Following this reasoning, it can be stated that an association may not indicate a true signal. Signal analysis and interpretation including via cohort study are still decisive in the signal detection process.

Conclusion

Although statistical analysis of SRS data sets is not the only method used in the signal detection process, it is currently gaining ground and is the most convenient way to mine large databases. GDQSDS, the computerized web-based system built on Guangdong's ADR database, is fast and offers a more standardized statistical signal detection tool. The system has three modules that support timely detection of four types of signals for marketed products; allow easy detection of changes in association scores; and prioritize, track, and document signals. A total of 5,407 rules for standardizing generic drug names and 8,715 rules for standardizing ADR names allow automatic completion of the standardization work of new cases by matching, which is a very important part of the signal detection process.

The IC measure was integrated into the system for the detection of associations from the database of drug monitoring in Guangdong and can be an extremely useful adjunct to the expert assessment of very large numbers of spontaneously reported ADRs. This web-based computerized quantitative signal detection system to mine ADR reports was established in China for the first time, and it can greatly enhance the efficiency of ADR monitoring and find new drug-ADR signals and outstanding safety problems. Quantitative signal detection systems will play a greater role in pharmacovigilance in China in the future based on a national regulatory database.

Acknowledgements We are grateful to Dr. E. P. van Puijenbroek (Netherlands Pharmacovigilance Foundation Lareb, the Netherlands) for his valuable comments. In addition, we gratefully acknowledge the two reviewers whose valuable comments and helpful remarks and suggestions for this paper led to a remarkable improvement.

Appendix

This appendix presents the web-based interfaces for logging in, association detection, dynamic IC values, and quarterly automatic warnings (images are screen prints or Chinese-to-English representational diagrams).

Figure 3 shows the log in interface. Different users can log in to the system using passwords.

Figure 4 is the user interface for the quantitative signal detection system in Guangdong (GDQSDS). On the left of Fig. 4b is the system menu. The "system setting" function is used to allocate different users with different privileges. "ADR tables" can perform statistical descriptions and export different statistical tables according to different requirements. Data coding, rules database, data audit, etc., are included in "ADR information management." Drug instructions and known ADRs are stored in "Knowledge databases." In addition, there are six buttons for the four types of association detection, the medium or strong associations list, and the quarterly automatic warnings on the Systems menu. On the right side of Fig. 4b is the display interface. For association detection, on the upper part of the display interface, there are four textboxes (generic name, ADR, and the range for date received) and two buttons (search and export). Under these buttons is the output window, and related tables or graphs will be shown in this window.

Figure 5 is the dynamic IC values interface. On the left of this interface is the System menu. An output window is on the right, which displays the IC values in graphic and tabular format at the same time.

Fig. 3 Log in interface

a

b

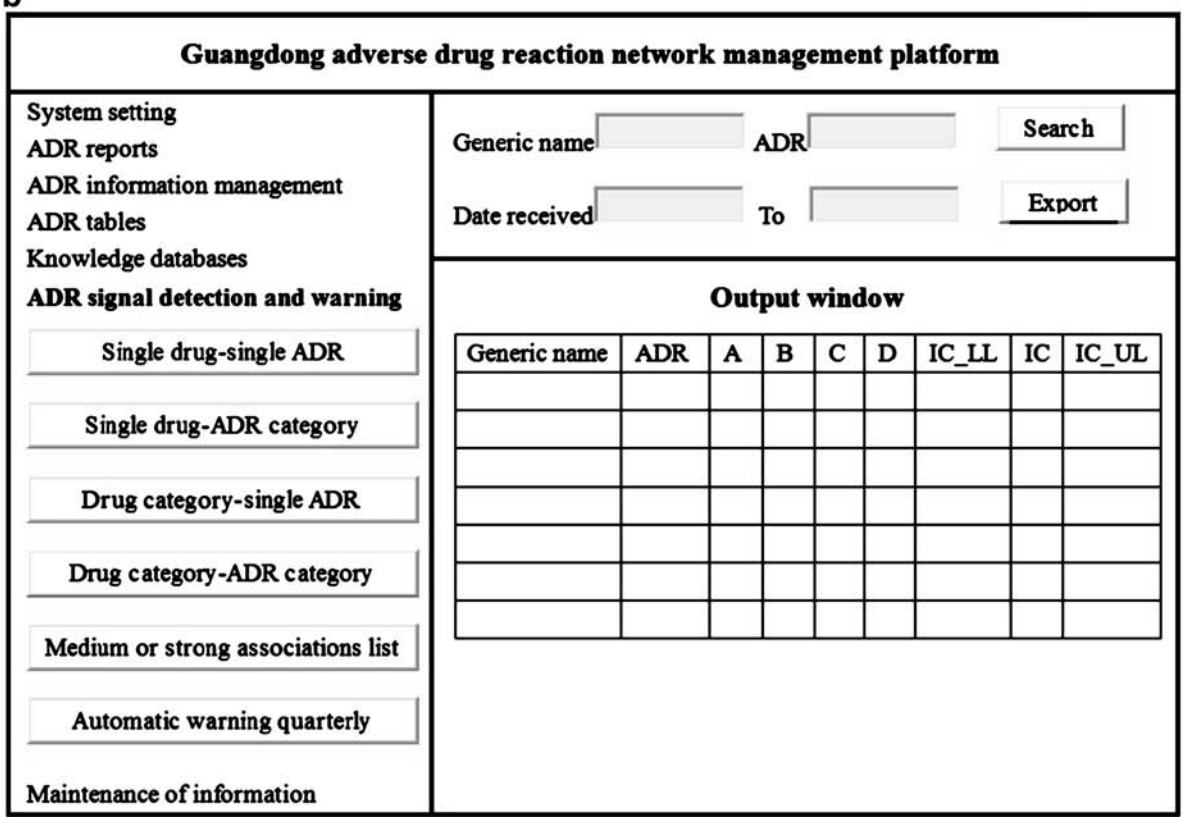

Fig. 4a, b User interface for the quantitative signal detection system. **a** Screen print of the Chinese version. **b** Representative diagram in English

a

b

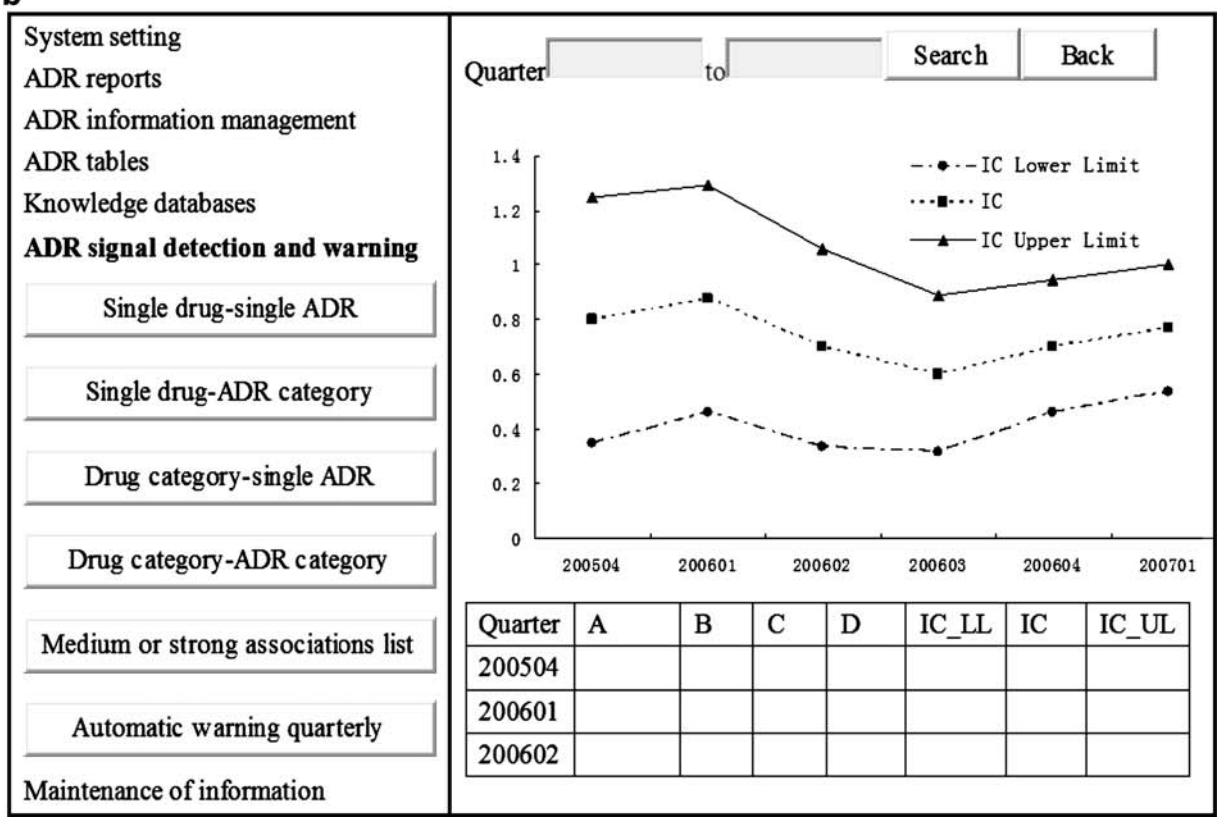

Fig. 5a, b User interface of dynamic IC values. **a** Screen print of the Chinese version. **b** Representative diagram in English

a

b

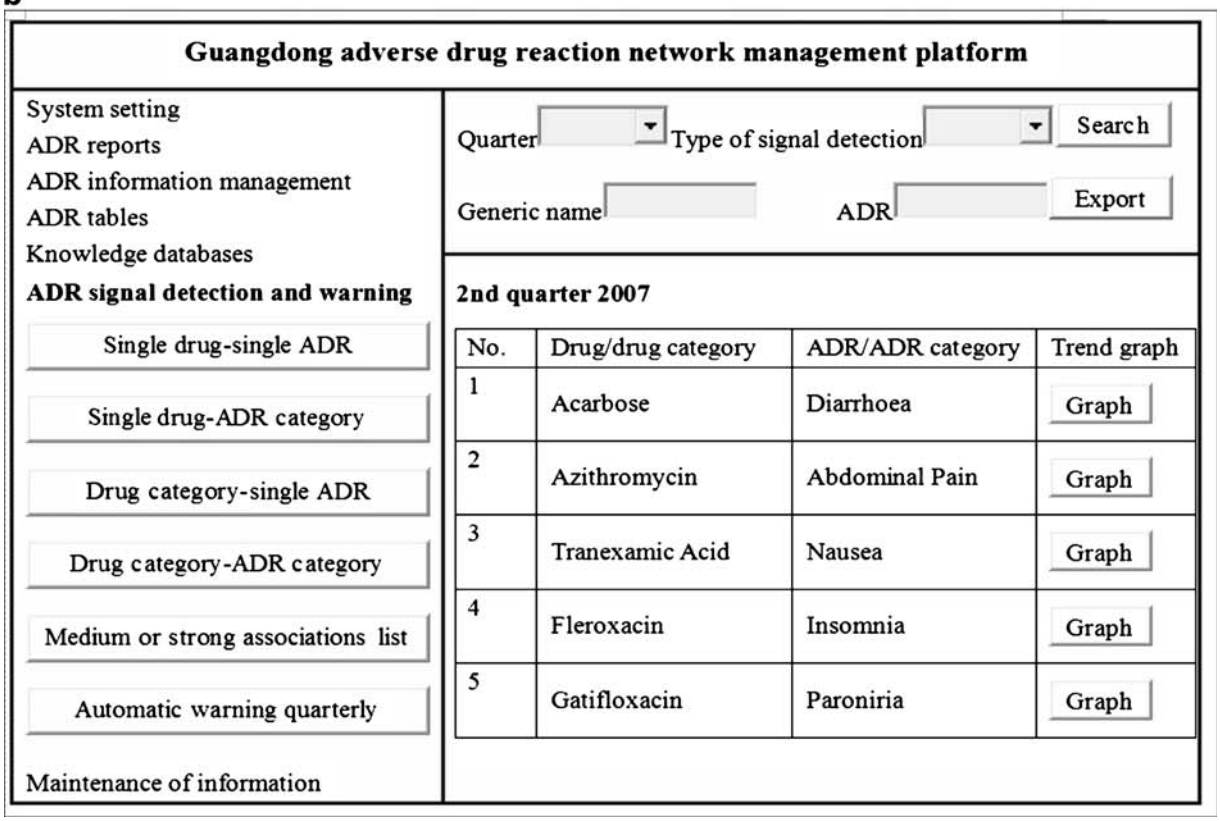

Fig. 6a, b User interface of automatic warning quarterly. **a** Screen print of the Chinese version. **b** Representative diagram in English

Figure 6 is the quarterly automatic warning interface. On the bottom-right display table, generic name and ADR term are shown. When the graph button is clicked, the output window in Fig. 5 will appear. The related warning associations can be browsed by selecting "Quarter" and "Type of signal detection" items.

References

1. Van Puijenbroek EP, Diemont WL, van Grootheest K (2003) Application of quantitative signal detection in the Dutch Spontaneous Reporting System for adverse drug reactions. Drug Saf 26 (5):293–301
2. Evans SJW, Waller PC, Davis S (2001) Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf 10:483–486
3. Jones JK (2001) The role of data mining technology in the identification of signals of possible adverse drug reactions: value and limitations. Curr Ther Res 62(9):664–672
4. DuMouchel W (1999) Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat 53(3):177–190
5. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM (1998) A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 54:315–321
6. Bate A, Lindquist M, Edwards IR, Orre R (2002) A data mining approach for signal detection and analysis. Drug Saf 25(6):393–397
7. Hauben M, Zhou XF (2003) Quantitative methods in pharmacovigilance: focus on signal detection. Drug Saf 26(3):159–186
8. Wilson AM, Thabane L, Holbrook A (2003) Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol 57 (2):127–134
9. Van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC (2002) A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf 11(1):3–10
10. Li C, Xia J, Deng J, Jiang J (2008) A comparison of measures of disproportionality for signal detection on adverse drug reaction spontaneous reporting database of Guangdong province in China. Pharmacoepidemiol Drug Saf 17(6):593–600
11. Norén GN, Bate A, Orre R, Edwards IR (2006) Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. Stat Med 25(21):3740–3757
12. Lindquist M, Ståhl M, Bate A, Edwards IR, Fucik H, Nunes AM (1998) From association to alert-a revised approach to international signal analysis. Pharmacoepidemiol Drug Saf 8: S15–S25
13. Lindquist M, Ståhl M, Bate A, Edwards IR, Meyboom RH (2000) A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. Drug Saf 23:533–542
14. Orre R, Lansner A, Bate A, Lindquist M (2000) Bayesian neural networks with confidence estimations applied to data mining. Comput Stat Data Anal 34:473–493
15. Wei Z, Cheng N, He L, Du W, Xu J, Chen B, Wang Y (2006) Establishment of quantitative signal detection system on adverse drug reaction spontaneous reporting database of Shanghai. Fudan Univ J Med Sci 33(4):475–478
16. Cai H, Xia J, Xu D, Gao D, Yan Y (2006) A generic minimization random allocation and blinding system on web. J Biomed Inform 39:706–719
17. Chen X (2003) New pharmacology, 15th ed. The People's Public Health Publishing, Peking
18. Sichuan Meikang Medicine Software Company (2006) MCDEX clinical drug reference. Sichuan Technology Publishing, Chengdu
19. China National Adverse Drug Reaction Monitoring Centre (2003) WHO adverse drug reaction dictionary. China Medicine Technology Publishing, Peking
20. China Statistical Information Centre, Ministry of Health (2001) International classification of diseases, 10th revision. China Union Medical University Press, Peking