

Using VigiBase to Identify Substandard Medicines: Detection Capacity and Key Prerequisites

Kristina Juhlin · Ghazaleh Karimi · Maria Andér · Sara Camilli · Mukesh Dheda · Tan Siew Har · Rokiah Isahak · Su-Jung Lee · Sarah Vaughan · Pia Caduff · G. Niklas Norén

Published online: 17 February 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Background Substandard medicines, whether the result of intentional manipulation or lack of compliance with good manufacturing practice (GMP) or good distribution practice (GDP), pose a significant potential threat to patient safety. Spontaneous adverse drug reaction reporting systems can contribute to identification of quality problems that cause unwanted and/or harmful effects, and to identification of clusters of lack of efficacy. In 2011, the Uppsala Monitoring Centre (UMC) constructed a novel algorithm to identify reporting patterns suggestive of substandard medicines in spontaneous reporting, and applied it to VigiBase[®], the World Health Organization's global individual case safety report database. The algorithm identified some historical clusters related to substandard products, which were later able to be confirmed in the literature or by contact with national centres (NCs). As relevant and

detailed information is often lacking in the VigiBase reports but might be available at the reporting NC, further evaluation of the algorithm was undertaken with involvement from NCs.

Key Points

Some cases of substandard medicines can be identified from spontaneous reports.

Many of the suspected clusters in our study could not be conclusively evaluated, because of lack of information on the reports and the difficulty of contacting the primary reporter.

Important prerequisites for broad, prospective detection of substandard medicines in VigiBase[®] are not currently fulfilled.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-015-0271-2) contains supplementary material, which is available to authorized users.

K. Juhlin (✉) · G. Karimi · M. Andér · P. Caduff · G. N. Norén
Uppsala Monitoring Centre, Box 1051, 75140 Uppsala, Sweden
e-mail: kristina.juhlin@who-umc.org

S. Camilli
Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA

M. Dheda
Pharmacovigilance Unit, Medicines Regulatory Authority, National Department of Health, Pretoria, South Africa

T. S. Har
Vigilance and Compliance Branch, Health Products Regulation Group, Singapore, Singapore

R. Isahak
National Pharmaceutical Control Bureau, Ministry of Health, Petaling Jaya, Malaysia

S.-J. Lee
Pharmaceutical Safety Information, Ministry of Food and Drug Safety, Cheongwon-gun, Republic of Korea

S. Vaughan
Medicines and Healthcare Products Regulatory Agency, London, UK

G. N. Norén
Department of Mathematics, Stockholm University, Stockholm, Sweden

Objective To evaluate the effectiveness of an algorithm that identifies clusters of potentially substandard medicines, when these are assessed directly at the NC concerned.

Methods The algorithm identifies countries and time periods with disproportionately high reporting of product inadequacy. NCs with at least 20 clusters were eligible to participate in the study, and six NCs—those in the Republic of Korea, Malaysia, Singapore, South Africa, the UK and the USA—were selected, taking into account the geographical spread and prevalence of recent clusters. The clusters were systematically assessed at the NCs, following a standardized protocol, and then compiled centrally at the UMC. The clusters were classified as ‘confirmed’, ‘potential’ or ‘unlikely’ substandard products; or as ‘confirmed not substandard’ when confirmed by an investigation; or as ‘indecisive’ when the information available did not allow a sound assessment even at the NC.

Results The assessment of a total of 147 clusters resulted in 8 confirmed, 12 potential and 51 unlikely substandard products, and a further 19 clusters were confirmed as not substandard. Reflecting the difficulty of evaluating suspected substandard products retrospectively when additional information from the primary reporter, as well as samples, are no longer available, 57 clusters were classified as indecisive.

Conclusion While application of the algorithm to VigiBase allowed identification of some substandard medicines, some key prerequisites have been identified that need to be fulfilled at the national level for the algorithm to be useful in practice. Such key factors are fast handling and transfer of incoming reports into VigiBase, detailed information on the product and its distribution channels, the possibility of contacting primary reporters for further information, availability of samples of suspected products and laboratory capacity to analyse suspected products.

1 Introduction

Substandard medicines pose a significant potential hazard to patient safety. Inadequate medicinal product quality can be a consequence of lack of compliance with good manufacturing practice (GMP) or good distribution practice (GDP) standards, or a consequence of intentional tampering, which also includes false labelling or fake packaging.

Quality problems can be of only regional interest if they concern an end product with limited circulation, but they can also have a global impact. In January 2008, the US Centers for Disease Control and Prevention began a nationwide investigation of severe adverse drug reactions (ADRs) concerning acute hypersensitivity reactions to heparin sodium for injection; these had occurred within the

previous months, and a significant number of them were reported as having had a fatal outcome [1]. Further clusters occurred in other countries with products from different manufacturers. Investigations confirmed contamination of the heparin supply with oversulfated chondroitin sulfate originating in China and distributed in at least 12 countries [2–4]. This example shows the value of spontaneous ADR reporting systems in evaluating the clinical consequences of product quality issues and demonstrates how this type of data source contains information that could be used for identifying potentially substandard products. The increasing global drug sales over the internet contribute significantly to the dissemination of substandard, often counterfeited products that circumvent regulatory channels. Spontaneous reports could act as a useful complement to regulatory quality-control channels in identifying adverse reactions related to these products. Substandard medicines containing no or too low concentrations of the active ingredients, on the other hand, might be detected through clusters of reports on lack of efficacy. Spontaneous reports can therefore not only be a valuable source of information for detection of adverse reactions but also provide support for identification of quality issues.

In 2011, the Uppsala Monitoring Centre (UMC) led a project within the Monitoring Medicines initiative, funded by the European Commission [5], in which a novel algorithm was constructed to identify reporting patterns suggestive of substandard medicines in spontaneous ADR reporting. The algorithm was applied to VigiBase®, the World Health Organization’s (WHO’s) global individual case safety report (ICSR) database, and managed to identify examples of historical clusters that were subsequently confirmed as being related to substandard medicinal products by the countries concerned [5]. Lack of relevant and detailed information in ICSRs in VigiBase makes evaluation of such clusters at the UMC difficult. As this information is likely to be at least partly available at the national centres (NCs) for pharmacovigilance submitting ICSRs to VigiBase, further evaluation of the effectiveness of the algorithm was undertaken in 2013 and 2014 with involvement from NCs, and the results of this are presented in this article.

2 Methods

There is no universally agreed definition of the term ‘substandard product’. For the purpose of this study, a substandard product was set to mean either a substandard product as defined by the WHO:

Substandard medicines (also called out of specification (OOS) products) are genuine medicines produced by manufacturers authorized by the National

Medicines Regulatory Authority (NMRA) which do not meet quality specifications set for them by national standards.

or a counterfeit drug as defined by the WHO:

A counterfeit medicine is one which is deliberately and fraudulently mislabelled with respect to identity and/or source. Counterfeiting can apply to both branded and generic products and counterfeit products may include products with the correct ingredients or with the wrong ingredients, without active ingredients, with insufficient (inadequate quantities of) active ingredient(s) or with fake packaging.

A term that is collectively used for substandard and counterfeit products is ‘SSFFCs’ (substandard/spurious/falsefully labelled/falsified/counterfeit medical products), which is the acronym used by the WHO (<http://apps.who.int/gb/ssffc/>). In the following text, the terms ‘substandard product’ and ‘SSFFC’ are used as equivalents.

The overall approach to the empirical evaluation of the algorithm together with NCs is schematically described in Fig. 1.

Eligible for participation were NCs with a minimum of 20 clusters generated by the algorithm. Selection was also made on the basis of the geographical spread and prevalence of recent clusters. Six NCs participated in the study: those in the Republic of Korea, Malaysia, Singapore, South Africa, the UK and the USA.

2.1 SSFFC Detection Algorithm

The data source for this study was the WHO global ICSR database, VigiBase, as of 3 June 2013. In order to facilitate the follow-up of reports and to limit the study to the most recent data, only reports entered into VigiBase after 1 January 2001 were included. As of 3 June 2013, VigiBase included 8 million reports from 113 countries. Reports from studies, clinical trials, special monitoring programmes and the literature were excluded from the analysis.

A systematic review of the Medical Dictionary for Regulatory Activities (MedDRA), version 16.1, was performed, and 78 MedDRA preferred terms (PTs), such as

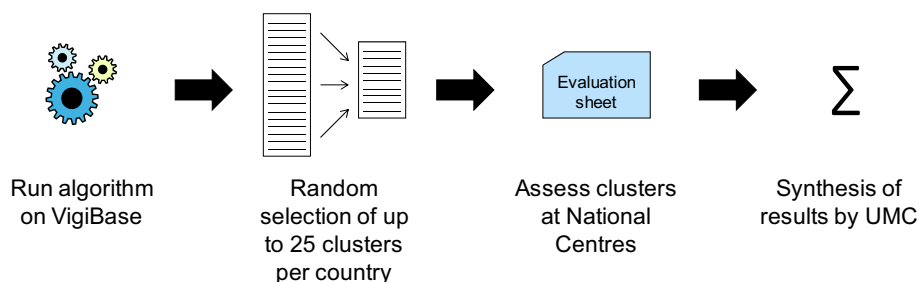
‘drug ineffective’ and ‘product quality issue’, indicative of potentially substandard medicines, were identified. The list was evaluated for clinical relevance and in terms of its accordance with the outlined definition of substandard medicines used in this study. For reports using the WHO Adverse Reaction Terminology, ADRs were mapped to MedDRA before inclusion in the study data set. The full list of terms and their frequency in the study data can be found online in Electronic Supplementary Material 1. We identified 281,000 reports that listed at least one of these terms, and we found that a few countries contributed more than 90 % of these reports. The five biggest contributors were the USA (77 %), Canada (9 %), Germany (2 %), Australia (2 %) and the UK (1 %).

2.1.1 Division of Data Into Clusters

As the basis for detection of suspected events related to substandard medicines, we used the previously developed algorithm [5] for identifying groups of reports on a drug with a higher than expected relative frequency of product inadequacy. These groups of reports are referred to as ‘clusters’ and are made up of the reports on a specific product, a year and a country. The expected count is based on the reporting frequency of product inadequacy only within reports on the same product, to control for the variation between different products’ therapeutic success rates. A ‘product’ in this context is defined as a unique trade name, including the generic names of generic products.

Each report is assigned to a country–year stratum, based on the country of origin of the report and the date of the event. In order to assign each report to the most accurate year stratum, the onset date of the reported event was used whenever available. However, for a large proportion of the VigiBase reports (35 %), no onset date was reported. An algorithm for setting an estimated onset date, corresponding to the most accurate date available, was therefore applied to the reports. When the onset date of the event was not available, the estimated onset date was set on the date when the patient stopped taking the drug. In the case of no reported drug stop date, the date when the NC first received the report was employed, and if none of these dates were

Fig. 1 Overall approach to the empirical evaluation. UMC Uppsala Monitoring Centre



available, the date when the report was first entered into VigiBase was used.

2.1.2 Identification of Stratum-Specific Excess Reporting

The association between a medicinal product and an event can be measured as a contrast between the number of observed events and the expected number of events. Consider the following cross-classification (see Table 1) of the reports in the database according to whether or not they list a given drug (X) and any of the previously defined PTs (Y).

Assuming that the reporting of the product and the medical event are independent of each other, the observed-to-expected (OE) ratio for the association between X and Y [6] can be expressed as:

$$OE = \frac{a}{(a+b) \cdot (a+c)/(a+b+c+d)} \quad (1)$$

Strong associations between the product of interest and product inadequacies in one specific country and year are identified by comparing the OE ratio in that subset with the OE ratios in other countries and other time periods. The OE ratio specific to the selected country/time stratum, OE_{cy} , is computed as:

$$OE_{cy} = \frac{O(s=cy)}{E(s=cy)} = \frac{O_{cy}}{E_{cy}} \quad (2)$$

where O_{cy} is the number of reports in the selected country/time stratum, and E_{cy} is the expected number in the stratum.

The complement is stratified by country and year, and the OE ratio for the complement is computed as a weighted average of the OE ratios for each country–time stratum, s , in the data for all strata except those from cy .

$$OE'_{cy} = \frac{\sum_{s \neq cy} (O_s/E_s) \cdot E_s}{\sum_{s \neq cy} E_s} = \frac{\sum_{s \neq cy} O_s}{\sum_{s \neq cy} E_s} \quad (3)$$

Using the weighted average when computing the expected count for the complement accounts for where and when the drug is used and where and when product inadequacy is reported. We search for countries and time periods where a medicinal product is more strongly associated with reports of product inadequacy than on average for the same medicinal product across other regions and time periods. In practice, we take the ratio

between the OE ratio for the country–time stratum of interest and that in the complement:

$$OE\Delta = \frac{OE_{cy}}{OE'_{cy}} = \frac{O_{cy}}{O'E_{cy}/E'} = \frac{O_{cy}}{E^*} \quad (4)$$

The resulting OE Δ is an OE ratio in its own right. OE ratios are volatile when the observed or expected number of events is small. To prevent highlighting of spurious associations, we employ a shrinkage transformation, using a shrinkage of 0.5, which moderates the ratio towards one:

$$OE = \frac{O + 0.5}{E + 0.5} \quad (5)$$

In this study, we use the previously described information component, IC [6], a bivariate measure based on the OE ratio of the medicinal product and event of interest with the following shrinkage transformation:

$$IC \approx \log_2 \frac{O + 0.5}{E + 0.5} \quad (6)$$

Positive values of the logarithmic measure indicate more reports than expected, whereas negative values indicate fewer reports than expected.

On the basis of the derived OE ratio, a comparative IC value, ICA Δ , is derived:

$$ICA\Delta = \log_2 \left(\frac{O_{cy} + 0.5}{E^* + 0.5} \right) \quad (7)$$

An ICA Δ above 0 indicates that the product has a greater relative frequency of reports on product inadequacy in the country and year of interest than in the database as a whole. A stratum is considered as a potential SSFFC cluster when the lower 95 % confidence interval of the ICA Δ exceeds 0.

As Eq. 4 shows, E^* is undefined for $E' = 0$, which is why the ICA Δ cannot be computed for these strata, and they will be excluded from the computation. This is not likely to influence the performance of the algorithm, as products with $E' = 0$ are such that there is no other country–time period where the product is reported. In other words, there is no relevant background to compare the reporting rate against. In practice, only 42 medicinal products have an E' of 0, and all of these are reported very rarely and in a single country in VigiBase. For $O' = 0$, the ICA Δ value depends solely on O_{cy} , an undesirable situation leading to possible spurious associations. To reduce the risk of spurious associations, drug–ADR combinations with $O' = 0$ were excluded from this analysis.

2.2 Empirical Evaluation with the National Centres

For each of the countries participating in the study, we selected 25 clusters with the lower 95 % confidence interval of the ICA Δ , ICA Δ_{025} , greater than 0. Clusters are

Table 1 Cross-classification of reports

	Y	Not Y
X	a	b
Not X	c	d

Table 2 Clusters included by country

Country	Number of clusters	Years included
Republic of Korea	25	2011–2013
Malaysia	24	2004–2013
Singapore	24	2002–2013
South Africa	25	2009–2013
UK	25 ^a	2011–2013
USA	25	2012–2013

^a Two clusters were identified as doublets, and hence only 24 clusters from the UK are included in the results

selected randomly, but more recent clusters are given priority over less recent ones. For example, if a country has 50 clusters from 2013, 25 will be randomly selected out of those 50. If a country has 20 clusters from 2013 and 20 from 2012, the 20 from 2013 will be selected together with five random clusters from 2012. In total, 147 clusters were assessed (see Table 2).

2.2.1 Cluster Assessment

The structured assessment of clusters performed by the participating NCs aims to distinguish between ‘confirmed’, ‘potential’ and ‘unlikely’ substandard products and to identify the root cause of the disproportional reporting of clusters dismissed as unlikely to be related to a substandard product. Cluster assessment was performed on the basis of a structured evaluation sheet (see Electronic Supplementary Material 2).

The sheet provided information on the product name and year of each cluster to enable identification of the cluster reports from the NC’s database, but it did not include the VigiBase report IDs, so as not to limit the assessment to reports available in VigiBase. In addition to this, it included a number of fields to be filled in by the assessor, most notably the classification of the cluster as a confirmed, potential or unlikely SSFFC. The sheet also included a number of categories for potential alternative explanations of the non-SSFFC clusters, such as ‘switch of brand’ and ‘medication error’, in order to support a structured approach to the cluster assessment and to detect patterns that might help to improve the performance of the algorithm.

The evaluation sheet was supplemented with two complementary questions sent to all countries:

- Has the cluster been investigated as a potential SSFFC at your NC previously?
- Would the cluster be of interest if available in ‘real time’?

All assessments from the NCs were collected and compiled by the UMC. To ensure consistency in the cluster

Table 3 Cluster classification categories

Classification	Description
Confirmed SSFFC	A known SSFFC product event
Potential SSFFC	There are plausible reasons to believe that this may be an SSFFC
Unlikely SSFFC	Compelling reasoning and supporting information that this is not likely related to an SSFFC
Confirmed non-SSFFC	Confirmed by investigation (by the national centre or manufacturer), including quality tests if relevant
Indecisive	Cannot be assessed with enough degree of certainty, because of lack of information

SSFFC substandard/spurious/falsey labelled/falsified/counterfeit medical product

Table 4 Cluster classifications from evaluation by the national centres

Classification	Number	Percentage
Confirmed SSFFC	8	5.4
Potential SSFFC	12	8.2
Unlikely SSFFC	51	34.7
Confirmed non-SSFFC	19	12.9
Indecisive	57	38.8
Total	147	100.0

SSFFC substandard/spurious/falsey labelled/falsified/counterfeit medical product

classifications, a review of the cluster classifications was made, in some cases a revised classification was proposed to the country and the revised classification was reviewed by each country. The root causes of dismissed and not assessable clusters were further analysed by the UMC, for future algorithm optimization. Table 3 lists the cluster classifications that were used.

3 Results

3.1 Algorithm Hit Rate

A total of 147 suspected SSFFC events were evaluated by the participating NCs. Compilation of the cluster evaluations from the NCs shows a hit rate of 5.4 % for confirmed SSFFC clusters and 13.6 % when potential SSFFCs are also included, as shown in Table 4.

When the clusters with too little information for a proper assessment to be made (‘indecisive’) were excluded, the hit rates were 8.9 % (for confirmed SSFFCs) and 22.2 % (when both confirmed and potential SSFFCs were included), respectively (see Table 5).

Table 5 Cluster classifications for assessable clusters from evaluation by the national centres

Classification	Number	Percentage
Confirmed SSFFC	8	8.9
Potential SSFFC	12	13.3
Unlikely SSFFC	51	56.7
Confirmed non-SSFFC	19	21.1
Total	90	100.0

SSFFC substandard/spurious/falsey labelled/falsified/counterfeit medical product

3.2 Potential Value of Algorithm in a Real-Time Setting

The responses to the supplementary questions show that 46 of the clusters had been previously investigated as potentially substandard products by the NCs. Though the majority (35) of these clusters were, in the end, dismissed or classified as indecisive, the NCs made the assessment that it would have been relevant to flag them for review in a real-time setting. Together with the clusters classified as potential or confirmed substandard products, this adds up to 55 clusters of interest for investigation, corresponding to 37 % of all of the assessed clusters.

Four of the six countries provided assessments of whether their respective clusters in the study would be of interest for an investigation in a real-time setting. The response was that 65 out of 97 clusters would be of interest, showing that there is greater interest in the candidate

clusters than the hit rates show. Among the 65 clusters that were reported as being of interest in real time, 20 had not been previously investigated by the NCs, and thus they represent previously unknown clusters that the algorithm uniquely detects.

3.3 True Positives

The clusters that were confirmed as corresponding to events related to a substandard or counterfeit product are listed in Table 6.

The clusters classified as potential SSFFCs are characterized by cases where there is a suspicion but insufficient information to conclude that this is a substandard medicine. These include, for example, previously known cases of suspiciously high reporting rates of product inadequacy where no product quality test results are available, or a case where a single report on a product packaging problem has not been confirmed.

Another characteristic of the potential clusters, compared with the indecisive clusters, is that they contain more specific descriptions of the lack of effect or product quality issues, or they provide some complementary information. One example of the latter is a Spiriva[®] (tiotropium) cluster for which the reported terms were simply ‘drug ineffective’ but which was reinforced by the information that a patient had noted that a capsule did not contain any powder. Examples of clusters where a more precise quality issue supported the case was a clopidogrel cluster with several reports mentioning a rubber smell, or Dulcolax[®]

Table 6 Confirmed SSFFC (substandard/spurious/falsey labelled/falsified/counterfeit medical product) clusters

Country, year ^a	Product (substance)	Comment
UK, 2012	Typhim Vi [®] (typhoid vaccine)	Batch G0530 recalled by manufacturer in September 2012 because it was identified as being at risk of antigen content below specifications
UK, 2012	(Levothyroxine)	Signal identified through pharmacovigilance monitoring in 2011 regarding product substitution issues between levothyroxine brands; subsequent analysis identified manufacturing issues with the Teva brand [7]
UK, 2011	(Levothyroxine)	(See comment in previous row)
USA, 2012	(Quetiapine)	Two cases related to confirmed SSFFC event where company recalled tablets after testing failed dissolution requirements at 3-month time point
USA, 2012	Xanax [®] (alprazolam)	Product confirmed by laboratory as counterfeit; the case was prosecuted and is now closed
USA, 2012	Lo/Ovral [®] (ethinylestradiol/norgestrel)	Company recalled 14 lots from US customers because of inexact counts of active or inert tablets that may have been out of sequence; FDA statement available at http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm289803.htm
USA, 2012	(Atorvastatin)	In Nov 2012, FDA recalled certain lots of Ranbaxy atorvastatin for reported glass particulates; FDA statement available at http://www.fda.gov/drugs/drugsafety/ucm329951.htm
USA, 2012	Ventolin [®] (salbutamol)	Mix of true quality problem discovered by company and lack of effect explained by inadequate administration technique, expired product, etc.

FDA US Food and Drug Administration, *SSFFC* substandard/spurious/falsey labelled/falsified/counterfeit medical product

^a Approximated year of onset

Table 7 False positive findings by category

Category	Number of unlikely SSFFCs	Number of confirmed non-SSFFCs	Percentage of false positives explained
Expected low efficacy	28	7	49
Switch of brand	8	1	13
Medication error	5	1	8
Poor responders	5	1	8
Interaction	5	0	7
Compliance	5	0	7
Switch of regimen	5	0	7
Irrational prescribing	1	0	1

(bisacodyl) for which 10 of 20 reports stated either “product hardly melted” or “the effect decreased compared to previous use”.

3.4 False Positive Findings by Category

The clusters categorized as unrelated to substandard medicines were divided into eight categories specified before the start of the study, depending on the suspected mechanism behind the lack of efficacy, as shown in Table 7, with the most common category (in 35 clusters) being ‘expected low efficacy’. For several of these clusters, the lack of efficacy was assessed as being due to natural progression of disease, antimicrobial resistance or virological failure.

Progression of disease was particularly often associated with oncology drugs, with two examples of this being several clusters of reports on Glivec® (imatinib) in Singapore and two clusters concerning Velcade® (bortezomib) in the UK. The time to progression (or the time of survival) depends on the stage of the cancer being treated, adjuvant therapies administered, co-morbidities of the patient, etc., and are therefore difficult to compare between different settings. Velcade is used in multiple myeloma in combination with other antineoplastic agents, and its efficacy varies greatly with the stage of the disease. Varying treatment regimens between countries, as well as changes in treatment guidelines leading to the use of a product for new indications or new patient groups in one country over time, may influence the reporting of a lack of effect for this type of drug, compared with the background rate, and hence could explain the unexpectedly high reporting rate for Velcade in the UK.

Another factor, not covered in the alternative explanations provided, is more meticulous reporting of a product through programmes for soliciting or encouraging reporting. One example is the Glivec clusters in Singapore, where the majority of the reports were submitted through such a programme. Though these programmes were not directed specifically at lack of effect, it might be expected

that lack of effect is subject to greater underreporting than other events, and so the relative increase is more significant.

Other common explanations included a switch of brand, where a change in the efficacy of the treatment could be caused by varying pharmacokinetic parameters of different products containing the same substance, poor responders and medication errors. For example, a surge in the number of reports of lack of efficacy of Mircera® (methoxy polyethylene glycol-epoetin beta) in Singapore had been found by the NC to likely reflect inappropriate use due to lack of understanding of its dosing soon after the product launch.

4 Discussion

Our empirical evaluation shows that some events where substandard medicines have led to patient harm can be detected through the existing global pharmacovigilance system. At the same time, a large proportion of the clusters of suspected product inadequacy in individual case reports lack enough information to allow for a conclusive classification, reflecting a significant challenge. The precision and timeliness in reporting suspected product inadequacy must improve in order for the global pharmacovigilance system to be of broader value in the detection of harmful substandard medicines.

As has been shown in previous studies [8], reports on lack of effect can often be attributed to causes other than substandard products. While the majority of the highlighted clusters in our study did not correspond to substandard medicines, the eight that did illustrate the potential for the method to act as an early screening tool. They include a counterfeit product, a product substitution issue coupled with manufacturing problems, failure to meet product specifications, contamination, delivering device failure and more. All in all, nearly a third of all highlighted clusters had been previously investigated as suspected substandard medicines by the respective pilot countries, and, even though these suspicions had sometimes been dismissed, it

shows the accuracy of the proposed method in detecting suspected substandard medicines. The future potential is further reinforced by the fact that out of the clusters that had not already been evaluated by the respective country, two in five would have been of interest for follow-up in real time. Many of the countries in this study already have active monitoring of substandard medicines, hence the potential value of using the algorithm may be even greater for countries with no or little current monitoring activity.

In considering the results, one should bear in mind the retrospective nature of our study and the risk that public awareness of cases of substandard medicines might have led to increased rates of reporting of product quality issues, after they were known. For example, for the Ranbaxy cluster (see Table 6), lots were recalled by the US Food and Drug Administration (FDA) in 2012 because of the presence of glass particles [9], and the stimulated reporting for product quality issues afterwards might have led to its detection in this study. It should also be noted that some clusters that match confirmed cases of substandard medicines also include several reports with other explanations for the suspected product inadequacy and could represent coincidental findings.

Many of the highlighted clusters that were confirmed as not representing substandard medicines had already been evaluated and dismissed by the NCs. Around half of the false positives were related to drugs with low expected therapeutic success rates, such as antineoplastic agents administered as second- or third-line therapy to cancer patients in advanced stages. The method that was used does adjust for the overall reporting rate of lack of effect in VigiBase for the drug of interest, which provides a proxy for the expected rate of therapeutic failure. However, for many of the false positives, the increases in the relative reporting of suspected product inadequacies may have resulted from increased awareness and reporting in the context of post-authorization safety studies and solicited reports through patient assistance programmes, or from local variations in clinical practice. As an example, the proteasome inhibitor Velcade is indicated for the treatment of multiple myeloma and mantle cell lymphoma in combination with other antineoplastic agents. As its efficacy varies greatly with the stage of the disease, the patient population that is selected and the indication for treatment in a specific region will largely determine the rate of reporting of lack of effect. Therefore, the higher rate of reporting of product inadequacy for Velcade that was detected in the UK may be explained by its use in more ill patients.

A number of the highlighted clusters were noted to reflect medication errors or switches of brands or regimens, and were classified as false positives for the purpose of the study. It should be noted that these may be relevant findings in the broader context of medicine safety.

We did not have access to a reference set of known cases of harm to patients from substandard medicines. As a result, this study could not be designed to evaluate the proportion of real cases of substandard medicines that would be correctly recalled and how many would be missed. However, in an earlier phase of this research, we identified a number of historical cases of substandard medicines that would not have been detected by the method [5]. In some of those cases, there were no relevant reports in VigiBase—for example, if the country in question was not a member of the WHO programme for international drug monitoring, at the time. In other cases, the reports in VigiBase reflected an unexpected therapeutic response—for example, related to a contaminant, which would only be accounted for by the method if the reporter had explicitly noted the unexpected therapeutic response as a separate adverse reaction term on the report.

The major obstacle in our study, and perhaps the greatest challenge for prospective use, was the difficulty of completing a root cause analysis for the reports in the suspected clusters, which led to the classification of more than one in three clusters as being indecisive. This stems from the lack of information, such as the product name, batch number and case descriptions, available on the reports. A second challenge is the difficulty (or impossibility) of contacting the reporter for additional details. The time that elapses from when the adverse event occurs to when the individual case report is sent to the NC database and is then reported to VigiBase amplifies these challenges. This delay ranged from a few weeks to over a year and would be a major issue in real-world use, since it severely hinders the ability to take action on detected SSFFC cases. Another challenge is that many reports list only the active ingredient and not the trade name of the product used, which impedes quality investigations. This is perhaps further amplified by the practice in many countries of prescribing medicines at the level of the active ingredient and not the product, and of substituting products on the basis of their availability, in public health programmes.

If the long lead times for reporting decrease and the precision of information on reports, the availability of samples and laboratory capacity, and the capability to contact the original reporters improve in the future, analysis of pharmacovigilance data might be considered for real-world use, as a complement to other initiatives for detection of substandard medicines. If so, further improvements of the method should be explored. For example, there may be ways to filter out, or at least highlight, clusters related to ‘expected low efficacy’ on the basis of the overall relative reporting rate of lack of effect for that drug in VigiBase. Similarly, reports that originate with the same reporter, the same study, etc., may also be marked for transparency in the manual review. Another potential

extension of the method would be to adapt the selection of terms used to identify inadequate products depending on the substance (e.g. including ‘unintended pregnancy’ only for contraceptives), indication (looking for cases where the suspected ADR matches the indication for treatment) and formulation. It would also be desirable to have a more flexible division of reports into clusters where years or countries with few reports could be grouped together and countries with many reports could be divided into smaller regions. A potential further refinement of the cluster definition would be to separate reports on the same product on the basis of the drug formulation. Yet another area of future research may be natural language processing to identify relevant reports that do not list one of the relevant adverse event terms, on the basis of free text case descriptions.

Contamination of a product might be detected from clinical symptoms related to the contaminant. Examples are slimming products adulterated with thyroid hormone—causing tachycardia, arrhythmia, sleeplessness or diarrhoea—or herbal medicines produced from herbs harvested in an environment likely to cause contamination with heavy metals, such as roads with heavy traffic—resulting, for example, in lead poisoning of the patients taking them. The development of methods to identify uncharacteristic adverse events related to unexpected therapeutic effects of substandard medicines should be further explored. To this end, an extension of the method proposed by DuMouchel et al. [10] to include the geographic dimension might be explored; they searched for increased reporting rates of an adverse reaction, compared with the historical reporting rates of that adverse reaction with the drug of interest.

In our study, the computational analysis was performed in VigiBase, the WHO global database. The methodology can be applied to other settings—for example, to the database of a pharmaceutical company or of a larger country, which can be divided into smaller regions for the purpose of analysis. The advantage of a global database in this context is the statistical power to detect signals early, which derives from its size and scope. However, it is possible that shorter lead times in reporting and more direct access to the original reporter would decrease the proportion of indecisive clusters in company or country databases. A more uncertain venture would be to adapt and apply our method to social media content. This may provide an interesting alternative information source, not suffering from the delays described above. As has been noted before, social media might lend itself more naturally to causality assessment of substandard medicines, relying on root cause analysis, than to assessment of adverse reactions, which relies on epidemiological analysis [11]. However, root cause analysis would require that signals can be traced back to the source, which may present ethical and technical challenges.

5 Conclusion

The application of the algorithm to VigiBase allowed identification of some cases of substandard medicines; however, some key prerequisites for broad prospective detection of substandard medicines are not currently fulfilled. Such key factors are fast handling and transfer of incoming reports into VigiBase, detailed information on the product and its distribution channels, the possibility of contacting primary reporters for further information, the availability of samples of suspected products and laboratory capacity to analyse suspect products.

Acknowledgments The authors would like to thank Shanthi Pal and Michael Deats from the World Health Organization, and Sten Olsson and Ralph Edwards from the Uppsala Monitoring Centre, for their valuable feedback during the study; and the following people for their valuable contributions to the cluster evaluations at the national centres: Song-yi Han from the Ministry of Food and Drug Safety in Korea; Soo-youn Chung, Ye-jee Kim and Yeon-ju Woo from the Korea Institute of Drug Safety and Risk Management; Sally Soh, Tham Mun Yee, Tan-Koi Wei Chuen, Patricia Ng and Dorothy Toh from the Health Products Regulation Group in Singapore; Sameerah Shaikh Abdul Rahman, Norleen Mohamad Ali and Nafiza Ismail from the Ministry of Health in Malaysia; Farzana Khan, Ndimu Dowelani and N. O’Connor from the Medicines Regulatory Authority in South Africa; Phil Tregunno, Charlotte Goldsmith and Brian Burch from the Medicines and Healthcare Products Regulatory Agency in the UK; and Melissa Kim, Robert Ball, Solomon Iyasu, Scott Proestel, Robert Levin, Ilisa Bernstein, Thomas Christl and Eleni Anagnostiadis from the US Food and Drug Administration.

Funding Financial support for the algorithm development was provided through the Monitoring Medicines Project (FP7 Grant No. 223566). The empirical evaluation was conducted with financial support from the US Food and Drug Administration (Collaboration and Capacity Building Ensuring Global Product Safety and Quality cooperative Agreement No. 1U01FD005031-01).

Conflict of interest Kristina Juhlin, Ghazaleh Karimi, Maria And er, Mukesh Dheda, Tan Siew Har, Rokiah Isahak, Su-Jung Lee, Sarah Vaughan, Pia Caduff and G. Niklas Nor en have no conflicts of interest that are directly relevant to the content of this study. Sara Camilli reports that the study was partially financed by a grant from the US Food and Drug Administration, of which she is an employee.

Theme issue This article is part of a theme issue co-edited by Elliot G. Brown, Shanthi Pal and Sten Olsson. No external funding was used to support the publication of this theme issue.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Blossom DB, Kallen AJ, Patel PR, et al. Outbreak of adverse reactions associated with contaminated heparin. *N Engl J Med.* 2008;359(25):2674–84.

2. Kishimoto TK, Viswanathan K, Ganguly T, et al. Contaminated heparin associated with adverse clinical events and activation of the contact system. *N Engl J Med*. 2008;358(23):2457–67.
3. US Food and Drug Administration. Information on heparin. 2012. <http://www.fda.gov/drugs/drugsafety/postmarketdrugsafety/informationforpatientsandproviders/ucm112597.htm>. Accessed 16 Sept 2014.
4. European Medicines Agency. Questions and answers on heparins. 2008. http://www.ema.europa.eu/docs/en_GB/document_library/Medicine_QA/2009/11/WC500015169.pdf. Accessed 16 Sept 2014.
5. 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.pdf> 2011. Accessed 5 Feb 2015.
6. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res*. 2013;22(1):57–69.
7. Medicines and Healthcare Products Regulatory Agency. Levothyroxine tablet products: a review of clinical and quality considerations. 2013. <https://www.gov.uk/drug-device-alerts/drug-alert-levothyroxine-100mcg-tablets-recall-of-teva-uk-and-numark-branded-product>. Accessed 11 Feb 2015.
8. González CPV, Martínez RPD, Gutiérrez JLL, Pedraza RS, Figueras A. Algorithm for the evaluation of therapeutic failure reports—proposal and pilot analysis. *Pharmacoepidemiol Drug Saf*. 2013;22(2):199–206.
9. US Food and Drug Administration. FDA statement on the ranbaxy atorvastatin recall. 2012. <http://www.fda.gov/Drugs/DrugSafety/ucm329951.htm>. Accessed 5 Feb 2015.
10. DuMouchel W, Yuen N, Payvandi N, Booth W, Rut A, Fram D. Automated method for detecting increases in frequency of spontaneous adverse event reports over time. *J Biopharm Stat*. 2013;23(1):161–77.
11. Norén GN. Pharmacovigilance for a revolving world: prospects of patient-generated data on the internet. *Drug Saf*. 2014;37(10):761–4.