

## RESEARCH ARTICLE

# Reports of substandard medicines: a lexicographic analysis of the Brazilian Health Surveillance Report System

Mayra Bruna Fedalto<sup>1</sup>, Antonio Eduardo Matoso Mendes<sup>1</sup>, Fernanda Stumpf Tonin<sup>1,2</sup>,  
Fernando Fernandez-Llimos<sup>3</sup>, Roberto Pontarolo<sup>1\*</sup>

<sup>1</sup>Programa de Pós-graduação em Ciências Farmacêuticas, Universidade Federal do Paraná (UFPR), Curitiba, PR, Brasil

<sup>2</sup>Research Institute for Medicines (iMed.Ulisboa), Department of Social Pharmacy, Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal

<sup>3</sup>Laboratory of Pharmacology, Department of Drug Sciences, Faculty of Pharmacy, University of Porto, Porto, Portugal

\*Corresponding author: pontarolo@ufpr.br

## Abstract

Regulatory agencies are responsible for collecting and evaluating spontaneous reports of suspected problems related to medications, including those from substandard medicines (SM).

**Objectives:** The aim was to evaluate the profile of SM reports submitted to the Brazilian Health Surveillance Notification System (Notivisa) and classify these reports objectively by means of lexicographic analysis. **Methods:** Was extracted all SM reports available in Notivisa during the period 1 January 2007 to 31 December 2017. Descriptive statistics were performed and the reasons for SM reporting were standardized (using OpenRefine and Microsoft Excel). The following analyses were performed using IRAMuTeQ 0.7 alpha2: lexicographic analysis to obtain the frequency of active words; descending hierarchical classification (DHC) to categorize the active words into lexical classes; factorial correspondence analysis (FCA) to obtain graphs of the classes. Approved by the Ethics Committee of the Hospital do Trabalhador/SES/PR CAAE 81873417.3.0000.5225 (protocol number: 2.506.594). **Results:** A total of 61,775 reports were analyzed, most of them reported by hospitals (46%). The DHC of the reasons for SM produced four classes visualized in the FCA: (i) packaging problems (16%) mainly leakages/opening issues; (ii) inadequate drug identification (22%), such as illegible label information; (iii) stability and contamination issues (11%) such as presence of particles; (iv) damaged tablets/blisters (23%) mainly broken tablets. Most SM (52%) were solutions for parenteral use; sodium chloride (9%), glucose and dipyrrone (3%) were the products with most complaints. **Conclusions:** The reasons for SM reporting can be objectively classified into classes that represent the main problems submitted to Notivisa. This classification could guide the standardization of SM reporting and contribute to improving surveillance reporting systems worldwide.

**Keywords:** Bibliometrics. Public Health Surveillance. Substandard Medicine (Substandard Drugs).

## How to cite

Fedalto MB, Mendes AEM, Tonin FS, Fernandez-Llimos F, Pontarolo R. Reports of substandard medicines: a lexicographic analysis of the Brazilian Health Surveillance Report System. *Rev Ciênc Farm Básica Apl.* 2021;42:e753. <https://doi.org/10.4322/2179-443X.0753>

Financial support: MBF, FST - received scholarship the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. AEMM, FFL, RP - None.

Conflicts of interest: MBF, AEMM, FST, FFL, RP declare no conflict of interest.

The study was carried out at Universidade Federal do Paraná (UFPR), Curitiba, PR, Brasil.

Received on August 30, 2021. Accepted on October 18, 2021



Copyright © Fedalto et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

Substandard medicines (SM), defined by the World Health Organization as pharmaceutical products that do not meet their quality standards or specifications<sup>1</sup>, constitute a serious public health problem, especially in emerging countries with non-existent or emerging regulatory agencies<sup>2-9</sup>. Nonetheless, developed countries such as Canada<sup>10</sup>, the United States of America<sup>11-13</sup>, and the United Kingdom<sup>14</sup> often report issues related to SM. The use of SM can lead to adverse events, therapeutic failure, drug resistance, disease progression, and death. In recent years, an increase in SM-related incidents of over 30% has been documented, with around half of all SM detected in the legitimate supply chain affecting life-saving therapies<sup>15-19</sup>. Additionally, SM are responsible for a significant amount of marketed drug recalls due to quality deviations<sup>10,11,20</sup>, which causes financial losses for the consumer and the manufacturer and may diminish overall confidence in the healthcare system<sup>3</sup>.

The evaluation of spontaneous notifications (i.e. reports) of suspected problems related to medications, usually carried out by institutions, health professionals, managers, and citizens, is one of the main research branch activities of pharmacovigilance<sup>21,22</sup>. Regulatory agencies around the world are responsible for collecting and evaluating these spontaneous reports<sup>23</sup>. To design corrective actions and preventive public health policies that tackle SM-related issues, gathering accurate data and performing robust research are crucial<sup>21</sup>. In Brazil, substandard medicines reports are collected and classified according to their characteristics (e.g. type of report, notifier, outcomes) by the National Notification System for Health Surveillance (Notivisa) database<sup>24</sup>

Although methods for SM reporting have evolved, with recent studies describing the use of algorithm stratification analysis and data mining, or surveillance guided by geospatial analyses<sup>25-27</sup>, no standard model for the final analysis and classification of these reports exists. Usually, spontaneous SM reports are presented with descriptive information about the product and the problem found, written in natural language by the reporter (e.g. healthcare professional, researcher, patient). Previous studies attempted to evaluate and classify SM reports by means of manual data extraction and reading<sup>10,14,16,28</sup>. Depending on the number of reports, these manual and person-based techniques may be unrealistic, as well as exposed to different types of bias and heterogeneity. In this case, the use of lexicographic analysis, an innovative and broad approach to content analysis, could be an alternative to objectively identify the semantic and paradigmatic relationships existing within the vocabulary (lexicon) used in the SM reports<sup>29</sup>. This type of analyses allows insightful look at the studied variables, including the categorization process<sup>30</sup>. Thus, the aim of this study was to classify the different types of quality deviation of medicines reported to the Brazilian Pharmacovigilance System during a 11-year period, by means of lexicographic analysis.

## MATERIAL AND METHODS

### Data collection and standardization

This was an exploratory study of the SM reports available in the Brazilian National System of Notifications for Health Surveillance (Notivisa) from the Brazilian Drug Regulatory Agency (Anvisa), from 1 January 2007 to 31 December 2017. Access to the information was provided by the Health Surveillance Service of the Department of Health in Paraná State (Southern Brazil). An SM was here defined as a product with a problem observed prior to causing harm to a patient's health, such as suspected quality misuse, unregistered product, counterfeit product, or unauthorized company. Only SM reports classified by Notivisa as "Technical Complaint" and "Suspected Quality Deviation Product" were exported to Microsoft® Office Excel (Microsoft, Redmond, WA) for analyses.

The information collected from the SM reports included: date of report, reporter occupation and institution, region of reporter, origin of the product, product physical form (i.e., pharmaceutical form and administration route), active substance, and reason for

reporting (open field to allow an unstructured comment describing the SM characteristics). To normalize the information, pharmaceutical forms and administration routes were standardized based on the lists of drugs registered in Anvisa<sup>31</sup>, and active substances were standardized according to the ATC Index ([https://whocc.no/atc\\_ddd\\_index/](https://whocc.no/atc_ddd_index/)). Typing errors and wrong words in the field 'reason for reporting' were corrected and standardized using OpenRefine software (<https://openrefine.org/>) by applying the Text Facet tools.

## Data analysis

After standardization of the information contained in the SM reports, descriptive statistical analyses were performed for all the categorical variables and results were reported as absolute and relative frequencies.

The text containing the 'reason for reporting' for all SM reports was assembled into a single text file named the text corpus, which was analyzed in IRAMuTeQ 0.7 alpha 2 (Interface R for Analyses of Multidimensional Texts and Questionnaires)<sup>30</sup>.

Based on Reinert's method<sup>32-34</sup> this software uses algorithms that reorganize the original text corpus based on the repetition and sequence of words. Before the lexicographic analysis itself, the text is prepared according to the following steps: 1- Lemmatization of the text corpus, which consists of the replacement of a word by its root term (e.g., 'ampoules' by 'ampoule'; or 'industries' by 'industry'). This process removes the inflectional ending of the word in order to normalize the text and decreases its complexity without compromising the precision; 2- Reduction of the text corpus into text segments - represented by each notification from the database; and 3- Regrouping of the text segments into context units in order to contain a minimum number of two active words<sup>29,35,36</sup>.

After this process, the following analysis were conducted:

(i) Lexicographic analysis – determination of the frequency and distribution of active words (i.e. words with individual meaning such as nouns and verbs) and supplementary words (i.e. used to assist in the construction of the text such as articles);

(ii) Descending Hierarchical Classification (DHC) – categorization of the active words into similar lexical classes and subclasses, which were depicted in dendrograms. This analysis considers the frequency and position of a word occurring in the text, starting from a unique cluster and successively dividing it into subgroups. The algorithm starts by assuming that all active words initially belong to the same class (single cluster) and divide it sequentially, deriving the two most distinct word classes from the text. That is, an active word partition from the text corpus is created by maximizing interclass inertia and minimizing intraclass inertia. Inertia is defined as a measure of variation between individual profiles around the average profile. Thus, the greater the differences between the words, the greater the inertia between them. Row inertia and column inertia were used to assess associations between categories and contributions to variation in the data in a contingency table. Higher values generally indicate a stronger association and greater deviation from expected values. This interactive process is interrupted when the interclass inertia is not enhanced by a new data partition<sup>29,37</sup>. The extent of the connection between each active word and class or subclass was calculated by chi-square statistic ( $\chi^2$ ). The greater the chi-square value, the more likely the hypothesis of dependence between the active word and the class. However, this test does not show how these two variables are related, so a correspondence analysis should also be performed;

(iii) Correspondence analysis – used to create graphs by a correspondence factorial analysis made from the DHC, which allow the classes visualization and the proximity between similar according to their lexicon. This analysis decomposes the inertia by identifying a small number of mutually independent dimensions that represent the most important deviations from independence. Dimension 1 represents the largest amount of explained inertia from independence; dimension 2, the second largest, and so on. Dimensions are formed by

identifying those axes for which the distance between the profiles and axes is minimized while simultaneously maximizing the amount of explained inertia<sup>30,37</sup>. Categories with similar distribution are represented as points close in space, and categories that have very dissimilar distributions are positioned far apart. These plots reveal the contrasting use of vocabulary in the different lexical groups and the proximity of lexical items inside each of the groups; it is not a matter of counting occurrences, but of relations among words. The graphs are interpreted by the opposition of the axes<sup>32-34,38,39</sup>.

Robustness of IRAMuTeQ analyses were evaluated considering: (i) the minimal retention of 70% of the text segments in lexicographic analysis, (ii) a sum between axis factors of the correspondence analysis graphs of the subclasses close to 100, and (iii) chi-square values higher than 5 in the DHC analyses which represents good separation between the classes<sup>29,33</sup>.

In order to demonstrate the usability and robustness of lexicographical analysis to identify SM reports objectively, we conducted a validation analysis in which the 25 main active words obtained per class were used to classify all the reports from the Notivisa database. Names were assigned arbitrarily to the classes and subclasses of SM reports by the research team, considering the main active words obtained during the analyses and based on terms previously described in the literature<sup>10,14,16,28</sup>.

## RESULTS

Overall, 66,439 reports were initially extracted from Notivisa for the 11-year period evaluated, of which 61,775 (93.0%) were documented by the reporter as an SM report. Of these SM reports, most (68.3%) were documented from 2012 to the present. Reports were submitted mostly by medical centers (45.6%) and pharmacists (11.4%). The majority of SM mentioned in the reports were produced in Brazil (76.1%); among imported products, almost 15% were manufactured in India. Most of these products were solutions (47.1%) or tablets (27.6%), with 52.0% for parenteral administration (52.0%). The total number of reports included 1202 different active substances. Among all the products, problems with sodium chloride (9.3%), followed by glucose (3.2%) and dipyrone (1.9%), were the most commonly reported. Further information on the SM is available in Table 1.

**Table 1** Characteristics of substandard medicines reports

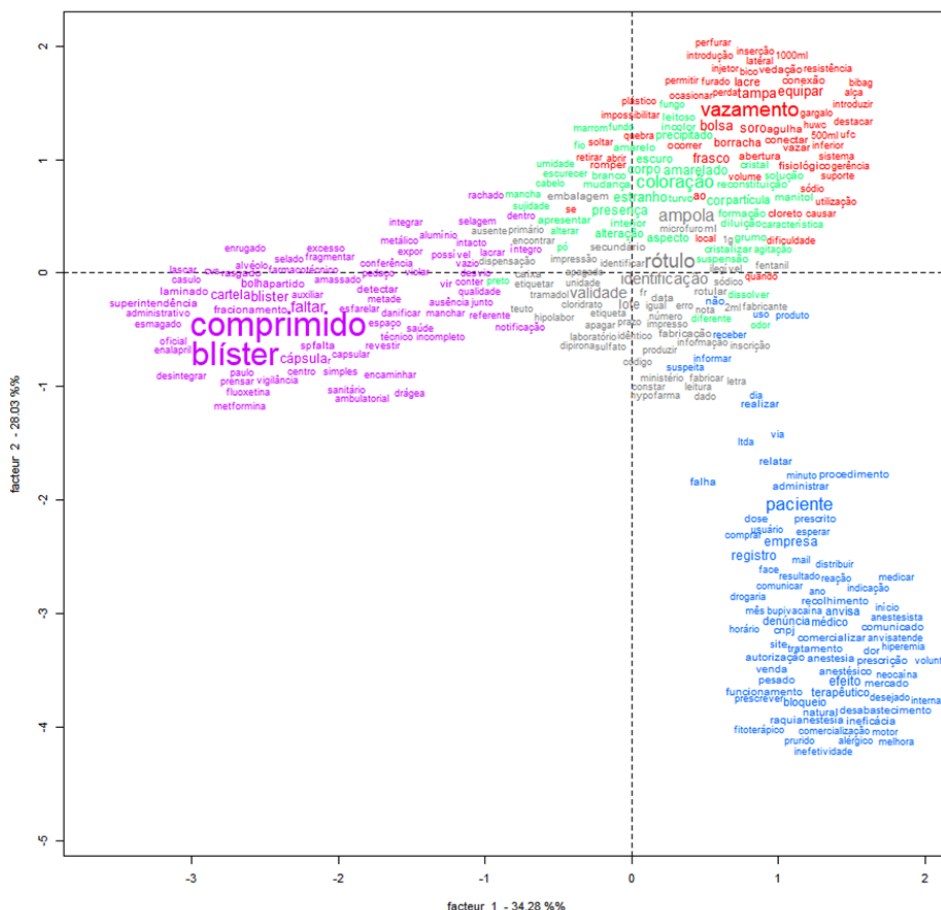
<i>Characteristics</i>	<i>n (%) notifications</i>
Year of notification	
2017–2016	14,005 (22.7%)
2015–2014	14,519 (23.5%)
2013–2012	13,680 (22.1%)
2011–2010	11,055 (17.9%)
2009–2007	8516 (13.8%)
Reporter's institution	
Hospital/medical services	28,160 (45.6%)
Anvisa and regional surveillance	11428 (18.5%)
Social assistance	4503 (7.3%)
Research center/university	4372 (7.1%)
Pharmacy or drugstore	1642 (2.7%)
Industry	761 (1.2%)
Distributor	126 (0.2%)
Others	17986 (17.4%)
Main pharmaceutical form	
Liquid solution	29,074 (47.1%)
Tablet/pill	17,032 (27.6%)
Powder	6821 (11.0%)
Capsule	3428 (5.5%)
Others/ Not Reported	5420 (8.8%)
Actives	

Table 1 Continued...

<b>Characteristics</b>	<b>n (%) notifications</b>
Synthetics	39,721 (64.3%)
Fluid therapy	10,970 (17.8%)
Biological	2130 (3.4%)
Herbal	226 (0.4%)
No therapeutic use	1283 (2.1%)
Not Reported	7445 (12.0%)
Reporting region	
North	1226 (2.0%)
Northeast	7745 (12.5%)
Central-West	5958 (9.6%)
Southeast	29,716 (48.1%)
South	16,281 (26.4%)
Not Reported	849 (1.4%)
Main notifier profession	
Pharmacist	7029 (11.4%)
Engineer	60 (0.1%)
Physician	56 (0.1%)
Nurse	53 (0.1%)
Others/ Not Reported	54577 (88.3%)
Origin of medicines	
National	46,994 (76.1%)
Imported	1705 (2.8%)
India	250 (14.7%)
Others/ Not Reported	13076 (21.1%)
Main administration route	
Parenteral	32,143 (52.0%)
Oral	22,883 (37.0%)
Topical	1081 (1.7%)
Others/ Not Reported	5668 (9.3%)
Top 5 general actives	
Sodium chloride	5773 (9.3%)
Glucose	2003 (3.2%)
Dipyron	1192 (1.9%)
Water for injection	981 (1.6%)
Omeprazole	961 (1.6%)
Others/ Not Reported	50866 (82.3%)

An initial lexicographical analysis of the text corpus with the 'reasons for reporting' revealed a total of 24,205 active words (82.0%) and 5305 additional words (18.0%). However, after standardization of the active words, the final lexicographic analysis of the corrected text corpus revealed 22,553 (81.0%) active words. The DHC analysis of the active words produced five lexical classes, with  $\chi^2$  values over 5. A list of the top 25 active words per class with their respective  $\chi^2$  values are available in Table S1 and S2 in Supplementary Material. Figure 1 shows the results of the correspondence analysis, with categorization of the active words in the five lexical classes. An in-depth analysis of these classes demonstrated that SM reports were correctly classified into Classes 1, 2, 3, and 5. Class 4 corresponded to adverse event reports that were probably misclassified by the reporter as SM, but correctly identified by the lexicographic analyses.

Class 1, representing 16.0% of all reports, corresponds to 'Packaging problems' and contains terms such as 'leakage' and 'difficulty opening'. Class 2 (21.7% of reports) corresponds to inappropriate 'Identification and content', being represented by terms such as 'illegible information'. Class 3 (10.9% of reports) refers to 'Contamination and stability' issues, containing terms such as 'presence of particles'. Class 5 (22.9% of reports) refers to 'Damaged tablets/blisters'.



**Figure 1.** Correspondence analysis plot of the most frequent active words in each of the lexical groups. Sizes of active words are proportional to their frequency in the text. Colors correspond to the lexical classes (Class 1: red; Class 2: grey; Class 3: green; Class 5: purple; Exclusion (class 4): blue). The four groups closest to each other (1, 2, 3 and 5) refer to reports of substandard medicines. Class 4, more distant, that is, less related to the other groups, corresponds to notifications that were excluded from this study.

The new DHC analysis of Classes 1, 2, 3, and 5 corresponding to SM reports generated 10 subclasses, according to the dendrogram presented in Figure 2. Subclasses of Class 1 ('Packaging problems') were classified into problems with 'Aperture', 'Bulk', and 'Leakage', covering reports such as difficulty in opening the drug, seal disintegration, improper seals, packaging holes, and leakage problems. Subclasses of Class 2 ('Identification and Content') were named 'Identification', 'Content', and 'Label', containing issues such as missing information on the label or vial, unreadable information (e.g. batch, date of manufacture, expiration date, dosage, name of the asset, among others), printing errors, empty medicine packaging, wrong number of vials, and unlabeled or unglued medicine. Class 3 was divided into 'Contamination' and 'Stability', characterized by changes in color or appearance, and the presence of particles or foreign bodies. Class 5 ('Damaged tablets/blisters') consists of the subclasses 'Quantity and integrity' and 'Amount of blisters', with reported problems including broken tablets, blister packs without or missing tablets/capsules, blisters with decayed aluminum foil, and wrinkled blisters. The frequency, description, main characteristics, and active words for each subclass are presented in Table 2. The lists of the top 25 active words with the highest  $\chi^2$  for each subclass are available in Tables S3, S4 and S5 in Supplementary Material. The correspondence analysis of Classes 1 and 2 (Figure S1 and Figure S2, respectively) showed good separation (sum of the axis close to 100) and are presented in Supplementary Material.

**Table 2** Frequency and description of quality problems for each class of substandard medicines

Subclass	n (%)*	Description	Main active words	PF n (%)	AR n (%)	ATC Ther. n (%)
<b>CLASS 1 Packaging problems</b>	10,608 (16.0%)	-	-	Solution n = 8488 (80.0%)	Parenteral n = 8520 (80.3%)	B05 n = 5350 (49.9%)
1.1 Aperture	3110 (4.7%)	Difficulty opening or breaking ampoule. Risk of breakage when opened. Risk at the time of use.	Ampoule/Difficulty/ Aperture/Breaking/Risk/ When/Open/Use	Solution n = 2151 (69.2%)	Parenteral n = 2728 (87.8%)	B05 n = 1036 (33.3%)
		Inserting the needle into the ampoule, the sealing rubber breaks up, falling into the medicine/sealing rubber loose on medication.	Rubber/Needle/Equipo/ Introduce/Medicine/ Fragment/Fall/Extricate			
		Difficulty connecting to equipment.	Equipo/Difficulty/ Connect			
1.2 Bulk	2536 (3.8%)	Lower volume of solution/ volume below that specified.	Volume/Lower/Solution/ Less/Specified/Below	Solution n = 2281 (89.9%)	Parenteral n = 2330 (91.9%)	B05 n = 1887 (74.4%)
		Punctured.	Bag/Punctured			
		Empty packaging or bag.	Bag/Packaging/Empty			
1.3 Leakage	2352 (3.5%)	Leakage of the product. Leakage in the sealing region, in stopper, in bottleneck, in lateral solder.	Leakage/Product/Sealing /Stopper/Region/ Bottleneck/Lateral/ Solder/Body/Junction	Solution n = 1902 (80.9%)	Parenteral n = 1325 (56.3%)	B05 n = 1018 (43.3%)
		Bottle with crack.	Cracked/Crack			
		Bottle with integral seal showing clue of leakage.	Leakage/Present/Seal/ Integrate/Clue			
		Leakage of content affecting visualization of label.	Leakage/Content/ Visualization			
<b>CLASS 2 Identification and content</b>	14,399 (21.7%)	-	-	Solution n = 9275 (64.4%)	Parenteral n = 10,690 (74.2%)	J01 n = 2015 (14.0%)
2.1 Identification	3590 (5.4%)	No batch number, expiration date, date of manufacture, drug name, company, or illegible information.	Batch/Expiration date/ Fabrication/Date/ Number/Information/ Company/Illegible	Solution n = 1951 (54.3%)	Parenteral n = 2114 (58.9%)	J01 n = 500 (13.9%)
		Error printing labels.	Error/Print/Label			
2.2 Content	4040 (6.1%)	Sealed packaging, box, or ampoule without contents inside.	Packaging/Seal/Box/ Empty/Inside/Content/ Ampoule	Solution n = 2727 (67.5%)	Parenteral n = 2955 (73.1%)	B05 n = 556 (13.8%)
		Sealed packaging/boxes without medication or with less quantity. When opening the secondary packaging, a smaller amount of ampoules were detected.	Box/Packing/Amount/ Secondary/Detect/Lower / Inferior/Open/Ampoule			
2.3 Label	3503 (5.3%)	Label or tag taken off.	Label/Tag/Take off	Solution n = 2303 (65.7%)	Parenteral n = 2954 (84.3%)	J01 n = 582 (16.6%)
		Lack of identification. No identification label.	Label/Identification			
<b>CLASS 3 Contamination and stability</b>	7218 (10.9%)	-	-	Solution n = 4036 (55.9%)	Parenteral n = 5226 (72.4%)	B05 n = 1558 (21.6%)
3.1 Stability	4013 (6.0%)	Color change, coloration. Color change from colorless, light to yellow. Solution yellowish.	Coloring/To present/ Alteration/Yellowish/ Yellow/Color	Solution n = 1692 (42.2%)	Parenteral n = 2794 (69.6%)	J01 n = 978 (24.4%)
		Color change after dilution or reconstitution.	Alteration/Color/ Dilution/Reconstitute			
		Different aspect, yellowish aspect, turbid, milky.	Aspect/Different/Milky/ Turbid/Yellowish			
3.2 Contamination	2418 (3.6%)	Presence of foreign body in the solution. Black dots in solution in closed, sealed ampoule.	Foreign/Body/Presence/ Ampoule/Seal/Black	Solution n = 1833 (75.8%)	Parenteral n = 1852 (76.6%)	B05 n = 896 (37.0%)
		Presence of piece of glass/plastic.	Presence/Plastic/Piece			
		Suspended particles in ampoules.	Particle/Ampoule			

Table 2 Continued...

Subclass	n (%)*	Description	Main active words	PF n (%)	AR n (%)	ATC Ther. n (%)
		Presence of fungus in solution.	Presence/Fungus			
<b>CLASS 5 Damaged tablets/blisters</b>	15,219 (22.9%)	-	-	Tablet n = 12,186 (80.1%)	Oral n = 14,553 (95.6%)	N06 N = 1381 (9.1%)
5.1 Quantity and integrity	12,537 (18.9%)	Blister with missing tablet. Sealed blister without tablet.	Tablet/Blister/Cartouche/Seal	Tablet n = 10,673 (85.1%)	Oral n = 12,436 (99.2%)	C09 n = 1234 (9.8%)
		Two tablets in the same blister.	Tablet/Bubble/Inside			
		Broken, crumbled, missing piece. Wrinkled, cracked tablet in full blister.	Tablet/Blister/Broken/Piece/Crumble/Wrinkled/Cracked			
		Crumpled blister. Cartouche with broken metal part.	Blister/Wrinkled/Cartouche/Metallic			
5.2 Amount of blisters	2169 (3.3%)	Excess blisters in the box.	Box/Excess	Tablet n = 1296 (59.8%)	Oral n = 1672 (77.1%)	N06 n = 354 (16.3%)

PF: pharmaceutical form; AR: administration route; ATC Ther.: ATC therapeutic; B05: blood substitutes and perfusion solutions; C09: agents acting on the renin-angiotensin system; J01: antibacterials for systemic use; N06: psychoanaleptics. \*Percentages of classes and subclasses do not sum to 100% because some reports were not eligible for classification.

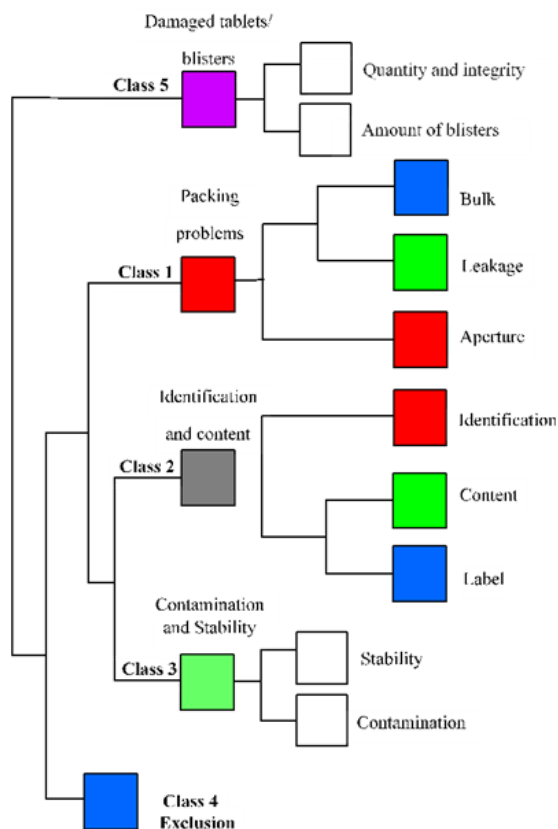


Figure 2. Schematic dendrograms of classes and subclasses of substandard medicines, formed by the descending hierarchical classification analysis of the active words of the analyzed lexicon. The colors of the classes and subclasses were chosen by IRAMuTeQ and correspond to those used in the correspondence analysis. Correspondence analysis plots were only generated for three or more lexical classes; therefore, 'Damaged tablets/blisters' and 'Contamination and stability' subclasses are represented without colors

The validation analysis allowed the correct classification of 78.4% of SM reports in their respective class. The remaining reports could not be objectively classified because they presented an equal probability of belonging to more than one lexical class.



## DISCUSSION

Our study was able to objectively classify the main reasons for quality deviation from more than 60,000 SM reports of the Brazilian Health Surveillance Notification System, dividing them into four main classes and 10 subclasses. The results of the analyses were found to be robust, with over 70% retention of the text segments in the lexicographic analysis, along with  $\chi^2$  values over 5 in the DHC and sum between axis factors close to 100 in the correspondence analysis. Our analyses were performed using data from the Brazilian National System database, and we cannot guarantee that our results extend completely to other reporting databases.

Considering the large amount of heterogeneity and bias to which SM reporting reasons are subjected (e.g. open fields to be fulfilled by the reporter), content analysis is a difficult task; however, lexicographic analysis an innovative, broader, secure and credible data analysis approach for mapping text databases<sup>25,40</sup> facilitated the objective classification of verbose SM reports, which could guide actions in drug quality surveillance. This analysis allows to employ statistical calculations on qualitative data; the vocabulary is identified and quantified in the frequency and also in relation to its position in the text. This is subjected to statistical calculations for later interpretation, overcoming the dichotomy between quantitative and qualitative research<sup>41</sup>.

The main reasons for SM reporting included packaging problems, inadequate drug identification and content, stability, contamination, and damaged tablets/blisters. These classes are similar to those reported in previous studies using other analytical methods, which confirm them as the major reasons for quality deviations of medical products worldwide. Juhlin et al.<sup>42</sup> reported that an inappropriate amount of tablets and the presence of particles were the most frequent quality problems in manufactured drugs, while Caudron et al.<sup>43</sup> classified quality problems into irregular filling of vials, contamination, mislabeling, poor stability, and packaging issues, and Taberner et al.<sup>15</sup> reported packaging problems and chemical changes.

Problems related to 'damaged tablets/blisters' were the most prevalent among the SM reports analyzed, which is consistent with the pharmaceutical market scenario, where solid dosage forms account for more than 70% of all preparations<sup>44</sup>. Although tablets require fewer precautions in terms of microbiological quality standards compared to parenteral formulations, the consequences of poor quality of this pharmaceutical form cannot be ignored. There are records of deaths in Pakistan due to contaminated tablets of antimalarial drugs<sup>17</sup>, also warned about by Almuzaini et al.<sup>10</sup> in 2014. Injectable drugs have the highest quality requirements (including sterility of the solution), yet SM reports for these products are also frequent<sup>45</sup>.

Conversely to previous studies, where most SM reports were submitted by people working in the pharmaceutical industry, physicians, and pharmacists<sup>10</sup>, we found that almost half of the Brazilian reports were submitted by hospitals, clinics, and other medical care centers. This may occur given the organization of the national healthcare system, where sentinel hospitals have a strategic position and important role in pharmacovigilance activities<sup>46,47</sup>. For imported medicines, India was the main manufacturing country. Studies suggest that about 12–25% of drugs supplied globally by India are either contaminated, substandard, or falsified<sup>48</sup>. Another study demonstrated that Indian drug manufacturers export lower-quality medicines to low-income countries compared to higher-income countries or even distribution in India itself. A potential reason for this may be the weaker and flawed regulatory control systems existing in emerging countries<sup>49</sup>. Other studies also highlight that complaints about drug quality seem to be more prevalent in countries with the worst regulatory control<sup>6,50</sup>. Although substandard medicines exist around the world, bad-natured manufacturers have a particular preference for where regulatory efforts are absent or minimal<sup>40</sup>. Caudron et al.<sup>43</sup> drew attention to the fact that manufacturers can execute production processes parallel to the main production line, adjusting their pattern according to the recipient country. This raises concerns about how regulatory agencies are actually ensuring the quality of medicines.

The volume of SM reported in Brazil is higher than in other countries<sup>10,14</sup>. The volume of SM reported in Brazil is higher than in other countries. According to a 9-year analysis of the Canadian regulatory agency, between 2005 to 2013, 649 substandard drugs were found<sup>7</sup>,

while in the United Kingdom, in a 11-year study, only 280 defective drugs were reported<sup>11</sup>. We are not sure if this higher prevalence is real or is associated with weaker SM reporting systems in other emerging countries, which may underestimate the amount of SM drugs in these countries. Although some studies highlight a global increase in SM reports over recent years<sup>10,14,19</sup>, it is not possible to conclude if this rise is due to better detection of SM by the regulatory agencies or to an increase in SM production. In this scenario, further studies to develop or improve classification systems for SM reports around the world are needed. To facilitate rapid SM report analysis, our study advocates for the creation of additional multiple-choice fields in electronic reporting systems, namely the classes resulting from our DHC analysis. This standardization could also reduce typing errors and optimize form-filling time.

Our study has some limitations. Some of the reports could not be classified into classes and subclasses because they had similar probabilities of belonging to more than one class. Further analyses with more active words could be performed to solve this problem. This study focused on the analyses of SM reports; thus, data should be interpreted in this scenario. No extrapolation of the results to other type of reports (e.g. spontaneous notifications of suspected adverse drug reactions, counterfeit drugs) were performed.

## CONCLUSIONS

By using a lexicographic analysis approach on a large pharmacovigilance database, we identified the reasons for SM reports and classified them objectively into four main classes and 10 subclasses. This classification could guide the standardization of SM reporting and contribute to improving surveillance reporting systems worldwide.

## ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001. The authors thank the Health Surveillance Service of the Department of Health in Paraná State (Southern Brazil) for the collection of data in Notivisa.

## REFERENCES

1. World Health Organization. Substandard and falsified medical products [Internet]. Geneva: WHO; 2018 [cited 2018 Dec 18]. Available from: <https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products>
2. Johnston A, Holt DW. Substandard drugs: a potential crisis for public health. *Br J Clin Pharmacol*. 2014;78(2):218-43. <http://dx.doi.org/10.1111/bcp.12298>. PMID:24286459.
3. Khan AN, Khar RK. Current scenario of spurious and substandard medicines in India: a systematic review. *Indian J Pharm Sci*. 2015;77(1):2-7. <http://dx.doi.org/10.4103/0250-474X.151550>. PMID:25767312.
4. Ravinetto R, Vandenberg D, Macé C, Pouget C, Renchon B, Rigal J, Schiavetti B, Caudron JM. Fighting poor-quality medicines in low- and middle-income countries: the importance of advocacy and pedagogy. *J Pharm Policy Pract*. 2016;9(1):36. <http://dx.doi.org/10.1186/s40545-016-0088-0>. PMID:27843547.
5. Giralt AN, Schiavetti B, Meessen B, Pouget C, Caudron JM, Marchal B, Massat P, Thys S, Ravinetto R. Quality assurance of medicines supplied to low-income and middle-income countries: poor products in shiny boxes? *BMJ Glob Health*. 2017;2(2):e000172. <http://dx.doi.org/10.1136/bmjgh-2016-000172>. PMID:28589013.
6. Wafula F, Dolinger A, Daniels B, Mwaura N, Bedoya G, Rogo K, Goicoechea A, Das J, Olayo B. Examining the quality of medicines at kenyan healthcare facilities: a validation of an alternative post-market surveillance model that uses standardized patients. *Drugs Real World Outcomes*. 2017;4(1):53-63. <http://dx.doi.org/10.1007/s40801-016-0100-7>. PMID:27888478.
7. Sakuda M, Yoshida N, Takaoka T, Sanada T, Rahman MS, Tanimoto T, Zin T, Kimura K, Tsuboi H. Substandard and falsified medicines in Myanmar. *Pharmacy*. 2020;8(1):45. <http://dx.doi.org/10.3390/pharmacy8010045>. PMID:32204459.

8. Rojas-Cortes R. Substandard, falsified and unregistered medicines in Latin America, 2017-2018. *Rev Panam Salud Publica*. 2020;44:e125. <http://dx.doi.org/10.26633/RPSP.2020.125>. PMID:33033498.
9. Khurelbat D, Dorj G, Sunderland B, Sanjiv T, Bayarsaikhan E, Damdinjav D, Dorj G, Jigjidsuren A, Lkhagvasuren O, Erdenetsetseg B. A cross-sectional analysis of falsified, counterfeit and substandard medicines in a low-middle income country. *BMC Public Health*. 2020;20(1):743. <http://dx.doi.org/10.1186/s12889-020-08897-x>. PMID:32434489.
10. Almuzaini T, Sammons H, Choonara I. Quality of medicines in Canada: a retrospective review of risk communication documents (2005-2013). *BMJ Open*. 2014;4(10):e006088. <http://dx.doi.org/10.1136/bmjopen-2014-006088>. PMID:25361839.
11. Nordt SP, Minns A, Tomaszewski C, Cantrell FL, Clark RF. Retrospective review of digoxin exposures to a poison control system following recall of Digitek(R) tablets. *Am J Cardiovasc Drugs*. 2010;10(4):261-3. <http://dx.doi.org/10.2165/11537640-000000000-00000>. PMID:20653332.
12. Davies E. More US citizens die from meningitis as scope of inquiry widens. *BMJ*. 2012;345(2):e7095. <http://dx.doi.org/10.1136/bmj.e7095>. PMID:23086963.
13. Paxton M. Current challenges with supply-chain integrity and the threat to the quality of marketed drugs. *Clin Pharmacol Ther*. 2011;89(2):316-9. <http://dx.doi.org/10.1038/clpt.2010.288>. PMID:21178987.
14. Almuzaini T, Sammons H, Choonara I. Substandard and falsified medicines in the UK: a retrospective review of drug alerts (2001-2011). *BMJ Open*. 2013;3(7):e002924. <http://dx.doi.org/10.1136/bmjopen-2013-002924>. PMID:23883882.
15. Taberner P, Fernandez FM, Green M, Guerin PJ, Newton PN. Mind the gaps--the epidemiology of poor-quality anti-malarials in the malarious world--analysis of the WorldWide Antimalarial Resistance Network database. *Malar J*. 2014;13(1):139. <http://dx.doi.org/10.1186/1475-2875-13-139>. PMID:24712972.
16. Visacri MB, de Souza CM, Sato CM, Granja S, Marialva M, Mazzola PG, Moriel P. Adverse drug reactions and quality deviations monitored by spontaneous reports. *Saudi Pharm J*. 2015;23(2):130-7. <http://dx.doi.org/10.1016/j.jsps.2014.06.008>. PMID:25972731.
17. Arie S. Contaminated drugs are held responsible for 120 deaths in Pakistan. *BMJ*. 2012;344(3):e951. <http://dx.doi.org/10.1136/bmj.e951>. PMID:22315298.
18. Renschler JP, Walters KM, Newton PN, Laxminarayan R. Estimated under-five deaths associated with poor-quality antimalarials in sub-Saharan Africa. *Am J Trop Med Hyg*. 2015;92(6, Suppl.):119-26. <http://dx.doi.org/10.4269/ajtmh.14-0725>. PMID:25897068.
19. Mackey TK, Liang BA, York P, Kubic T. Counterfeit drug penetration into global legitimate medicine supply chains: a global assessment. *Am J Trop Med Hyg*. 2015;92(6, Suppl.):59-67. <http://dx.doi.org/10.4269/ajtmh.14-0389>. PMID:25897059.
20. Machado RAT, Berardo BFR, Loureiro R. Quality of medicines in Portugal: a retrospective review of medicine recalls (2000-2015). *PDA J Pharm Sci Technol*. 2018;72(1):44-9. <http://dx.doi.org/10.5731/pdajpst.2017.007567>. PMID:29030530.
21. World Health Organization. The importance of pharmacovigilance: safety monitoring of Medicinal Products. Geneva: WHO; 2002. p. 52.
22. Pisani E, Nistor AL, Hasnida A, Parmaksiz K, Xu J, Kok MO. Identifying market risk for substandard and falsified medicines: an analytic framework based on qualitative research in China, Indonesia, Turkey and Romania. *Wellcome Open Res*. 2019;4:70. <http://dx.doi.org/10.12688/wellcomeopenres.15236.1>. PMID:31131333.
23. Leal MM, Sanz MM, Ferrando JRC, Martinez-Martinez F. A comparative analysis of the pharmacovigilance systems of Brazil, Spain, the European Union and the United States based on the information provided by their regulatory agency websites. *Daru*. 2019;27(1):379-87. <http://dx.doi.org/10.1007/s40199-019-00249-4>. PMID:30778915.
24. Choo XN, Pavord ID. Morbidity associated with oral corticosteroids in patients with severe asthma. *Thorax*. 2016;71(4):302-4. <http://dx.doi.org/10.1136/thoraxjnl-2015-208242>. PMID:26903592.
25. Trippe ZA, Brendani B, Meier C, Lewis D. Identification of substandard medicines via disproportionality analysis of individual case safety reports. *Drug Saf*. 2017;40(4):293-303. <http://dx.doi.org/10.1007/s40264-016-0499-5>. PMID:28130773.
26. Cuomo RE, Mackey TK. An exploration of counterfeit medicine surveillance strategies guided by geospatial analysis: lessons learned from counterfeit Avastin detection in the US drug supply chain. *BMJ Open*. 2014;4(12):e006657. <http://dx.doi.org/10.1136/bmjopen-2014-006657>. PMID:25468507.

27. Pinheiro LC, Candore G, Zaccaria C, Slattery J, Arlett P. An algorithm to detect unexpected increases in frequency of reports of adverse events in EudraVigilance. *Pharmacoepidemiol Drug Saf.* 2018;27(1):38-45. <http://dx.doi.org/10.1002/pds.4344>. PMID:29143393.
28. Reis MAT, Berardo BFR, Loureiro R. Quality of medicines in Portugal: a retrospective review of medicine Recalls (2000-2015). *PDA J Pharm Sci Technol.* 2018;72(1):44-9. <http://dx.doi.org/10.5731/pdajpst.2017.007567>. PMID:29030530.
29. Lebart L, Salem A, Bécue M. *Statistical analysis of texts.* Lleida: 2000.
30. Kami M, Larocca LM, Chaves M, Lowen I, Souza V, Goto D. Working in the street clinic: use of IRAMUTEQ software on the support of qualitative research. *Esc Anna Nery Rev Enferm.* 2016;20. <http://dx.doi.org/10.5935/1414-8145.20160069>.
31. Agência Nacional de Vigilância Sanitária – ANVISA. Regularização de produtos: medicamentos [Internet]. 2018 [cited 2018 May 20]. Available from: <http://portal.anvisa.gov.br/registros-e-autorizacoes/medicamentos/produtos>
32. Reinert A. Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Cah Anal Donnees.* 1983;8(2):187-98.
33. Reinert M. Classification descendante hierarchique et analyse lexicale par contexte application au corpus des poesies d'A. Rimbaud. *Bull Methodol Sociol.* 1987;13(1):53-90. <http://dx.doi.org/10.1177/075910638701300107>.
34. Reinert M. Alceste une methodologie d'analyse des donnees textuelles et une application: aurelia De Gerard De Nerval. *Bull Methodol Sociol.* 1990;26(1):24-54. <http://dx.doi.org/10.1177/075910639002600103>.
35. De Alba M. ALCESTE method and its application to the study of social representations of the urban space: the case of Mexico city. *Pap Soc Represent.* 2004;13:1-20.
36. Lemaire B. Limites de la lemmatisation pour l'extraction de significations. In: 9e Journées Internationales d'Analyse Statistique des Données Textuelles; 2008; Lyon. Lyon: Presses Universitaires de Lyon; 2008. p. 725–32.
37. Sourial N, Wolfson C, Zhu B, Quail J, Fletcher J, Karunanathan S, Bandeen-Roche K, Béland F, Bergman H. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *J Clin Epidemiol.* 2010;63(6):638-46. <http://dx.doi.org/10.1016/j.jclinepi.2009.08.008>. PMID:19896800.
38. Baroni M, Evert S. Statistical methods for corpus exploitation. In: Lüdeling A, Kytö M, editors. *Corpus linguistics: an international handbook.* Berlin: De Gruyter Mouton; 2006.
39. Clausen SE. *Applied correspondence analysis: an introduction (quantitative applications in the social sciences).* 1st ed. Thousand Oaks: SAGE; 1998. <http://dx.doi.org/10.4135/9781412983426>.
40. Souza MAR, Wall ML, Thuler A, Lowen IMV, Peres AM. The use of IRAMUTEQ software for data analysis in qualitative research. *Rev Esc Enferm USP.* 2018;52:e03353. <http://dx.doi.org/10.1590/S1980-220X2017015003353>. PMID:30304198.
41. Lahlou S. Text mining methods: an answer to Chartier and Meunier. *Pap Soc Represent.* 2012;20(38):1-38.7.
42. Juhlin K, et al. Using VigiBase to identify substandard medicines: detection capacity and key prerequisites. *Drug Saf.* 2015;38(4):373-82. <http://dx.doi.org/10.1007/s40264-015-0271-2>. PMID:25687792.
43. Caudron JM, Ford N, Henkens M, Macé C, Kiddle-Monroe R, Pinel J. Substandard medicines in resource-poor settings: a problem that can no longer be ignored. *Trop Med Int Health.* 2008;13(8):1062-72. <http://dx.doi.org/10.1111/j.1365-3156.2008.02106.x>. PMID:18631318.
44. Kaur H. Processing technologies for pharmaceutical tablets: a review. *Int Res J Pharm.* 2012;3(7):20-3.
45. Brueckner D, Krahenbuhl S, Zuber U, Bonkat G, Braissant O. An alternative sterility assessment for parenteral drug products using isothermal microcalorimetry. *J Appl Microbiol.* 2017;123(3):773-9. <http://dx.doi.org/10.1111/jam.13520>. PMID:28776899.
46. Sáfiadi MA, Berezin EN, Munford V, Almeida FJ, de Moraes JC, Pinheiro CF, Racz ML. Hospital-based surveillance to evaluate the impact of rotavirus vaccination in Sao Paulo, Brazil. *Pediatr Infect Dis J.* 2010;29(11):1019-22. <http://dx.doi.org/10.1097/INF.0b013e3181e7886a>. PMID:20543761.
47. Kuwabara CCT, Évora YDM, Oliveira MMB. Risk management in technovigilance: construction and validation of a medical-hospital product evaluation instrument. *Rev Lat Am Enfermagem.* 2010;18(5):943-51. <http://dx.doi.org/10.1590/S0104-11692010000500015>. PMID:21120414.

48. Medicines PFS. Counterfeit drug incident encyclopedia [Internet]. Partnership for Safe Medicines; 2013 [cited 2018 Dec 31]. Available from: <https://www.safemedicines.org/resources-for-healthcare-professionals/counterfeit-drug-incident-encyclopedia>
49. Bate R, Jin GZ, Mathur A, Attaran A. Poor quality drugs and global trade: a pilot study. *Am J Health Econ.* 2014;2(3):373-98. [http://dx.doi.org/10.1162/AJHE\\_a\\_00049](http://dx.doi.org/10.1162/AJHE_a_00049).
50. World Health Organization. Survey of the quality of antimalarials in Sub-Saharan Africa. Geneva: WHO; 2010.

### **Authors' contributions**

MBF, FFL and RP conceived the original idea. MBF collected the data. MBF, FST and AEMM conceived and planned the experiments. AEMM performed the IRAMuTeQ analysis. FFL performed the statistical analyses. MBF, FST and FFL contributed to the interpretation of the results. RP encouraged MBF to investigate substandard medicines and FST, FFL and RP supervised the findings of this work. MBF wrote the manuscript with support from FST. All authors provided critical feedback, discussed the results and helped shape the research, analysis and manuscript.

**SUPPLEMENTARY MATERIAL**

The complementary information is available at OSF platform (DOI 10.17605/[OSF.IO/KSDZY](https://doi.org/10.17605/OSF.IO/KSDZY))