Received: 22 February 2019 | Revised: 27 June 2019 | Accepted: 17 July 2019

DOI: 10.1111/jcpt.13021

ORIGINAL ARTICLE

Journal of Clinical Pharmacy and Therapeutics WILEY

Detection of cluster adverse drug events in the spontaneous reporting system of China using a disproportionality filter algorithm

Meng Wang MS[1,2] | Xiaofei Ye PhD[1] | Yiming Ruan MS[1] | Jinfang Xu PhD[1] | Xiaojing Guo PhD[1] | Duo Dong PhD[3] | Hongyun Feng PhD[3] | Jia He MD, PhD[1]

[1]Department of Health Statistics, Second Military Medical University, Shanghai, China

[2]General Hospital of Southern Theatre Command of PLA, Guangzhou, China

[3]National Center for ADR Monitoring, Beijing, China

Correspondence
Jia He, Department of Health Statistics, Second Military Medical University, No. 800 Xiangyin Road, Shanghai 200433, China.
Email: hejia63@yeah.net

Hongyun Feng, National Center for ADR Monitoring, Beijing 100045, China.
Email: even_fzh@aliyun.com

Funding information
This study was supported by a grant from the Fourth Round of Three-year Action Plan on Public Health Discipline and Talent Program: Evidence-based Public Health and Health Economics, Grant/Award Number: 15GWZK0901; National Nature Science Foundation of China, Grant/Award Numbers: 81703296; Nature Science Foundation of Shanghai, Grant/Award Number: 18ZR1449500; Shanghai municipal commission of health and family planning fund for excellent young scholars, Grant/Award Number: 2018YQ47.

Abstract

What is known and objective: Cluster adverse drug events (CADEs) are multiple ADEs with similar clinical manifestations involving the same drug, manufactured by the same company, that occur within a short time period. The disproportionality filter algorithm (DFA) is a promising tool for the identification of historical clusters related to ADEs. The Chinese spontaneous adverse drug reaction reporting system (SRS) may serve as an important database for the detection of CADEs. The objective of this study was to evaluate the usefulness of DFA as an approach to identify CADEs using SRS.

Methods: Suspected adverse drug reaction (ADR) reports collected by the Chinese SRS in 2014-2015 were examined to identify potential CADEs. The reports were divided into 48 15-day subsets. Disproportionate excess reporting of ADEs for drugs from specific companies may be a signal for CADEs. The clusters were categorized as 'confirmed', 'potential', 'unlikely', 'indecisive' or 'information-loss' ADEs when evaluated by report units. Furthermore, early warning information in 2014-2015 from the Chinese early warning system (EWS) classified as 'concern', 'monitoring', 'ignorance' or 'rest' was compared with DFA to explore the applicability of the novel algorithm in Chinese SRS.

Results and discussion: In total, 2294 CADEs, comprising of 380 confirmed, 1753 potential, 15 unlikely and 59 indecisive clusters, were generated; 87 clusters were missing additional information. There were 263 'significant' clusters with DFA, but only 26 'significant' clusters in EWS. The sensitivity of DFA was 88.5%, but the specificity and positive predictive value were low.

What is new and conclusion: Spontaneous adverse drug reaction reporting system in China may be a potential database for the identification of CADEs engaging the DFA. This could supplement the EWS of CADEs in China. The DFA may be of value in detecting CADEs with high sensitivity, although expert screening is required given the low specificity and positive predictive value.

KEYWORDS

cluster adverse drug events, disproportionality filter algorithm, spontaneous adverse drug reaction reporting system

Meng Wang, Xiaofei Ye, Yiming Ruan and Jinfang Xu authors contributed equally to the study.

1 | WHAT IS KNOWN AND OBJECTIVE

Medicine can be a double-edged sword for public health. Adverse drug reactions (ADRs) and adverse drug events (ADEs) are a major global health threat.[1] It is estimated that 3%-7% of all hospital admissions and a substantial increase in morbidity and mortality are attributed to ADEs.[2] A cluster of multiple ADEs poses a significant potential hazard to patient safety. Such serious lessons are common occurrences in the history of pharmaceutical development, such as the many disabilities caused by thalidomide and the cases of acute renal failure induced by medicines manufactured by 'Qiqihar' in China.[3,4] To minimize the impact of ADEs, specifically cluster adverse drug events (CADEs), which are characterized by multiple similar ADEs that occur centrally with the same drug manufactured by the same manufacturer within a short time period,[5] the signals of the CADEs need to be discovered, identified and controlled as early as possible.

Disproportionality analysis (DPA) is a data mining technique employed in pharmacovigilance for the identification and verification of ADRs. The most widely used DPA methods in SRS include the reporting odds ratio (ROR), proportional ADR reporting ratio (PRR), information component (IC) and the Multi-item Gamma Poisson Shrinker (MGPS).[6-9] In 2011, the Uppsala Monitoring Center (UMC) conceptualized and described a novel algorithm, abbreviated DFA, based on the IC theory of DPA, which was used to identify some historical clusters.[10] In the study, a cluster signal was identified if countries and time periods displayed disproportionately high reporting of medicine inadequacy. China has developed an early warning system (EWS) for CADEs.[11] However, manual confirmation of these events takes time and effort and the standards for signal generation are subjective.

With the rapid progress of the information era, SRSs, which are used for post-marketing drug safety surveillance, have earned a reputation for offering a fast and efficient way to detect ADRs.[12] In the last few decades, SRS in China has been updated and improved and a massive amount of data has been assembled; the number of ADRs reported each year continues to increase. In 2017, the number of reports collected was approximately 1.42 million or so.[13] Spontaneous reports can, therefore, be a valuable source of information for the detection of CADEs.

Although the capacity of DFA has been confirmed, its practicality for identifying CADEs within a drug manufacturing company and time period in SRS is unknown. The objective of this study was to examine the performance of the DFA as a promising approach for identifying CADEs using the SRS database in China as a potential data source.

2 | METHODS

2.1 | Data source and data cleaning

The data source of this study was the SRS database in China from 01 January 2014 to 31 December 2015. The database was established by the China Food and Drug Administration (CFDA) almost 30 years ago and, as of 2017, contained approximately 12.18 million reports.[13,14] The database includes suspected ADR reports submitted by relevant organizations (drug manufacturing companies, drug distributors or medical institutions) or individuals. In 2014-2015, the database included a total of 2 703 902 reports from each province and 58 variables. China also has an EWS of CADEs based on daily monitoring. From this EWS, 8521 clusters from 2014 to 2015 were available for comparison and evaluation of the effectiveness of DFA.

To ensure the credibility of the study, all ADE names were mapped to the preferred term (PT) listed in the World Health Organization Adverse Reaction Terminology (WHO-ART) published in 2002. The names of drug products were converted into the terms listed in the Chinese Pharmacopoeia. Reports with no or incomplete records for ADE name, drug name or manufacturing company name were excluded from the analysis. The original reports, which probably contained several types of medicines or different ADEs, were split into a variety of single drug-ADE pairs. Concerning the violation of normativity among reports (ie '*', '#' or ','), which results in data invalidity, all meaningless or uncorrelated characters were deleted.

2.2 | DFA

2.2.1 | Division of data

Disproportionality filter algorithm proposed by the UMC was derived from classic IC method[15] with a subtle introduction of layered thought.[16] DFA identifies groups of reports on a drug with disproportionately excess reporting of ADEs. These groups of reports, which consist of reports on a specific medicine within a company and an interval of 15 days, are defined as clusters.

Each drug-event pair of reports was classified to a company-15 days stratum, based on the drug company and the date of the ADE. A period of 15 days was adopted as the time division standard as it is the maximum monitoring interval in Chinese EWS for CADEs.[17] The time division of each cluster started from 1 January 2014 with a cluster set every 15 days until 31 December 2015. Therefore, reports were divided into 48 clusters in each company.

To ensure a sufficient quantity of the drug-event pairs for analysis in each stratum, we deleted companies that produced fewer than three types of medicines or which were related to fewer than 200 reports. Finally, we identified 3 540 260 drug-event pairs as the data for the analysis.

TABLE 1 Cross table of reports of reports

	Targeted ADE	Not targeted ADE
Targeted drug	a	b
Not targeted drug	c	d

Note: a denotes the number of targeted ADEs caused by a targeted drug; b denotes the number of ADEs not of interest caused by a targeted drug; c denotes the number of targeted ADEs caused by a drug not of interest; and d denotes the number of ADEs not interested caused by drug not interested.

TABLE 2 Suspected ADEs reported as number of strata of companies with different ADR event numbers

Event number	Frequency	Percentage (%)	Cumulative frequency	Cumulative percentage (%)
200-	852 594	24.08	852 594	24.08
2000-	507 228	14.33	1 359 822	38.41
4000-	267 310	7.55	1 627 132	45.96
6000-	314 635	8.89	1 941 767	54.85
8000-	210 128	5.94	2 151 895	60.78
10000-	240 003	6.78	2 391 898	67.56
12000-	127 733	3.61	2 519 631	71.17
14000-	60 559	1.71	2 580 190	72.88
16000-	118 171	3.34	2 698 361	76.22
18000-	56 883	1.61	2 755 244	77.83
20000-	63 560	1.80	2 818 804	79.62
22000-	46 712	1.32	2 865 516	80.94
24000-	24 398	0.69	2 889 914	81.63
26000-	81 141	2.29	2 971 055	83.92
28000-	29 696	0.84	3 000 751	84.76
30000-	539 509	15.24	3 540 260	100.00

2.2.2 | Identification of stratum

The association between a drug and an event can be measured as the ratio of the observed to the expected frequency of events. Based on whether or not a drug and an event were expected, a cross table of reports was constructed (Table 1).

Given the assumption that the reporting of the drug and the event are independent of each other, the calculation in the cross-classification table of a specific drug-event pair of interest can be expressed in a quantitative relative estimate of disproportionality. The observed-to-expected (OE) ratio for the association between drug and event of interest was expressed as follows:

$$OE = \frac{a}{(a+b)\times(a+c)/(a+b+c+d)}$$

By comparing the OE ratio in that subset with the OE ratios in other enterprises and other time periods, we can estimate associations between the drugs of interest and other drugs produced by one specific manufacturing enterprise and each 15-day period. Similarly, the OE ratio in complementary stratum, stratified by manufacturing company and time interval, can be computed as a weighted average of the OE ratios for each company-time stratum in the data for all strata, except those from the selected stratum. The weighted average may be helpful to illustrate the expected distribution of drug use and suspected ADEs reports. We searched for companies and time periods in which a specific ADE occurred more often than on average for the same medicine compared with companies and time periods.

Kristina et al[10] considered that, on the basis of the derived OE ratio, a comparative IC value (IC_{Δ}), a bivariate measure based on the OE ratio of the drug and events of interest, can be derived to estimate the possibility of a CADE.

An IC_{Δ} above 0 indicates that the drug has a greater relative frequency of reports than drugs produced by the company of interest than in the database as a whole. When the lower 95% confidence interval of the IC_{Δ} exceeds 0, the specific stratum is considered as a CADE cluster.

2.3 | Assessment of suspected CADEs

Each report was assessed as 'confirmed', 'potential', 'unlikely', 'indecisive' or 'information-loss' CADEs by the reporting units from different provinces in China. When the reports were sent to the SRS database, the reporting units provided a pre-evaluation about the relationship between drugs and ADEs. The cluster assessment result was collected in one variable in the database. The definition of each level of the suspected CADEs was expressed as follows:

- Confirmed CADE: A known CADE

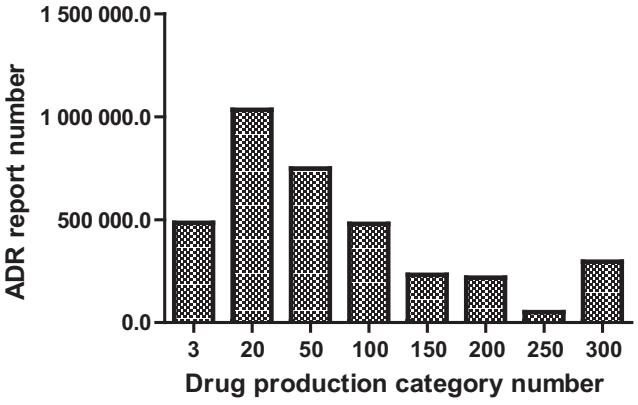


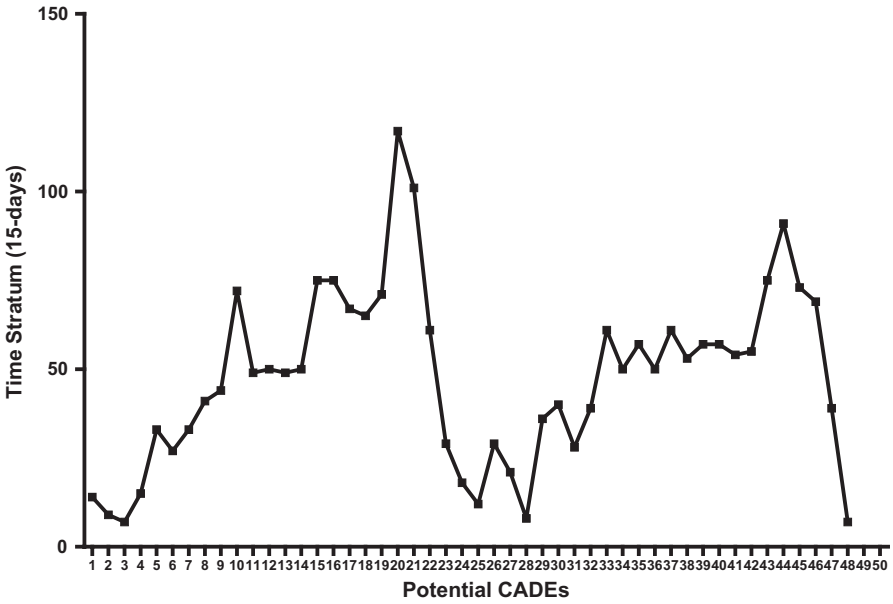
FIGURE 1 Suspected ADR report numbers of company strata for different drug production category numbers

TABLE 3 Top 10 CADEs in descending order by the lower 95% confidence interval of IC_{Δ}

No.	Manufacturer location	Drug-ADRs	Time to onset	IC_{Δ} 95% BMDL
1	Hebei Province	Drug A—Headache	2014-02-27	2.182
2	Jiangsu Province	Drug B—Fever	2015-12-14	2.060
3	Xian Province	Drug C—Headache	2014-08-07	1.961
4	Sichuan Province	Drug D—Palpitate	2015-10-16	1.955
5	Hebei Province	Drug E—Diarrhoea	2015-03-30	1.813
6	Tianjin Municipality	Drug F—Diarrhoea	2015-05-07	1.795
7	Hebei Province	Drug E—Diarrhoea	2014-06-20	1.784
8	Jilin Province	Drug G—Nausea	2015-09-22	1.762
9	Tianjin Municipality	Drug F—Diarrhoea	2014-07-11	1.702
10	Zhejiang Province	Drug H—Liver dysfunction	2014-07-01	1.666

Note: The details of the manufacturers and the real drug name could not be disclosed in the current report for confidentiality reasons.

FIGURE 2 Number of potential CADE clusters in different 15-d time strata



- Potential CADE: There were plausible reasons to believe that this may be a CADE
- Unlikely CADE: Compelling reasoning and supporting information that this was not likely related to a CADE
- Indecisive: Could not be assessed with a sufficient degree of certainty, because of lack of information
- Missing: Not been assessed by the reporting unit

3 | RESULTS

3.1 | Reports appearing in the database

In total, 1665 drug manufacturing entities were selected as the subjects of investigation. Division of these entities by the number of their suspected ADEs showed that those with 200-2000 suspected ADEs in the database comprised the largest proportion, contributing to 24.08% of the reports in the database. Those with 24 000-26 000 reports account for the smallest proportion (0.69%) (Table 2).

The difference in the number of produced drug categories was considerable between different companies. The minimum drug production category number was 3, and the maximum was 528 (Figure 1).

3.2 | DPA results

In total, 2294 potential CADEs were selected through 48 DPA (Top 10 IC_{Δ} are shown in Table 3). Analysis of reports between 15 October 2014 and 31 October 2014 generated 117 potential CADE clusters, the most frequent among the 48 analyses. The minimum was seven clusters, which were selected from the reports between 1 February 2014 and 15 February 2014 (Figure 2).

3.3 | Performance of DFA

Of the 2294 suspected CADEs, 380 clusters were confirmed as CADEs (16.56% hit rate), 1753 clusters as potential and 15 clusters

TABLE 4 Classification for assessable clusters from evaluation by report units

Classification	Number	Percentage (%)
Confirmed	380	16.56
Potential	1753	76.42
Unlikely	15	0.65
Indecisive	59	0.70
Missing	87	3.79
Total	2294	1.00

TABLE 5 Example of EWS and DFA matching

Manufacturer	DRUG name	EWS documented	Significant in EWS	DFA detected	Significant by DFA	Match result
A	DRUG 1	Yes	Yes	Yes	Yes	Yes
B	DRUG 2	Yes	No	No	-	-
B	DRUG 2	Duplicate deleted	Deleted	No	-	-
C	DRUG 3	No	No	Yes	Yes	-
C	DRUG 3	No	No	Duplicate deleted	Deleted	-
D	DRUG 4	Yes	No	Yes	Yes	No
E	DRUG 5	Yes	Yes	Yes	No	No
.....	.....	.....	.....	.....	.....	.....

TABLE 6 Cross table of method evaluation

		EWS		
		Significant signals	Insignificant signals	Count
DFA	Significant CADEs	23	148	171
	Insignificant CADEs	3	106	109
Count		26	254	280

as unlikely; 59 clusters were undecided, and information was unavailable for 87 clusters (Table 4).

To compare CADEs detected by DFA with the early warning signals reported in the national EWS, we considered only information on drug and manufacturer. Consequently, 516 of 2294 clusters with DFA and 1619 of 8521 signals in EWS were chosen, and 280 clusters were identified in both; an example is presented in Table 5. We regarded confirmed and potential clusters with DFA and considered concern and monitoring in EWS as significant clusters. The detailed results are displayed in Table 6. The sensitivity of DFA was 88.5%, but the specificity and positive predictive values were low (41.7% and 13.5%, respectively).

4 | DISCUSSION

According to Kristina et al,[10] the DFA is efficient for the identification of substandard medicines. Using their method as the reference, our study is, to the best of our knowledge, the first to detect CADEs by DPA in SRS in China. Through our analysis, some CADEs that have been confirmed by national EWS were detected. Kristina et al[10] showed that an IC_{Δ} above 0 was a filtering criterion indicating that ADE reports have a greater relative frequency in the companies and year of interest than in the database as a whole. To preventing potential false-positive cases, we applied more stringent standards to identify CADEs when the lower 95% confidence interval of the IC_{Δ} exceeded 0. At present, a gold standard has not been established for the identification of CADEs. Further studies under different circumstances would be helpful.

An abundance of information is provided in SRS in China. We selected drug manufacturing enterprises as a stratum to explore the quality of drugs from the perspective of producers. This is important for early warning and control of the CADEs. The variable of batch number was our first consideration for stratified analysis. Batch number can determine the drug production and distribution history, which can be used to monitor the quality of drug and control the ADEs. However, the drug batch number information in our database was incomplete and not standardized, which limited its use. In addition, the time stratum was divided into 48 periods of 15 days; however, some days in our period may be omitted, such as months with more or fewer than 30 days. These days were not allocated to a cluster. In addition, manufacturers with <3 products were not be monitored by DFA. Therefore, the precision and timeliness of reports must be improved for the global pharmacovigilance system to be of broader value in the detection of CADEs.

To compare the CADEs, we identified those from the national EWS and eliminated duplicates by the consideration of some basic information, as this may have had altered the effectiveness of the method. The current results indicated that DFA had a high sensitivity; that is the true positive rate was high. However, the specificity and positive predictive value were low. Thus, the method is more suitable for preliminary detection, with further confirmation required by the relevant professionals.

5 | WHAT IS NEW AND CONCLUSION

The SRS database in China can serve as a potential source for the identification of CADEs and may be a feasible complement for early surveillance in pharmacovigilance. The application of DFA to Chinese SRS is a beneficial attempt and allowed the detection of some clusters of ADEs. With regard to the current flaws in the results, the data quality requires further improvement and simulations in different sceneries may be of value. For broad prospective detection of CADEs, it is certainly of value to apply DFA to real-world research for the performance evaluation of drugs.

ACKNOWLEDGEMENTS

The authors would like to thank Hongyun Feng and Wei Yang from the National Center for ADR Monitoring, for supplying the data extracted from SRS and EWS in China and for their valuable feedback during the study. We would also like to thank the researchers from the Uppsala Monitoring Centre for their developing and sharing of the DFA.

CONFLICT OF INTEREST

Meng Wang, Xiaofei Ye, Yiming Ruan, Jinfang Xu, Xiaojing Guo, Duo Dong, Hongyun Feng and Jia He have no conflicts of interest that are directly relevant to the content of this study.

REFERENCES

1. Begum SS, Mansoor M, Sandeep A, et al. Tools to improve reporting of adverse drug reactions–A review. *Int J Pharm Sci Rev Res*. 2013;23(1):262-265.
2. Montastruc J-L, Sommet A, Bagheri H, Lapeyre-Mestre M. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *Br J Clin Pharmacol*. 2011;72(6):905-908.
3. Warkany J. Why I doubted that thalidomide was the cause of the epidemic of limb defects of 1959 to 1961. *Teratology*. 1988;38(3):217-219.
4. Potential health & safety impacts from pharmaceuticals and supplements containing Chinese-sourced raw ingredients. 2010. https://www.uscc.gov/sites/default/files/Research/NSD_BIO_Pharma_Report--Revised_FINAL_for_PDF--14_%20April_2010.pdf. Accessed May 18, 2018.
5. Feng HY, Wen-Chao GE, Xia XD, et al. Construction and practice of early warning system sharing platform on cluster adverse drug events. *Chin J Pharmacovigilance*. 2015;12(12):727-730.
6. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483.
7. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med*. 2010;25(21):3740-3757.
8. Wechwithan S, Suwankesawong W, Sornsrivichai V, McNeil EB, Jiraphongsa C, Chongsuvivatwong V. Signal detection for Thai traditional medicine Examination of national pharmacovigilance data using reporting odds ratio and reported population attributable risk. *Regul Toxicol Pharmacol*. 2014;70(1):407-412.
9. Dumouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999;53(3):177-190.
10. Juhlin K, Karimi G, Andér M, et al. Using vigibase to identify substandard medicines: detection capacity and key prerequisites. *Drug Saf*. 2015;38(4):373-382.
11. Feng HY, Yang L, Song QJ, et al. Proposal and practice of building the early warning system on cluster adverse drug events based on the adverse drug reaction monitoring. *Chin J Pharmacovigilance*. 2015;12(11):665-668.
12. van Puijenbroek EP, Bate A, Leufkens HG, et al. A comparison of measures of disproportionality for signal detection on adverse drug reaction spontaneous reporting database of Guangdong province in China. *Pharmacoepidemiol Drug Saf*. 2002;11(1):3-10.
13. China Food and Drug Administration. 2017 annual report for national adverse drug reaction monitoring released. 2018. http://www.cdr-adr.org.cn/xwdt/201804/t20180419_20011.html. Accessed April 19, 2018.
14. Hou Y, Li X, Wu G, Ye X. National ADR monitoring system in China. *Drug Saf*. 2016;39(11):1043-1051.
15. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998;54(4):315-321.
16. Khodabakhshi G, Juhlin K, Norén GN. Monitoring medicines (FP7 grant no 223566): D8-substandard medicines oversee a pilot project aimed at development of tools to identify reports indicating substandard medicines. 2011. http://www.monitoringmedicines.org/graphics/27523.pdf2011. Accessed Feb 5, 2015.
17. Feng HY, Hou YF, Wu GZ, et al. The establishing and running of pre-warning system on adverse event cluster signals. *Chin J Pharmacovigilance*. 2012;9(12):745-748.

How to cite this article: Wang M, Ye X, Ruan Y, et al. Detection of cluster adverse drug events in the spontaneous reporting system of China using a disproportionality filter algorithm. *J Clin Pharm Ther*. 2019;44:952–957. https://doi.org/10.1111/jcpt.13021