Drug Safety (2020) 43:479–487
https://doi.org/10.1007/s40264-020-00911-w

ORIGINAL RESEARCH ARTICLE

Disproportionality Analysis for Pharmacovigilance Signal Detection in Small Databases or Subsets: Recommendations for Limiting False-Positive Associations

Ola Caster[1] · Yasunori Aoki[1,2] · Lucie M. Gattepaille[1] · Birgitta Grundmark[1]

Published online: 1 February 2020

Abstract

Introduction Uncovering safety signals through the collection and assessment of individual case reports remains a core pharmacovigilance activity. Despite the widespread use of disproportionality analysis in signal detection, recommendations are lacking on the minimum size of databases or subsets of databases required to yield robust results.

Objective This study aims to investigate the relationship between database size and robustness of disproportionality analysis, with regards to limiting spurious associations.

Methods Three types of subsets were created from the global database VigiBase: random subsets (500 replicates each of 11 fixed subset sizes between 250 and 100,000 reports), country-specific subsets (all 131 countries available in the original VigiBase extract) and subsets based on the Anatomical Therapeutic Chemical classification. For each subset, a spuriousness rate was computed as the ratio between the number of drug–event combinations highlighted by disproportionality analysis in a permuted version of the subset and the corresponding number in the original subset. In the permuted data, all true reporting associations between drugs and adverse events were broken. Subsets with fewer than five original associations were excluded. Additionally, the set of disproportionately over-reported drug–event combinations in three specific countries at three different time points were clinically assessed for labelledness. These time points corresponded to database sizes of less than 10,000, 5000 and 1000 reports, respectively. All disproportionality analysis was based on the Information Component (IC), implemented as $IC_{025} > 0$.

Results Spuriousness rates were below 0.15 for all 110 included countries regardless of subset size, with only seven countries (6%) exceeding the empirical threshold of 0.10 observed for large subsets. All 21 excluded countries had < 500 reports. For random subsets containing 3000–5000 or more reports, the higher end of observed spuriousness rates was close to 0.10. In the clinical assessment, the proportion of labelled or otherwise known drug–event combinations was very high (87–100%) across all countries and time points studied.

Conclusions To mitigate the risk of highlighting spurious associations with disproportionality analysis, a minimum size of 500 reports is recommended for national databases. For databases or subsets that are not country-specific, our recommendation is 5000 reports. This study does not consider sensitivity, which is expected to be poor in smaller databases.

1 Introduction

The collection and assessment of individual case reports remains key to detecting safety signals for marketed medicinal products [1, 2]. Signal detection from individual case reports, both in the scientific and the regulatory context, ultimately relies on accurate assessment by trained pharmacovigilance professionals. However, supporting statistical and computational methods have played an increasingly important role over the last decades, partly because databases have grown larger. Despite the availability of numerous more advanced methods [3–7], disproportionality analysis [8] is still the predominant one.

✉ Birgitta Grundmark
birgitta.grundmark@who-umc.org

[1] Uppsala Monitoring Centre, Box 1051, 751 40 Uppsala, Sweden

[2] National Institute of Informatics, Tokyo, Japan

Key Points

Standard disproportionality analysis applied in national databases containing as few as 500 individual case reports does not yield higher rates of spurious associations than in larger national databases. For databases, or subsets of databases, that are not country-specific, our results suggest 5000 reports as a suitable lower limit to avoid excessive rates of false-positive associations.

These results extend our knowledge about disproportionality analysis. They should be relevant for anyone currently using, or planning to use, disproportionality analysis in small collections of individual case reports, such as national or regional pharmacovigilance centres with low reporting volumes.

This study does not consider the issue of disproportionality analysis failing to identify true safety signals. As this is most likely a bigger concern for small databases, case-by-case review of all incoming reports remains a highly relevant alternative or complement to disproportionality analysis in such settings.

Disproportionality analysis is primarily a tool to generate hypotheses on possible causal relations between drugs and adverse effects, to be followed up by clinical assessment of the underlying individual case reports. It is based on the contrast between observed and expected numbers of reports, for any given combination of drug and adverse event. While disproportionality analysis is generally recommended and necessary for large databases [9–12], more precise guidance on the required database size for using such analysis is lacking. A common sentiment is that methodological limitations of disproportionality analysis get more relevant the smaller the database [10], but as far as we are aware, this has not been studied specifically. Increased knowledge in this area would assist organisations that hold smaller databases, for example countries with low reporting volumes, in their design of appropriate signal detection strategies. It could also be relevant for very large databases that are subjected to subgroup analyses, since individual subsets may still be small [13, 14].

This study aims to investigate the effect of database size on the robustness and relevance of disproportionality analysis, and to provide practical recommendations for smaller database or subset sizes in various situations. Here, robustness should be understood as absence of excessive rates of spurious or irrelevant associations. We put focus on national databases, which are studied using both statistical permutation techniques and clinical assessment.

2 Data and Methods

2.1 Data

All analyses were based on a frozen version of VigiBase, the WHO Global Database of Individual Case Safety Reports, as of 2 January 2018 [15]. Suspected duplicate reports [16] and reports with country of occurrence different from country of reporting (so called foreign reports) were excluded, yielding a total of 16,036,530 reports. Only drugs characterised as suspected or interacting were considered.

2.2 Creation of VigiBase Subsets

To be able to study the properties and output from disproportionality analysis in smaller databases, different kinds of subsets of VigiBase were used.

2.2.1 Random Subsets

We randomly sampled 500 subsets from all VigiBase reports, for each of the following subset sizes: 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 7500, 10,000 and 100,000 reports. No individual report was included more than once in each subset. The range between 250 and 10,000 reports was expected to cover any possible recommendations for a lower database size, and 100,000 reports was included as a control size representing a large database. While these random subsets do not correspond to any naturally occurring collections of reports, this approach allows for a controlled and structured assessment of the relation between database size and the properties of disproportionality analysis.

2.2.2 Country-Specific Subsets

National databases of individual case reports represent a highly relevant potential use case for disproportionality analysis in practice. In this version of VigiBase, 131 different countries of origin were represented, each one yielding a country-specific subset. Because each report has only a single country of origin, all these subsets are mutually exclusive. It should be noted that, in general, the database held locally in a country will differ from the corresponding subset of VigiBase. For some countries, the difference may be relatively large due to reporting backlogs or policies to submit only certain kinds of reports. Nevertheless, these country-specific subsets retain the basic property of being coherent collections of reports from a defined regulatory and cultural context, and therefore should be reasonable proxies for the purposes of this study.

2.2.3 ATC-Based Subsets

The third and final type of subsets studied was based on the ATC (Anatomical Therapeutic Chemical) classification of reported drugs [17]. Disproportionality analysis in such subsets could be practically relevant; for example, when a large database is accessible via an advanced search and analysis interface. Furthermore, the hierarchical structure of the ATC classification might offer an interesting case study, as generally the subset size decreases with the depth of the hierarchy.

Each reported drug was mapped, at substance level, to all listed ATC codes at the first, second, third and fourth level. For example, a report containing streptomycin, which is listed under the ATC codes A07AA04 and J01GA01, would be included in the following ATC-based subsets: A and J at the first level, A07 and J01 at the second level, A07A and J01G at the third level and A07AA and J01GA at the fourth level. If the report contained further drugs, it would be similarly included in the subsets corresponding to the ATC codes of those drugs.

2.3 Disproportionality Analysis

All disproportionality analyses were performed for drug–event combinations (DECs) defined as pairs of reported drugs at the preferred base (i.e. substance) level of the WHODrug Global terminology, and reported adverse events at the preferred term (PT) level of MedDRA (Medical Dictionary for Regulatory Activities) version 20.1. The measure of disproportionality used was the Information Component (IC) [18, 19], defined in the following way:

$$\mathrm{IC} = \log_2 \frac{n_{\mathrm{DE}} + 0.5}{\frac{n_{\mathrm{D}} n_{\mathrm{E}}}{n} + 0.5},$$

where n_{DE} is the number of reports on the DEC, n_{D} is the number of reports on the drug, n_{E} is the number of reports on the event and n is the total number of reports; all referring to the specific VigiBase subset under consideration. n_{DE} and $(n_{\mathrm{D}} n_{\mathrm{E}})/n$ are the observed and expected numbers of reports, respectively. As in standard IC analysis, a DEC was considered disproportionately over-reported if the lower endpoint of the 95% credibility interval for the IC was positive (i.e. if $\mathrm{IC}_{025} > 0$) [19]. Note that this can never occur if n_{DE} is < 3, regardless of the total size of the data set.

2.4 Permutation Analysis

To systematically study the prevalence of spurious associations highlighted by disproportionality analysis in the subsets of different types and sizes described in Sect. 2.2, the nonparametric statistical technique of permutation analysis was used [20, 21]. A permuted version of each data set was created, in which the listed drugs of a given report were paired with the listed adverse events of a randomly selected report. This breaks all true underlying associations between drugs and adverse events, while retaining the structure of the data. In particular, the total number of reports as well as all marginal counts for drugs and adverse events remain the same. Additionally, the correlation structures both among the drugs and among the adverse events are preserved, as are the distributions of the number of drugs and adverse events per report.

To get a measure of false-positive disproportionality analysis findings that accounts for the widely varying sizes of the subsets, we define the *spuriousness rate* as the number of DECs for which $\mathrm{IC}_{025} > 0$ in the permuted version of a data set divided by the number of DECs for which $\mathrm{IC}_{025} > 0$ in the original data set. To avoid the uncertainty of dividing by a very small number, this spuriousness rate was only computed for subsets in which at least five DECs were disproportionately over-reported in the original version of the data.

2.5 Clinical Assessment for Selected Countries

Complementing the permutation analyses, three individual countries' data were studied in-depth by manual clinical assessment of actual lists of DECs highlighted by disproportionality analysis. This allowed for characterisation of properties beyond the rate of spurious findings, for the important use case of disproportionality analysis in small country-specific subsets.

Any country with a total number of reports between 5000 and 10,000 was considered. Out of the ten countries found eligible, one (Tunisia) was randomly selected and two more (Indonesia and Brazil) were subjectively added to obtain a geographically diverse sample. For each selected country, one 'current' (as of 2 January 2018) and two backdated lists of DECs were generated. The latter were based on reports up to the years before the country surpassed 5000 and 1000 reports, respectively, in VigiBase. These lists contained all DECs that were reported disproportionately often according to the previously defined criterion $\mathrm{IC}_{025} > 0$, and their respective basic reporting statistics.

Clinical assessment was performed by an experienced pharmacovigilance assessor (BG). Each DEC was primarily classified as 'labelled or otherwise known' or not. For the latter group, it was also noted whether there was an obvious or plausible explanation why the DEC was not labelled or known. Such explanations included adverse event terms relating to lack of effect or medication errors, and too unspecific drugs or adverse event terms. In this process, the assessor made use of established medical knowledge and publicly available documentation, primarily current Summaries of Product Characteristics from Europe. Individual reports were not assessed.

The primary outcome metric was the proportion of DECs classified as labelled or otherwise known, as those highlighted DECs can be considered robust findings with respect to the performance of disproportionality analysis.

3 Results

3.1 Characteristics of VigiBase Subsets

A high-level overview of the characteristics of the three types of VigiBase subsets investigated is provided in Table 1. The random subsets generally display little variability within a given subset size, which is expected. On the contrary, the subsets based on countries and ATC groups display huge variability in all metrics considered. For example, the smallest country subset contains only seven reports, whereas the largest contains over 7.5 million reports.

For the random subsets, there seems to be very limited potential usefulness of disproportionality analysis for sizes below 2000–3000 reports, as indicated by the low numbers of DECs highlighted with $IC_{025} > 0$. More relevant from a practical perspective is to study this property within the group of country-specific subsets, especially for countries with low numbers of reports. Those results are presented in Fig. 1, indicating that there are in general more disproportional DECs for the country-specific subsets than for the random subsets of corresponding size. For example, at 1000 reports, countries generate approximately 50–80 disproportional DECs compared with a median of 15 (range 5–30) for the random subsets. This is not too surprising, since reports from an individual country should form a more coherent collection than a random sample from all of VigiBase. Still, even for the countries, < 500–1000 reports imply very few disproportional DECs.

3.2 Rate of Spurious Associations

All results concerning spuriousness rates for disproportionality analysis in VigiBase subsets of various types and sizes, generated with the permutation analysis described in Sect. 2.4, are presented in Fig. 2. The spuriousness rate for the 500 random large subsets of 100,000 reports was tightly distributed with a median of 0.085 and a range between 0.077 and 0.094 (see Fig. 2a). Considering this subset size a control group, an empirical threshold of 0.10 was set to distinguish between normal and elevated spuriousness rates. As presented in Table 1, some subsets were excluded because of having too few disproportional DECs in the original (non-permuted) data: all random subsets of 250 reports and some of 500 and 750 reports were excluded; 16% (21 of 131) of the country-specific subsets were excluded, all containing fewer than 500 reports; and between 2 and 20% of ATC-based subsets were excluded for levels 2–4.

From Fig. 2a, for randomly generated subsets, smaller size implies greater variability. To obtain a distribution of spuriousness rates with a higher end close to the 0.10 threshold, somewhere between 3000 and 5000 reports are required.

Table 1 Overview of VigiBase subsets used in the investigations of the properties of disproportionality analysis. Where applicable, variability is reported as median (min–max)

Type of subset	No. of subsets[a]	No. of reports	No. of reported DECs	No. of DECs with $IC_{025} > 0$
Random 250	500/0	250 (fixed)	728 (566–1067)	0 (0–4)
Random 500	500/118	500 (fixed)	1434 (1200–2182)	3 (0–11)
Random 750	500/461	750 (fixed)	2143 (1888–3131)	8 (2–20)
Random 1 k	500/500	1000 (fixed)	2826 (2467–4972)	15 (5–30)
Random 2 k	500/500	2000 (fixed)	5463 (4838–6633)	62 (37–92)
Random 3 k	500/500	3000 (fixed)	7966 (7296–9627)	128 (97–177)
Random 4 k	500/500	4000 (fixed)	10,372 (9563–11,753)	205 (161–249)
Random 5 k	500/500	5000 (fixed)	12,632 (11,899–14,539)	291 (245–338)
Random 7.5 k	500/500	7500 (fixed)	18,097 (17,142–19,839)	533 (477–593)
Random 10 k	500/500	10,000 (fixed)	23,176 (22,036–26,072)	795 (716–877)
Random 100 k	500/500	100,000 (fixed)	142,286 (139,472–145,252)	10,550 (10,249–10,812)
Country-specific	131/110	2802 (7–7,652,319)	2844 (8–1,823,144)	173 (0–260,568)
ATC level 1	14/14	1,532,465 (151,017–3,423,029)	722,124 (147,388–1,025,528)	86,347 (11,299–123,187)
ATC level 2	98/96	146,849 (3–1,704,151)	143,098 (15–809,020)	10,622 (0–100,821)
ATC level 3	287/269	27,292 (8–1,704,151)	33,218 (11–487,295)	1434 (0–53,700)
ATC level 4	980/786	4491 (1–913,516)	7475 (1–311,556)	203 (0–29,394)

DEC drug–event combination, *IC* information component

[a] Total number of subsets/number of subsets included in permutation analysis

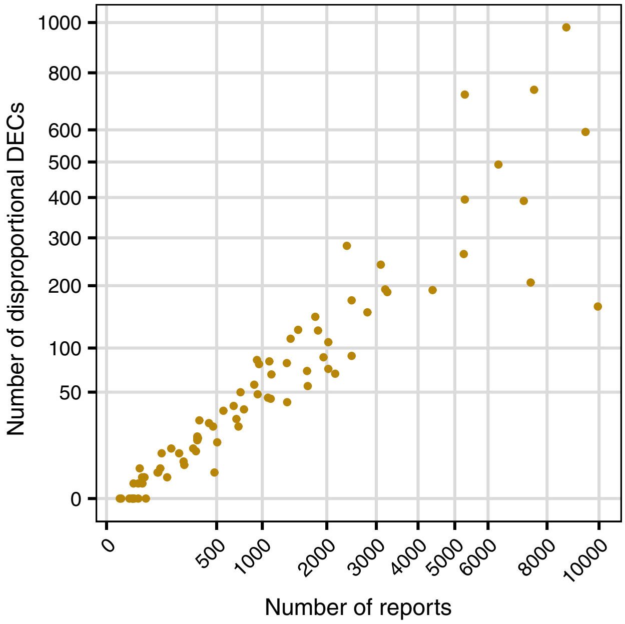

Fig. 1 The relation between the number of disproportional drug–event combinations (defined as $IC_{025} > 0$) and the size for country-specific subsets of VigiBase. Only countries with 10,000 or fewer reports are included. Note that both the *x* and *y* axis have been subjected to a square root transformation, to enhance the clarity of the displayed data

For country-specific subsets, shown in Fig. 2b, the results are quite different: all countries except one are below or just slightly above the threshold, with a single minor outlier with a spuriousness rate of 0.14. Only 7 (6%) of all included countries are above the 0.10 threshold. This difference cannot be attributed solely to the fact that countries provide real rather than random collections of reports: subsets based on ATC groups at level 4 (see Fig. 2c) are also real and cover approximately the same range of report counts as the countries. However, 12% of those subsets are above the 0.10 spuriousness rate threshold and some values are extremely high. If the comparison is restricted only to subsets of sizes between 500 and 10,000 reports, 7% of countries are above the threshold compared with 20% for ATC groups of level 4.

Subsets based on ATC groups at levels 1 and 2 display low spuriousness rates, which is expected from their generally high numbers of reports (see Fig. 2c). Even at level 3, results are very different from those presented earlier for level 4: only seven (8%) subsets in the 500–10,000 report range are above the 0.10 threshold. [Among all level 3 subsets, the corresponding number is 12 (4%)].

3.3 Clinical Assessment for Selected Countries

Table 2 presents overall characteristics of the nine lists of disproportionately over-reported DECs from Tunisia, Brazil and Indonesia at different time points, as well as the results from the clinical assessment of those DECs. The major finding is that across all countries and time points, the proportion of labelled or otherwise known DECs is very high (87–100%). This suggests that the output from disproportionality analysis in these small national subsets of VigiBase is robust, in the sense that most highlighted DECs correspond to established causal associations. At the same time, it leaves little opportunity for identifying new signals: in the current lists from Tunisia, Brazil and Indonesia, there were only 10 (5%), 56 (12%) and 21 (5%) DECs, respectively, that were not labelled/known.

The results for Brazil deviate slightly in the sense that for most of the DECs that were not labelled or otherwise known, an explanation was identified. The predominant explanation was an event term that suggested lack of effect. This was, however, not the case for the smallest Brazilian list, dating back to 2002.

4 Discussion

Disproportionality analysis is the most common statistical approach to support the detection of safety signals for marketed medicines from databases of individual case reports. This study provides important new information on the impact of database size on the robustness of the disproportionality analysis output. Contrary to common belief, our results suggest that disproportionality analysis can be used even in small databases without an elevated risk of highlighting spurious associations. However, at least for national databases, the potential for detecting new safety signals decreases with the number of reports. Based on our study, it is possible to formulate practical recommendations that will hopefully complement existing signal detection guidelines, in particular for national and regional databases with low volumes of reporting.

Using different kinds of subsets of VigiBase—randomly sampled subsets, country-specific subsets and subsets based on ATC codes of reported drugs—we have investigated the effect of database size on the rate of spurious associations for drug–event combinations (DECs) highlighted with the specific disproportionality algorithm $IC_{025} > 0$. Spuriousness rates were estimated by creating a permuted version of each data set, void of true associations but retaining the structure of the corresponding original data set. In addition, we have clinically assessed lists of DECs highlighted by disproportionality analysis for three different countries at three different time points that correspond to three levels of database size: between 5000 and 10,000 reports, < 5000 reports and < 1000 reports.

Our major finding is that, in contrast to random and ATC-based subsets, country-specific subsets were consistently

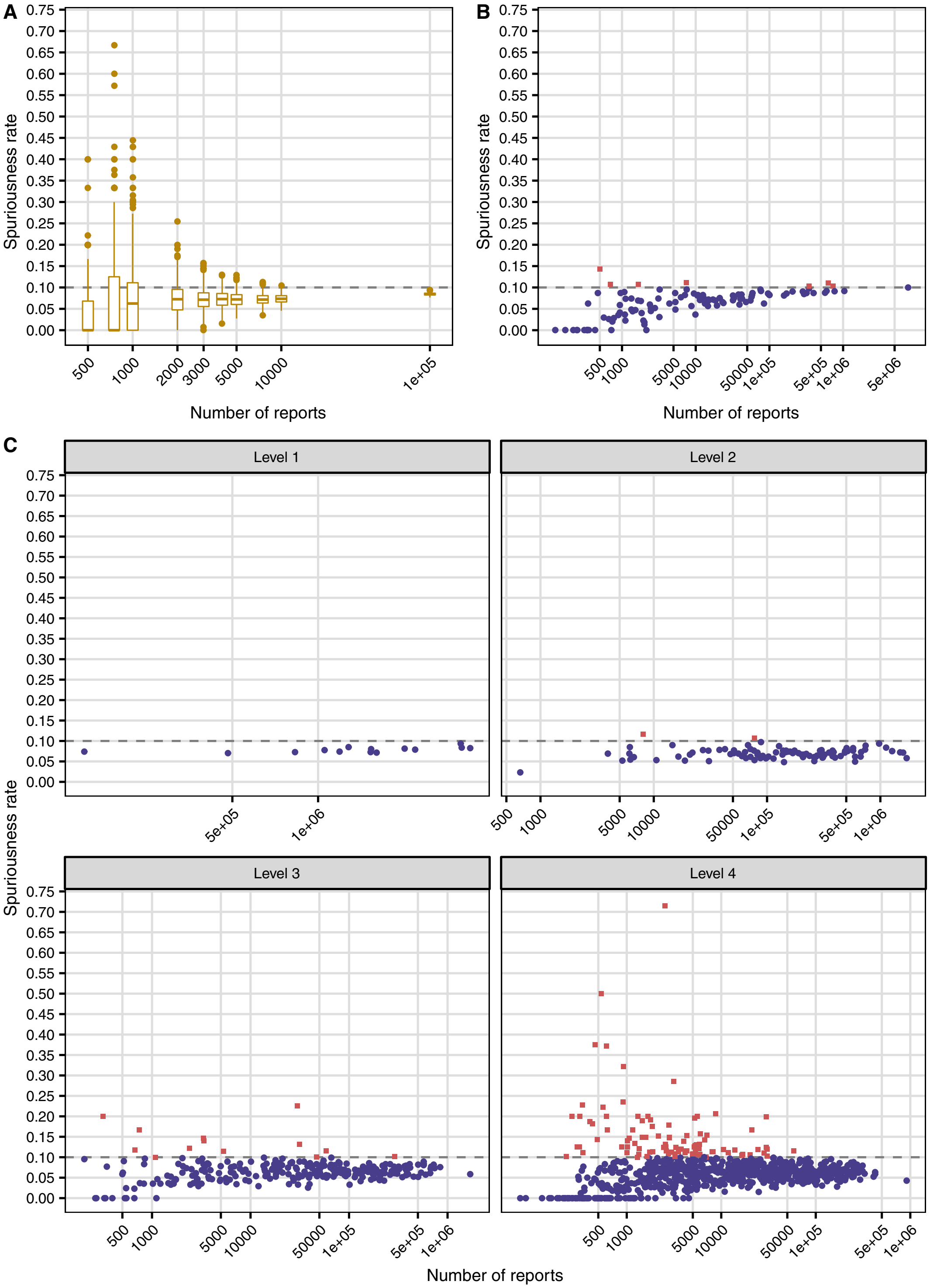

Fig. 2 The rate of spuriously highlighted drug–event combinations by disproportionality analysis (defined as $IC_{025} > 0$) for different types and sizes of VigiBase subsets. **a** shows box plots for randomly generated subsets of sizes between 500 and 100,000 reports. Each box is based on 500 subsets, except for those with 500 reports (118 included subsets) and 750 reports (461 included subsets). **b** shows results for country-specific subsets. Of 131 countries, 21 (16%) were excluded, all with fewer than 500 reports. **c** displays results for subsets based on ATC groups; 2%, 6% and 20% of subsets were excluded at level 2, 3 and 4, respectively. The horizontal lines at 0.10 indicate an empirical threshold for normal spuriousness rates derived from large subsets; individual points in (**b**, **c**) above and below this threshold are drawn as red squares and blue circles, respectively. Note that all *x* axes are logarithmic and adjusted to the data of the individual panels

very robust to false-positive associations. Among the 110 assessable countries, only 6% were above the empirical spuriousness rate threshold based on large databases and only a single country deviated more than marginally. These results were corroborated by those from the clinical assessment: even for the smallest back-dated subsets of reports, most DECs highlighted by disproportionality analysis were known associations. Because we were unable to reliably estimate spuriousness rates for several of the countries with < 500 reports in VigiBase, our recommendation is still to not use disproportionality analysis in national or regional databases below 500 reports. Also, below this size, the total number of disproportional DECs was very low (< 50, as seen in Fig. 1), offering minimal potential for discovering safety signals.

For small databases or subsets of larger databases that are not based on an individual country or other regional entity, our results for random and ATC-based subsets can be used to form generic recommendations. From Fig. 2a, a lower limit somewhere between 3000 and 5000 reports seems appropriate, as the variability of the spuriousness rates rapidly increases below those sizes. This agrees reasonably well with the results for ATC-based groups at level 4 (see Fig. 2c), where extreme spuriousness rates disappear from around 5000 reports. Also, above this size, the proportion of subsets that exceed the empirical 10% spuriousness rate threshold is 33 of 471 (7%), which is very close to the proportion for all countries. All results considered, we find 5000 reports to be an appropriate generic recommendation.

All our results suggest that disproportionality analysis can be used in considerably smaller databases and subsets than we had expected. Our trivial and plausible explanation is that statistical associations cannot appear without the underlying data supporting them: all common disproportionality analysis algorithms in use require at least three reports on an individual DEC for it to be highlighted, and those reports must be submitted by someone for some reason. In our clinical assessment of data from Tunisia, Brazil and Indonesia, we did not see any evidence of entirely nonsensical DECs being highlighted, not even when the total number of reports was < 1000. Nevertheless, it is very important to realise that even if disproportionality analysis may be robust in small databases, it may not contribute significantly to signal detection compared with qualitative methods.

By design, our study is limited to investigating robustness of disproportionality analysis in terms of avoiding false-positive (spurious) associations, which is a basic prerequisite for an effective first-pass screening tool in databases of spontaneous reports. The practical consequence of a false-positive association will typically be a waste of manual resources required to refute this association at a later stage. Its severity will vary greatly depending on the context but should not be neglected for national or regional pharmacovigilance centres in generally resource-limited settings, or where medical expertise is scarce. False-negative associations (i.e. true safety signals not detected by disproportionality analysis) are at least as important but in general more difficult to investigate. Doing this properly requires reference sets of time-stamped emerging safety issues [22], which are difficult to obtain in general, and practically infeasible on a country-by-country basis. Even the reference sets that do exist are very unlikely to provide a complete gold standard of all true positives, which obviously limits the possibilities of directly assessing false-positive and false-negative associations alike. Our permutation analysis provides a proxy for the assessment of false positives, but unfortunately an analogous approach for false negatives is much more challenging to devise. Intuitively though, smaller databases should suffer more from false-negative associations than larger databases. Our results indirectly support this notion by establishing a clear relationship between the number of reports and the number of DECs highlighted by disproportionality analysis, for the country-specific subsets. This puts an upper bound on the number of true associations that actually can be detected by means of disproportionality analysis.

Data homogeneity is another factor not studied here that could influence the performance of disproportionality analysis. If the data contains a limited number of similar drugs or adverse events, the background reporting rates used to compute the expected number of reports may be inappropriately high. The same can happen if one or a few DECs account for a high proportion of reports, an effect often called masking. Both of these issues can potentially cause an inflation in the false-negative rate, and it has been suggested that small databases might be particularly vulnerable to this [10]. We recommend being mindful of these aspects and possibly investigating frequency distributions of drugs and adverse events, especially if analysing subsets of a database where skewness may be expected. Additional research should be conducted to investigate in more detail the impact of low numbers of reports on the false-negative rate and how this relates to data homogeneity. False negatives are important since, in contrast to false positives, they cannot easily be identified during subsequent manual review.

A limitation of our clinical assessment is that the classification of DECs as either known or not was made using current knowledge. This may have overestimated the proportion of known DECs, at least for the lists based on back-dated subsets. Additionally, for accessibility and language reasons we have used European product labels. The adverse effects included therein might not be known in the selected countries, or there could be particular circumstances that would make some of the DECs classified as known by us to be considered signals in the local setting. Nevertheless, the fact that a disproportionately over-reported DEC has been included in a product label anywhere validates the highlighted statistical association as such. Lastly, all our

Table 2 Results from the clinical assessment of current and backdated lists from Tunisia, Brazil and Indonesia of drug–event combinations reported disproportionately often

Combination list	No. of reports	No. of DECs with $IC_{025} > 0$	No. of labelled/known DECs	
			Yes	No (with/without explanation)
Tunisia current[a]	7189	201	191 (95%)	10 (0/10)
Tunisia 2010	4209	100	99 (99%)	1 (0/1)
Tunisia 2000	634	10	9 (90%)	1 (0/1)
Brazil current[a]	6064	479	423 (88%)	56 (39/17)
Brazil 2015	4297	315	273 (87%)	42 (30/12)
Brazil 2002	932	29	29 (100%)	0 (0/0)
Indonesia current[a]	6925	389	368 (95%)	21 (1/20)
Indonesia 2013	4156	191	189 (99%)	2 (0/2)
Indonesia 1976[b]	564	21	20 (95%)	1 (0/1)

DEC drug–event combination, *IC* information component

[a] As of the end of the data extract (i.e. 2 January 2018)

[b] This is data collected prior to Indonesia joining the WHO Programme for International Drug Monitoring in 1990, which has been retroactively added to VigiBase

analyses are based on a single measure of disproportionality, the IC, and a single algorithm, $IC_{025} > 0$. While this is a limitation, all commonly used measures are similar enough for drug-event analysis that we would expect our results to generalise reasonably well beyond the IC [23]. As for the choice of algorithm, we have used the standard form of IC analysis because this seemed the most relevant from a practical perspective. Using other algorithms, either with the IC or another measure of disproportionality, might give slightly different results, especially if using a non-standard algorithm with either very high or very low propensity to highlight disproportionate reporting.

5 Conclusions

This study shows that disproportionality analysis can be used in small collections of individual case reports without great risk of generating excessive numbers of spurious findings. Based on our results obtained with the disproportionality analysis algorithm $IC_{025} > 0$, we recommend a lower size of at least 500 reports for national databases, and at least 5000 reports for databases or subsets of databases constructed in other ways. This study does not consider the issue of true safety signals not detected by disproportionality analysis, and does not suggest that disproportionality analysis will be equally effective and meaningful at all database sizes. For small databases, where the risk of false-negative associations is most likely higher, case-by-case review or systematic review of all reported drug-event combinations is probably advisable, resources permitting. Regardless of database size, it is imperative to acknowledge the primary role of disproportionality analysis as a tool for hypothesis generation, and the importance of subsequent manual clinical review.

Acknowledgements The authors are indebted to the national centres that make up the WHO Programme for International Drug Monitoring and contribute reports to VigiBase. Particular gratitude is paid to the centres of Tunisia, Brazil and Indonesia for allowing the publication of results based on their data in VigiBase. However, the opinions and conclusions of this study are not necessarily those of the various centres nor of the WHO. As of October 2018, Yasunori Aoki is a full-time employee of Astra Zeneca; and as of October 2019, Ola Caster is a full-time employee of Pharmetheus. However, all of their contributions to this work were made prior to their respective time of departure, as part of their employment at the Uppsala Monitoring Centre.

Compliance with Ethical Standards

Funding No external sources of funding were used to assist in the preparation of this study.

Conflict of interest Ola Caster, Yasunori Aoki, Lucie Gattepaille and Birgitta Grundmark declare that they have no conflicts of interest that are directly relevant to the content of this study.

Data sharing The datasets generated and analysed during the current study are not publicly available due to agreements between contributors of data to the database used (VigiBase) and the custodian of this database. National centres (mainly national drug regulatory authorities) constituting the WHO Programme for International Drug Monitoring (PIDM) contribute data to VigiBase and the Uppsala Monitoring Centre is the custodian in its capacity as WHO collaborating centre for international drug monitoring. Some subsets of the data may be available from the corresponding author on reasonable request.

References

1. Lane S, Lynn E, Shakir S. Investigation assessing the publicly available evidence supporting postmarketing withdrawals, revocations and suspensions of marketing authorisations in the EU since 2012. BMJ Open. 2018;8:e019759.
2. Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. BMC Med. 2016;14:10.
3. Caster O, Norén GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database. Stat Anal Data Min. 2010;3:197–208.
4. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012;4:125ra31-ra31.
5. Kulldorff M, Dashevsky I, Avery TR, Chan AK, Davis RL, Graham D, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiol Drug Saf. 2013;22:517–23.
6. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. 2013;93:539–46.
7. Caster O, Juhlin K, Watson S, Norén GN. Improved statistical signal detection in pharmacovigilance by combining multiple strength-of-evidence aspects in vigiRank. Drug Saf. 2014;37:617–28.
8. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18:427–36.
9. U.S. Food and Drug Administration (FDA). Guidance for industry—good pharmacovigilance practices and pharmacoepidemiologic assessment. 2005. https://www.fda.gov/media/71546/download. Accessed 25 June 2019.
10. CIOMS Working Group VIII. Practical aspects of signal detection in pharmacovigilance. Geneva: CIOMS; 2010.
11. Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. Good signal detection practices: evidence from IMI PROTECT. Drug Saf. 2016;39:469–90.
12. European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP). Module IX Addendum I—Methodological aspects of signal detection from spontaneous reports of suspected adverse reactions. 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-addendum-i-methodological-aspects-signal_en.pdf. Accessed 25 June 2019.
13. Hopstadius J, Norén GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. Drug Saf. 2008;31:1035–48.
14. Seabroke S, Candore G, Juhlin K, Quarcoo N, Wisniewski A, Arani R, et al. Performance of stratified and subgrouped disproportionality analyses in spontaneous databases. Drug Saf. 2016;39:355–64.
15. Lindquist M. VigiBase, the WHO global ICSR database system: basic facts. Drug Inf J. 2008;42:409–19.
16. Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. Data Min Knowl Discov. 2007;14:305–28.
17. WHO Collaborating Centre for Drug Statistics Methodology. ATC: Structure and principles. 2018. https://www.whocc.no/atc/structure_and_principles/. Accessed 20 Aug 2019.
18. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol. 1998;54:315–21.
19. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. Stat Methods Med Res. 2013;22:57–69.
20. Hopstadius J, Norén GN. Robust discovery of local patterns: subsets and stratification in adverse drug reaction surveillance. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium; Miami, Florida, USA: ACM; 2012. pp. 265–274.
21. Juhlin K, Star K, Norén GN. A method for data-driven exploration to pinpoint key features in medical data and facilitate expert review. Pharmacoepidemiol Drug Saf. 2017;26:1256–65.
22. Norén GN, Caster O, Juhlin K, Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. Drug Saf. 2014;37(9):655–9.
23. Candore G, Juhlin K, Manlik K, Thakrar B, Quarcoo N, Seabroke S, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. Drug Saf. 2015;38:577–87.