Drug Safety (2022) 46:39–52
https://doi.org/10.1007/s40264-022-01251-7

ORIGINAL RESEARCH ARTICLE

Individual Case Safety Report Replication: An Analysis of Case Reporting Transmission Networks

John van Stekelenborg[1] · Vijay Kara[2] · Roman Haack[3] · Ulrich Vogel[4] · Anju Garg[5] · Markus Krupp[6] · Kate Gofman[7] · Brian Dreyfus[8] · Manfred Hauben[9,10] · Andrew Bate[2,11,12]

Accepted: 18 October 2022 / Published online: 24 December 2022

Abstract

Introduction The basis of pharmacovigilance is provided by the exchange of Individual Case Safety Reports (ICSRs) between the recipient of the original report and other interested parties, which include Marketing Authorization Holders (MAHs) and Health Authorities (HAs). Different regulators have different reporting requirements for report transmission. This results in replication of each ICSR that will exist in multiple locations. Adding in the fact that each case will go through multiple versions, different recipients may receive different case versions at different times, potentially influencing patient safety decisions and potentially amplifying or obscuring safety signals inappropriately.

Objective The present study aimed to investigate the magnitude of replication, the variability among recipients, and the subsequent divergence across recipients of ICSRs.

Methods Seven participating TransCelerate Member Companies (MCs) queried their respective safety databases covering a 3-year period and provided aggregate ICSR submission statistics for expedited safety reports to an independent project manager. As measured in the US Food and Drug Administration (FDA)'s Adverse Event Reporting System (FAERS), ICSR volume for these seven MCs makes up approximately 20% of the total case volume. Aggregate metrics were calculated from the company data, specifically: (i) number of ICSR transmissions, (ii) average number of recipients (ANR) per case version transmitted, (iii) a submission selectivity metric, which measures the percentage of recipients not having received all sequential case version numbers, and (iv) percent of common ISCRs residing in two or more MAH databases.

Results The analysis reflects 2,539,802 case versions, distributed through 7,602,678 submissions. The overall mean replication rate is 3.0 submissions per case version. The distribution of the ANR replication measure was observed to be very long-tailed, with a significant fraction of case versions (~ 12.4% of all transmissions) being sent to ten or more HA recipients. Replication is higher than average for serious, unlisted, and literature cases, ranging from 3.5 to 6.1 submissions per version. Within the subset of ICSR versions sent to three recipients, a significant degree of variability in the actual recipients (i.e., HAs) was observed, indicating that there is not one single combination of the same three HAs predominantly receiving an ICSR. Submission selectivity increases with the case version. For case version 6, the range of the submission selectivity for the MAHs ranges from ~ 10% to over 50%, with a median of 30.2%. Within the participating MAHs, the percentage of cases that reside within at least two safety databases is approximately 2% across five databases. Further analysis of the data from three MAHs showed percentages of 13.4%, 15.6%, and 27.9% of ICSRs originating from HAs and any other partners such as other MAHs and other institutions.

Conclusion Replication of ICSRs and the variation of available safety information in recipient databases were quantified and shown to be substantial. Our work shows that multiple processors and medical reviewers will likely handle the same original ICSR as a result of replication. Aside from the obvious duplicate work, this phenomenon could conceivably lead to differing clinical assessments and decisions. If replication could be reduced or even eliminated, this would enable more focus on activities with a benefit for patient safety.

Extended author information available on the last page of the article

Key Points

ICSRs are replicated between different Health Authorities and other databases, resulting in the replication of case processing and medical assessment activities.

Our paper quantifies the level of replication and diverging information resulting from the replication process.

As this replication leads to significant levels of duplicate work and may conceivably lead to different safety decisions, mitigation of this phenomenon should be considered.

1 Introduction

Ensuring the safety of medicines and vaccines through systematic collection and review of adverse event (AE) reports is core to pharmacovigilance (PV) signal detection [1]. Overall, this is a shared responsibility between patients, healthcare professionals, academic researchers, regulatory bodies, and pharmaceutical companies. Pharmaceutical companies are legally responsible for collecting and evaluating AE reports and continuously monitoring the evolving evidence, taking appropriate action to maintain favorable benefit-risk profiles for the medical therapies they produce.

Pharmaceutical companies are required to share the reports that they receive with other parties in a timely manner, for example, with other companies when there are co-marketing agreements in place, and with Health Authorities (HAs) based on varying reporting criteria. Even without co-marketing agreements, some Marketing Authorization Holders (MAHs) forward a report for a non-company drug to the local HA and/or manufacturer (if known) according to local requirements. These individual case safety reports (ICSRs) are validated and processed, often in paper format, requiring a significant level of manual inputs and reconciliation between various entities in a timely fashion. Current safety data exchange and reporting mechanisms are complex due to variable regulatory requirements for report sharing [2]. The growth of safety data [3], paired with the lack of international harmonization around reporting requirements, can potentially lead to data quality problems and negatively impact public health.

While the long-standing quality of case *duplication* in *a single database* has been recognized as a significant problem in PV [4, 5], very little is known about the phenomenon of "*replication*." Replication of AE reports *between databases* occurs, as discussed above, when ICSRs are submitted to any recipient (e.g., HAs and companies) and then re-reported to multiple additional recipients. The net result is that each ICSR's case data reside in numerous databases, and each is repeatedly processed by numerous stakeholders. Replication may also lead to case duplication within databases where multiple pharmaceutical companies submit the same report to the same HA for regulatory compliance [5], which is a challenge in itself. Quantitatively and qualitatively speaking, differences can occur between safety databases and downstream analytical systems. Efficiency and completeness/correctness of information are probably affected by replication. However, no data quantifying replication has been published.

2 Study Objectives

The objective of this study is to estimate the extent and nature of primary replication of ICSRs, i.e., reporting of ICSRs by one or multiple Marketing Authorization/Clinical Trial Authorization (MAH/CTA) holders to *multiple* HAs. Additionally, replication across MAH databases is assessed. Within the framework of transmitting ICSRs, it is helpful to distinguish between "primary transmission" and "secondary transmission." For this paper, primary transmission is defined as the ICSR transmissions by the MAH/CTA holder of a pharmaceutical product. Secondary transmission is defined as further transmissions of these ICSRs from the recipients of the primary transmissions. The key recipients of primary transmission are HAs and partner companies. Additionally, in interventional clinical trials (ICTs), ICSRs are also sent by the CTA holder to Investigators and Ethics Committees/Institutional Review Boards (ECs/IRBs). For this paper, while ICSRs from clinical trials are in scope, *reporting destinations* other than HAs (e.g., Investigators and ECs/IRBs) are not, as the focus is on transmissions that will generally be stored in a reference database resulting in replication (Fig. 1).

Specifically, with regard to those MAHs that participated in the analysis, the replication assessment aims to:

(1) Retrospectively quantify the level of primary replication of ICSRs resulting from MAHs transmitting a given ICSR to multiple HAs.
(2) Retrospectively quantify the level of inter-company replication of ICSRs using the presence of common Worldwide Unique Identification Numbers (WUCIN) in multiple safety databases of the TransCelerate Member Companies participating in the project.
(3) Quantify the extent to which recipients have "complete information" regarding an ICSR, defined as having received all initial and subsequent case versions sub-

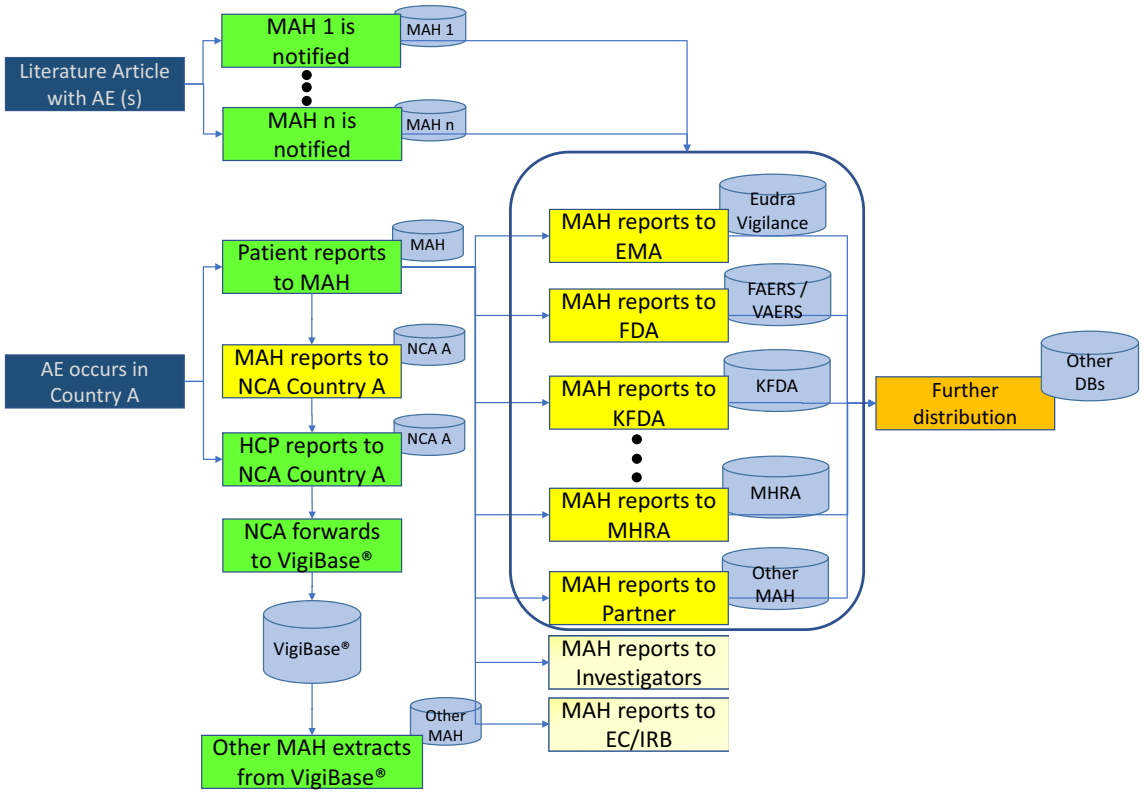

Fig. 1 Schematic representation of ICSR transmission. *AE* adverse event, *DB* database, *EC* ethics committee, *EMA* European Medicines Agency, *FAERs* FDA Adverse Event Reporting System, *FDA* Food and Drug Administration, *HA* health authority, *HCP* healthcare professional, *ICSR* individual case safety report, *ICT* interventional clinical trial, *IRB* institutional review board, *KFDA* Korean Food and Drug Administration, *MAH* Marketing Authorization Holder, *MHRA* Medicines and Healthcare Products Regulatory Agency, *VAERS* vaccine adverse event reporting system. The flowchart is for illustration purposes only and is not intended to be complete or representative of every ICSR. Legend: Yellow = Primary Transmissions; Yellow/Dashed = Primary Transmissions of ICT cases to EC/IRB/Investigators (out-of-scope); Orange = Secondary Transmissions; Green = Other Transmissions. Note that literature reports are usually identified by multiple pharmaceutical companies, which then forward the reports to multiple HAs; and literature reports may be independently identified by HAs themselves

mitted by MAHs, for example, due to new follow-up information obtained.

The net result is that any transmitted ICSR resides—with some lag time—in multiple databases, and each case is repeatedly processed and assessed by multiple stakeholders and replicated across the globe. This situation precludes a "ground truth" or "canonical version" of the original AE.

3 Methods

Participating TransCelerate Member Companies (hereafter referred to as MAHs) analyzed their internal safety data and provided summary tabulations for a 3-year period (generally, 1 January 2018–31 December 2020, although 1 June 2018–31 December 2020 in one case) to assess the number of HAs to whom an ISCR was submitted as an *expedited* report. Cases submitted in an aggregate report and not in an expedited manner were not in scope. The date range was selected to account for significant changes to ICSR reporting to HAs (e.g., centralized reporting to EudraVigilance in the European Union). Therefore, the sampled period remains generalizable to current PV practice.

3.1 Basic Definitions and Nomenclature

The following nomenclature is used for this paper:

- Case or ICSR: an individual case safety report submitted in an expedited manner that can have one or more versions.
- Case Version: a specific version of an ICSR (e.g., Version 1 is the first version of the report, Version 2 is an updated version of the same report that was submitted in an expedited manner, usually containing follow-up information).

- Submission: a Case Version submitted via expedited means to a specific HA.
- Recipient: an HA receiving a submission from a MAH/CTA holder.
- Primary Transmission: ICSR transmission by the MAH/CTA holder of a pharmaceutical product to the designated HA.
- Overall Replication Rate = number of submissions/numbers of case versions.
- Percentage of Total Case Versions as a Function of Number of Recipients = fraction of all case versions that are sent to one, two, three, etc. recipients.

The following is an example of these logically connected definitions:

- Upon receipt of an AE report, an ICSR with a unique identifier is created in an MAH's Safety Database (e.g., the internal case identifier is 100356).
- The MAH receives an additional three follow-up reports, resulting in four case versions of the ICSR. These are identified as 100356 (1), 100356 (2), 100356 (3), and 100356 (4).
- Due to different HA demands for follow-up information, the MAH submits version 1 to two HAs, version 2 to three HAs, version 3 to two HAs, and version 4 to one HA. Therefore, the total number of submissions is eight for this ICSR. Note that, while this is a hypothetical example, the number of recipients is driven by reporting rules for each recipient. Some recipients may require an updated version, while others do not. Likewise, new or updated information on relatedness, expectedness, or seriousness can trigger different distribution rules for different versions of the same case.

In the example above, four case versions are sent by the MAH. The distribution of recipient counts is:

- One HA recipient: 25% of the four case versions.
- Two HA recipients: 50% of the four case versions.
- Three HA recipients: 25% of the four case versions.

The terms “submission: and “transmission” are used interchangeably in this paper.

Stratifications

In addition to defining the "counting units" as above, data corresponding to several case stratifications were collected, to assess whether specific case characteristics selectively drive replication:

1. Case Type: Spontaneous or Non-Spontaneous.
2. Healthcare Professional (HCP) confirmed: Yes or No.
3. Seriousness: Serious or Non-serious.
4. Listedness (per Core Data Sheet [CDS]): Unlisted or Listed.
5. Literature: Yes or No.

3.2 Data Collection Methodology for Marketing Authorization Holders (MAHs)

A detailed data collection methodology was designed for this study to characterize and quantify the replication level in the Primary Transmission process. Under this approach, the participating MAHs generated four independent data outputs (Online Resource 1: ICSR Primary Transmission Data Collection Methodology (referred to as Online Resource 1 Tables 1, 2, 3, and 4)), which are part of the Electronic Supplemental Material (ESM). Each MAH was requested to provide the following:

- **Data output 1**: Case version volume
  - Number of case versions, stratified by the five stratification variables.
  - Number of case versions by version number from version 1 through 5 and aggregated for versions higher than 5.
- **Data output 2**: Replication
  - The average number of HA recipients of submissions by case version, for versions 1 through 5
  - The average number of HA recipients of submissions for versions 6 or higher
  - The average number of HA recipients as above, stratified by the five case attributes.
  - Full distribution: count of case versions submitted as a function of the number of HA recipients (note: five of seven MAHs provided this data)
- **Data output 3**: Submission Selectivity
  - Submission Selectivity Measure by case version (see Sect. 3.4 below)
- **Data output 4**: MAH Overlap
  - Percentage of case versions residing in two or more participating MAH Safety Databases (see Sect. 3.5 below)

Regarding data output 2, the choice to collect only the arithmetic mean (average) and submission counts for each MAH was driven by the need to keep anonymity among the MAHs. This limitation prevents further investigation of potential "interactions" between the various case attributes such as "Serious AND HCP Confirmed."

3.3 Submission Selectivity Measure

As discussed, different HAs may receive different case versions of the same case, resulting in a situation where some HAs are "blind" to at least some of the information at some time for the ICSR of interest. In other words, some HAs may have incomplete information when measured against the totality of available case versions. This study did not qualify the exact case attributes that are deviating most, and thus no impact to signal detection can be derived. However, an indicative measure has been developed which delivers a quantification of the incompleteness of case versions distributed. For the mathematical details of this submission selectivity measure, consult the ESM.

3.4 MAH Overlap Measure

Participating MAHs provided lists of WUCINs, which were compiled and analyzed for the presence of the same WUCIN in multiple MAH databases. Since WUCINs are used to uniquely identify a case throughout its lifecycle (including multiple transmissions), a WUCIN occurring in multiple MAH databases quantifies replication across MAHs.

4 Results

All seven MAHs provided the data described in the data collection methodology with the following exception:

- Two MAHs did not provide the list of WUCINs for Online Resource 1 Table 4 (ESM). Therefore, the results based on this table represent five MAHs.

ICSR volume reported for seven MCs makes up approximately 20% of case volume in the FAERS database.

Data from the seven MAHs covered 2,539,802 case versions, distributed through 7,602,678 submissions. This represents an overall replication rate of 7,602,678 / 2,539,802 = 3.0 transmissions per case version. For Data output 4, the five MAHs contributed 1,681,388 associated WUCINs in the raw format as otherwise available through publicly available information (e.g., FAERS).

For anonymization, the seven MAHs are arbitrarily designated as MAH01 through MAH07. These designations are used throughout the Results section.

4.1 Case Version Volume

A significant range in overall case version volume exists for the seven MAHs (approximately 100,000 to slightly over 1,000,000). The median number of case versions for the seven MAHs is 294,396, while the mean is 406,303 case versions. Figure 2 provides the median percentages for case versions across the seven MAHs combined for each case attribute, as well as by case version. In addition to large differences in case volume, significant variability is present across case attributes, for example, the relative abundance of spontaneous and clinical trial cases. The ESM shows the differences in each case attribute for the seven MAHs.

4.2 Replication

4.2.1 Average Number of Recipients

The results from Online Resource 1 Table 2 (ESM; Average Number of Recipients (ANR)) are summarized in Fig. 3.

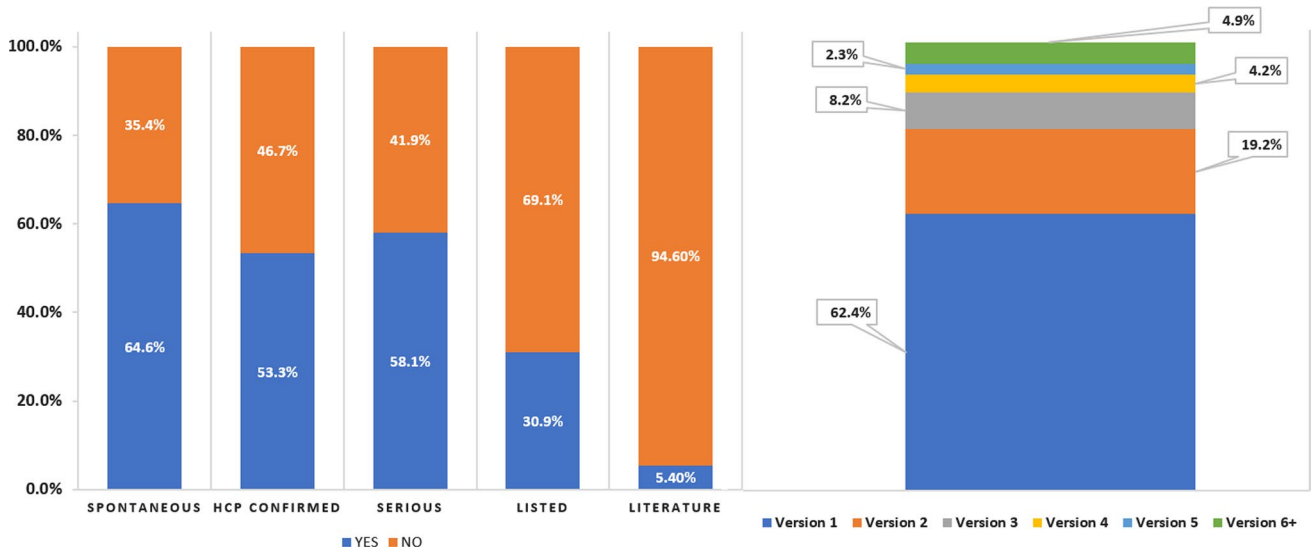

Fig. 2 Median percentage of case versions by case attribute and case version

Trends in the ANR are (near) monotonically increasing for four of the seven MAHs, (MAH01, MAH03, MAH04, and MAH07), indicating that additional follow-up reports beyond Version 1 are distributed to increasingly more HAs. Two MAHs (MAH02 and MAH05) have a maximum around Versions 2 or 3, with the ANRs tailing off for the higher versions. Finally, one of the MAHs (MAH06) shows a continuously decreasing trend as the version number increases. The magnitude of the replication level varies from approximately 1.5 recipients for MAH07, Version 1, to over five recipients for the highest case versions of MAH04.

The average level of replication (not weighted by case volume) across all MAHs is shown in Table 1. The level of replication varies significantly from this average, however, for certain case strata, as shown in Fig. 4. Replication is slightly higher for spontaneous, HCP confirmed, serious, unlisted, and literature-reported cases. These differences are likely driven by reporting requirements, which are different for these various case strata.

Table 1 Median and mean average number of recipients by case version

Version	Median ANR	Mean ANR
1	2.88	2.90
2	3.62	3.31
3	3.24	3.30
4	3.36	3.31
5	3.23	3.30
>5	3.14	3.55

ANR average number of recipients, *MAH* marketing authorization holder

The median and mean are not corrected for the differing case version volumes of the seven MAHs

4.2.2 Distribution of Number of Recipients

As discussed before, the simple arithmetic mean (or median) in combination with the submission count only provides a partial picture of the level of replication. In particular, the distribution of the fraction of case versions sent to a specific number of recipients is generally very long-tailed. For the five MAHs that provided this data, Fig. 5 shows the percentage of total case versions as a function of the number of recipients. While most case versions are submitted to fewer than ten recipients, the percentage of case versions sent to ten or more recipients (HAs) across these five MAHs is 12.4%. The individual MAH values for percent of case versions to ten or more recipients range from 0 to 26%. From the MAH perspective, the case versions submitted to ≥ ten recipients are approximately 12% of case versions but approximately 40% of submissions.

4.2.3 Analysis of Named Recipients

As the average number of recipients is three (for all cases) to four (for serious and/or unlisted cases), the question arises whether these three or four recipients are typically the same

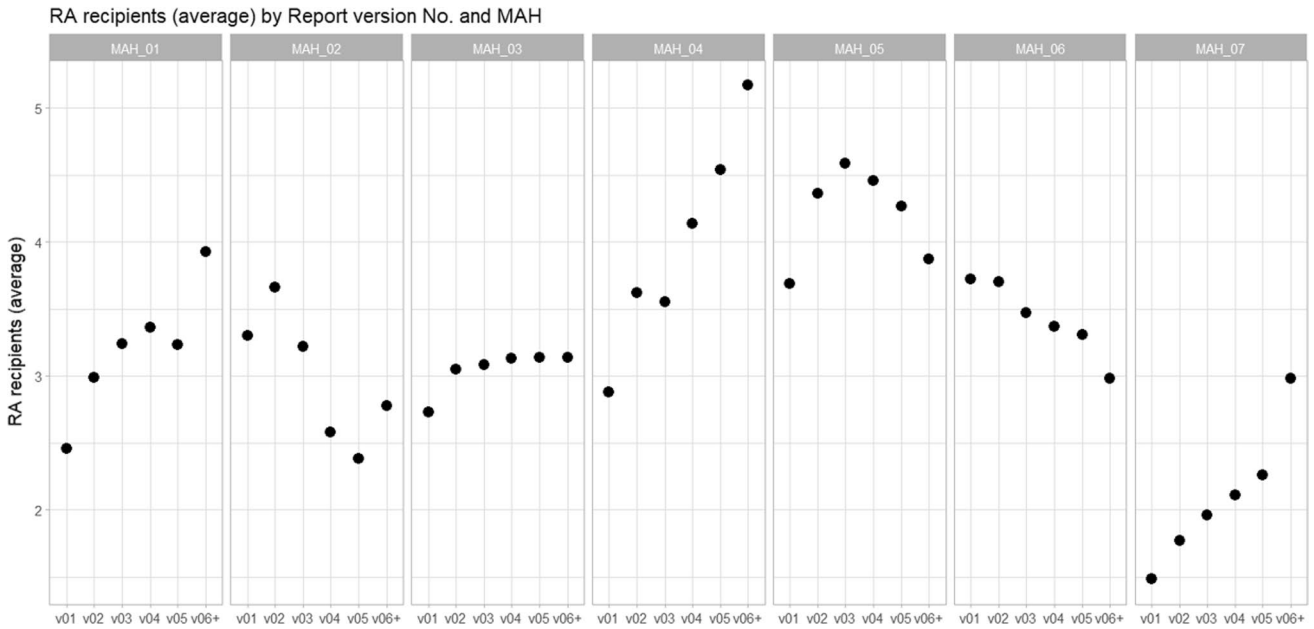

Fig. 3 Average number of recipients (ANR) by case version for each MAH

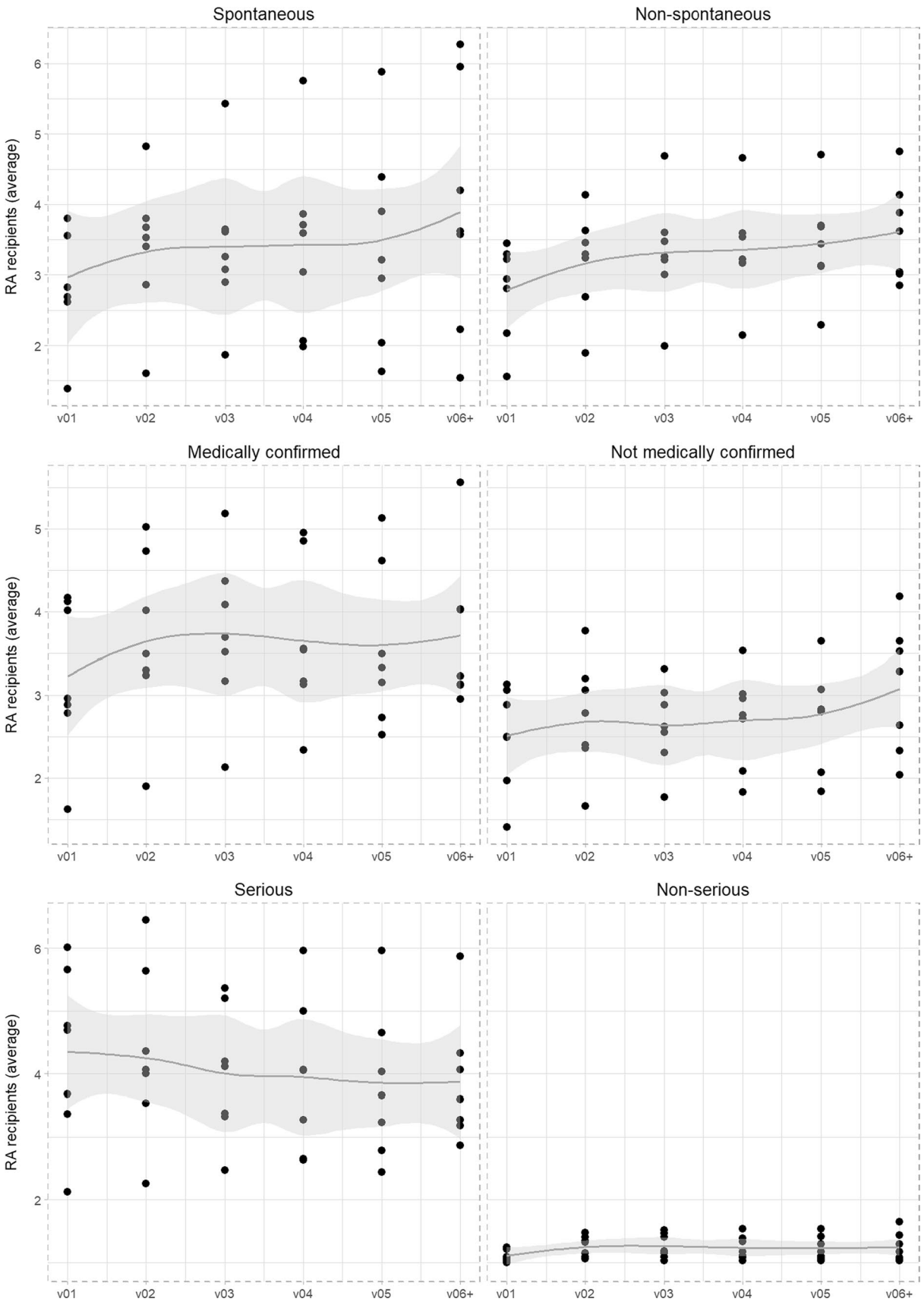

Fig. 4 Average number of recipients (ANR) by case version case attribute

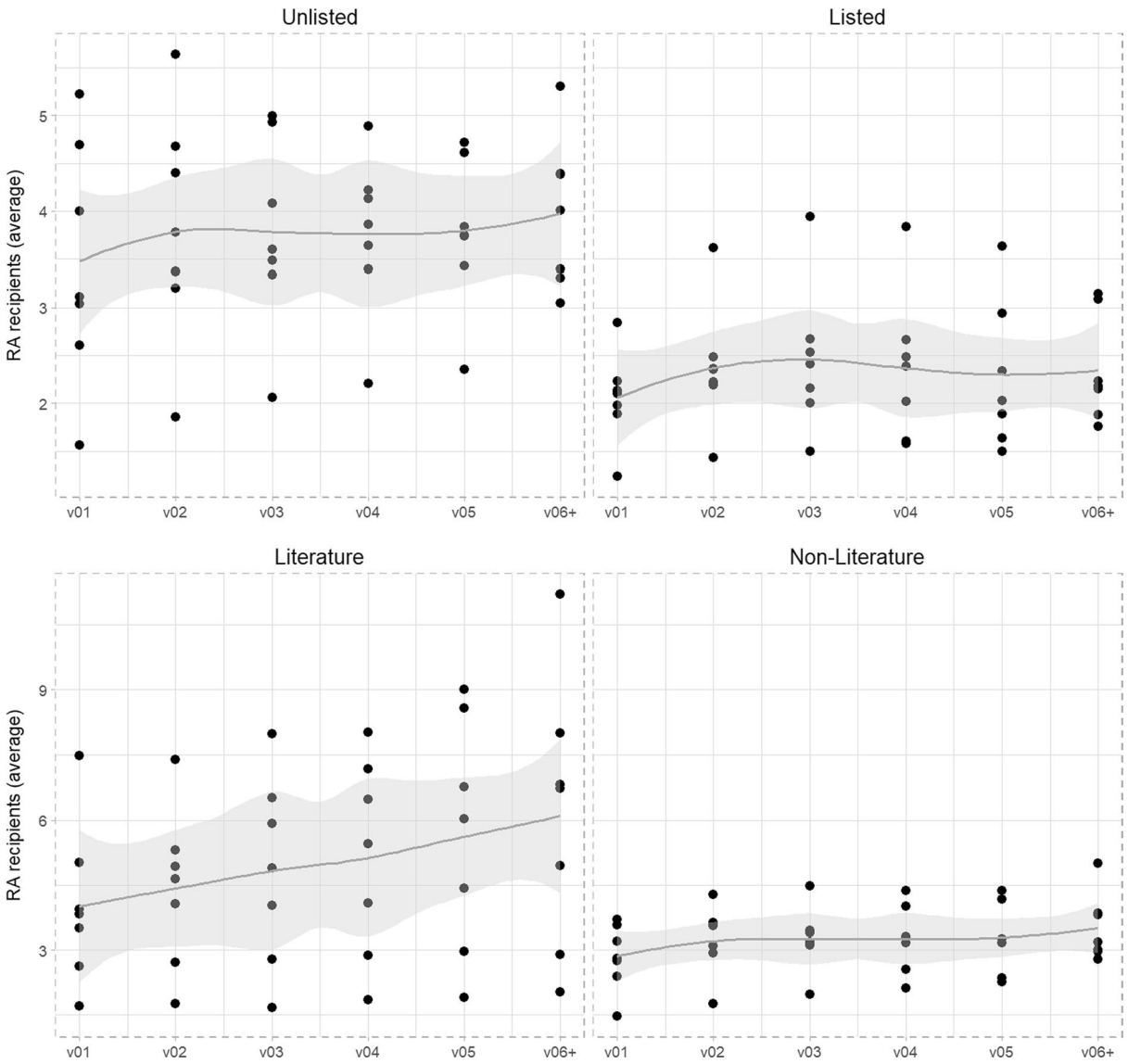

Fig. 4 (continued)

HAs receiving the case submissions. A post hoc analysis of recipient combinations was conducted in three MCs, measuring the relative percentages of occurrence of unique combinations of HA recipients. Table 2 shows these recipient combinations and their relative abundance for those ICSR submissions made to three recipients. Therefore, while an average of three MAHs receive a case version, there is considerable variability in the make-up of these recipients, which is reflective of a MAH's portfolio and marketing authorizations.

Another way of looking at this recipient-specific data is to measure what percentage of submissions is sent to the top three combinations. For submissions sent to three recipients, this percentage is 51.9%, 78.2%, and 27.9%, respectively, for MAH1, MAH2, and MAH3.

4.3 Submission Selectivity

The Submission Selectivity Measure is shown in Fig. 6 for versions 1 through 10, grouped by case version for all seven MAHs. On average, the magnitude of the Submission Selectivity increases for a given MAH, leveling off after version 5. Focusing on Version 6, the magnitude of Submission Selectivity ranges from approximately 10% to well over 50%. As shown in the smoothing line, the variation tends to cluster around 30% for the higher versions.

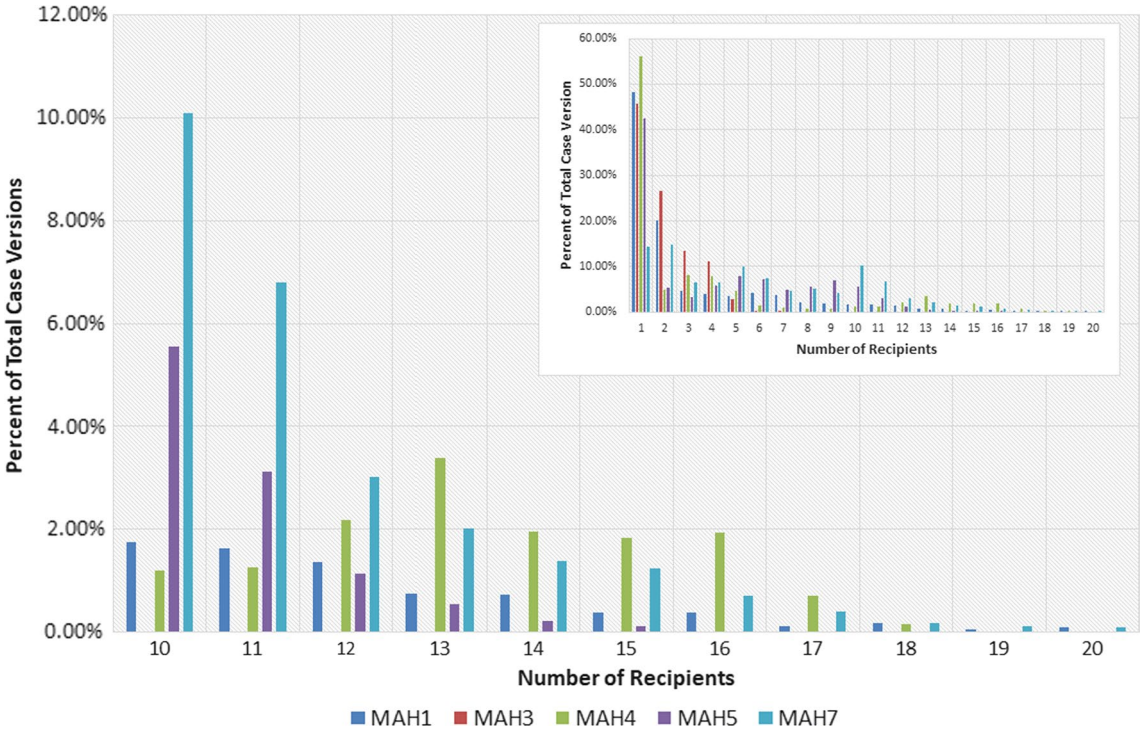

Fig. 5 Percent of total case versions vs. number of recipients. Main area = distribution tail. Picture-in-picture shows the complete distribution

Table 2 Top ten recipient combinations of Individual Case Safety Report (ICSR) submissions made to three recipients

MAH1		MAH2		MAH3	
HA recipients	Percent (%)	HA recipients	Percent (%)	HA recipients	Percent (%)
Canada; EMA; USA	23.9	EMA, N Macedonia, S Korea,	64.89	Azerbaijan, Ukraine, USA	12.49
Canada; EMA; S Korea	18.9	N Macedonia, S Korea, Taiwan	7.83	Jamaica, S Africa, Zimbabwe	8.91
Algeria; Canada; EMA	9.0	EMA, S Korea, Taiwan	5.50	Canada, Japan, S Africa	6.47
EMA; S Korea; USA	7.0	Canada, EMA, S Korea	4.90	Jamaica, Uganda, S Africa	3.93
China; EMA; S Korea	4.5	Canada, EMA, USA	2.50	Indonesia, Jamaica, D Africa	3.79
Canada; N Macedonia; S Korea	2.6	EMA, Japan, S Korea	2.47	Jamaica, Kenya, Uganda	3.61
Canada; S Korea; USA	2.1	EMA, S Korea, USA	1.38	Burkina Faso, Côte d'Ivoire, Togo	3.46
Canada; EMA; N Macedonia	1.9	Canada, N Macedonia, S Korea,	1.01	Russia, USA, S Africa	3.13
Canada; Japan; USA	1.6	N Macedonia, S Korea, USA	0.92	Jamaica, S Korea, S Africa	2.88
Canada; Russia; USA	1.5	China, EMA, S Korea	0.56	Moldova, Uganda, S Africa	2.18

4.4 MAH Overlap

The overlap between company (MAH) safety databases is shown in Fig. 7 for the five MAHs that provided the data requested in Online Resource 1 Table 4 (ESM). The vast majority of cases exist only in one company database, with a small percentage (approximately 2%) of cases that are shared ("replicated") in two or more company databases.

As the level of inter-MAH replication seemed rather low at 2%, an additional post hoc text analysis was done on the WUCINs provided by three of the five MAHs, where the sender's identifier was extracted from the WUCIN and subsequently classified as "MAH" versus "Non-MAH." MAH cases can generally be identified within the WUCIN through unique identifiers for the sender. This analysis showed that a significant percentage of WUCINs originate with partners and HAs: the percentages of Non-MAH WUCINs for the three participating MAHs are 13.4%, 15.6%, and 27.9%.

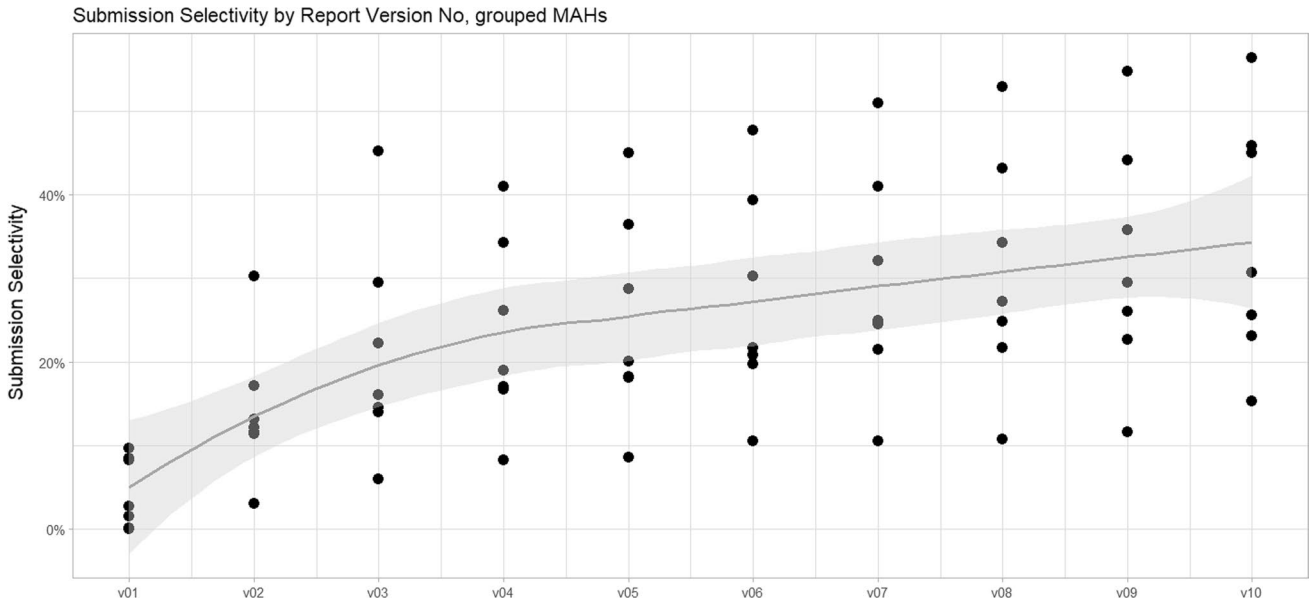

Fig. 6 Submission selectivity by case version across the seven MAHs combined

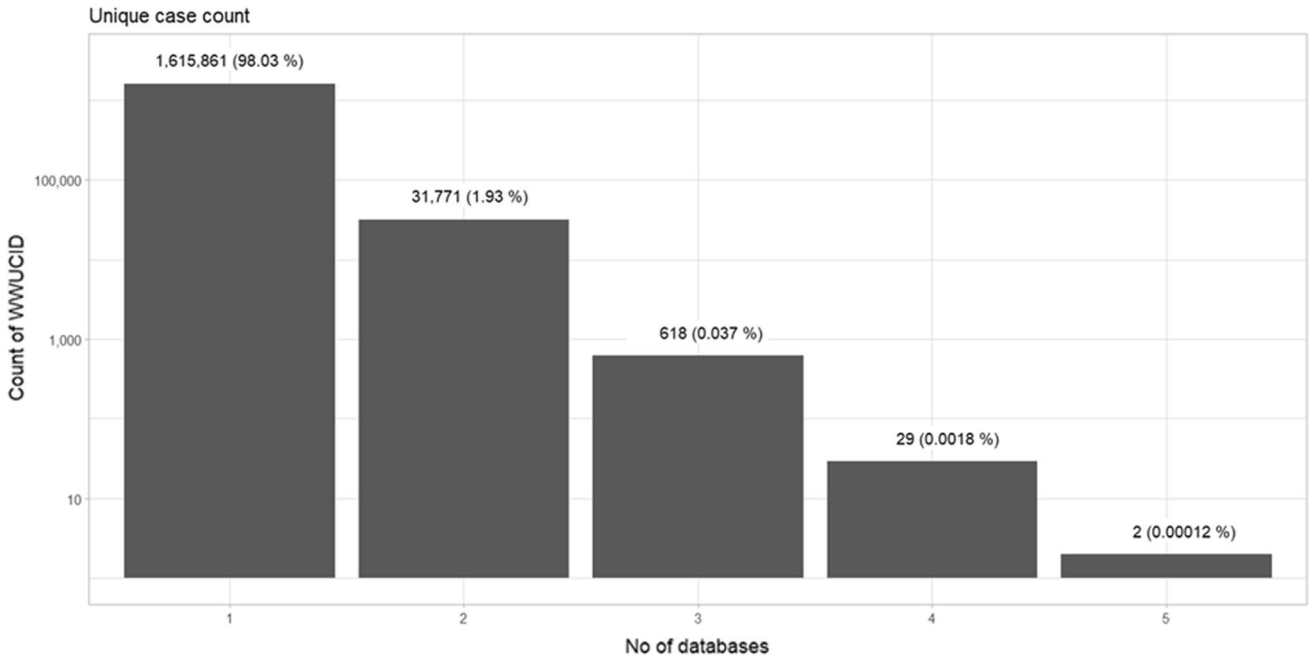

Fig. 7 Histogram of the WUCIN count and number of MAH databases that contain them

5 Discussion

5.1 Summary of Results

On average, the replication level of a case version is approximately 3.0, while serious unlisted cases are replicated on average in approximately four HA databases. Within this average of three recipients, the make-up of actual HAs is variable within a MAH and also between MAHs. In other words, even for an "average case version" that is submitted to three HAs, the actual recipients are not predominantly the same HAs. It should be clear that the variability in the specific HAs receiving a given report is much higher than simply inferred from the mean number

of three recipients. In general, replication is higher for Spontaneous cases, HCP confirmed cases, Serious cases, Unlisted cases, and Literature cases. Considering both the submission volume and the actual ANR value, the most significant drivers for replication are Serious and Unlisted cases. This is perhaps not surprising because most HAs require at least serious unlisted cases to be reported in an expedited manner.

The other main observation concerns a very long-tailed distribution of the number of recipients, providing evidence that 12.4% of cases exist in ten or more HA databases.

The variability of information received by MAHs—as measured by the Submission Selectivity measure—shows that the Selectivity for Version 1 of an ICSR is close to 0, which implies that this version exists in the same manner across HA databases, at least initially. However, as case versions increase, the Submission Selectivity level increases, meaning that different HAs have different information in their respective databases. While explanations of this phenomenon are somewhat speculative (e.g., for a given HA, the report may no longer meet all the submission reporting rules), the inescapable fact remains that there is very rapidly no longer a "canonical version" with complete information regarding a case that resides in every MAH's database. In other words, content becomes more divergent in all HA databases as the number of case versions increases, although how divergent is unknown.

The WUCIN analysis shows that a relatively low level (approximately 2%) of replication of unique cases exists across the safety databases of the participating MAHs. This, however, is representative of only the five MAHs that provided their data and therefore dependent on the co-marketing agreements between these MCs. Since large pharma companies often enter into dozens or even hundreds of these agreements, the additional text analysis on the WUCINs is deemed more representative of the extent of co-occurring cases in MAH databases. The overlap percentages in the databases of the three participating MCs (13.4%, 15.6%, and 27.9%) demonstrate that a significant fraction of cases originated from a partner's or HA's database.

5.2 Causes, Implications, and Mitigations of Individual Case Safety Report (ICSR) Replication

Across the participating MAHs, the level of replication in primary transmission for the most important cases (serious/unlisted) in PV results in a situation where each case version, on average, resides in four HA databases. The actual submission distribution is long-tailed, and over 10% of case versions are sent to ten or more recipients. As shown through the analysis of variation, these HA databases will not always have the same information, especially for higher versions where some versions may exist in one HA database but not another one. This is likely driven by specific reporting rules that may result in one recipient receiving a new version and another not receiving it. Other discrepancies may result in selective downgrades or nullifications of a case in certain countries, for example, if a product marketing authorization is changed, that then makes reporting inapplicable in a certain country. This situation results from the primary transmission process alone and is presumably exacerbated in scenarios where case versions are forwarded by their primary recipients to secondary ones. It should be noted that all analysis results are aimed at the "primary transmission" step, the first transmission of an ICSRs from the original report recipient (assumed to be a MAH) to the second recipient (assumed to be a HA). Additional transmissions may, of course, occur from the second recipient to further organizations, a step termed "secondary transmission." This secondary transmission step inevitably increases the phenomenon of replication and fracturing of information. The magnitude of this secondary transmissions step was not directly measured in our analysis; however, the post hoc text analysis of WUCINs undertaken by three of the five MAHs provides some insights retrospectively into the path taken by an ICSR through which a WUCIN number has been added by each recipient. This magnitude could be more directly estimated by analyzing secondary reporting rules (e.g., from a receiving HA to a partner HA) or by direct measurement similar to our analysis.

While this secondary transmission process was not in scope for this paper, it is fair to say that an imperfect status quo in the PV world has been in effect for a while, with rework driven by the level of replication and a lack of canonical information apparently unavoidable and significant. Although consolidation of ICSR reporting (e.g., reporting for all products authorized within the European Economic Area; EEA) has attempted to simplify and reduce the replication issue regionally, it is apparent that further improvements in ICSR reporting processes may be in order as PV frameworks globally continue to mature and drive the creation of new databases to monitor the safety of products.

Several possible options to avoid replication and variation of information may be contemplated and investigated, including:

- For literature cases, which represent a relatively small ICSR volume, and which are less time-critical, but where the highest levels of replication were observed in this study, a possible solution is managing the exchange of references through a single gateway.
- The creation of a global solution, though not necessarily in the sense of a single centralized database, a distrib-

uted system would be conceivable as well, along the lines of the Sentinel Distributed Database [6]. Such an ICSR system, intended for use by both MAHs and regulators and in which each report is held once with the necessary safeguards for ICSR read/write privileges, would require buy-in from all stakeholders, including lawmakers, medical professionals, HAs, and MAHs. With harmonized approaches, a willingness to give up some level of control is exchanged for efficiency and completeness. The consensus around "big data" analytics is that data storage is affordable, and the value of analytics on data more than covers the cost of replicating and centralizing it for easy analytics. Looking farther ahead, the general concept of transmitting cases back and forth might also be outdated [7]; an alternative approach could be one fee-based institution where MAHs book-in, take up, and maintain cases similar to a collaboration platform; economies of scale allow for growing artificial intelligence (AI) support and advanced technology instead of micro-steps performed internally by MAHs, meaningful qualitative signal detection is possible in one large place where providers, MAHs, and HAs can adhere to the FAIR principles of data management (Findability, Accessibility, Interoperability, and Reusability) [8]. However, it is acknowledged that any type of (de) centralized canonical database is aspirational and requires a long process of consensus building. As current PV operational choices on reporting are based on local requirements, these reporting requirements would need to manifest themselves in corresponding extraction rules for retrieving the data from the global platform.

- Alternatively, a significant degree of unification of worldwide regulatory requirements would go a very long way in assuring at least the similarity of content in each MAH's database. The requirement to send every case version to every HA, for example, coupled with intelligent systems to handle that reporting volume with automation, could be considered in the absence of one unique database. Obviously, this only addresses the question of diverging content and not the issue of multiple recipients. The creation of case-type specific solutions could be considered, specifically for literature cases. While literature cases do not represent a large case volume, this type of case does result in a high degree of replication. A specific approach that could prevent the inclusion or transmission of already existing literature references is the application of similarity measures for text comparison. In a similar field, Levenshtein distance [9] was applied for measuring redundancy in Electronic Health Records [10].
- Other more advanced technologies that allow the identification and tracking of a "canonical" ICSR through some type of immutable tagging without the need for a central repository could be considered as well.

Admittedly, significant consensus among the pharmaceutical manufacturers and regulators on this subject would be a necessary condition even to start investigating the possibilities in this area. However, it is in the direct interest of patient safety that complete and timely information is always available to all stakeholders, and a global distributed shared database, centralized database, harmonization of reporting rules, or a trackable ICSR may function as a significant step towards that goal. Whether the phenomenon of "replication" by itself is sufficient to trigger any concrete action towards any of the solutions proposed above remains to be seen.

5.3 Study Limitations

Finally, there are limitations of this study:

- The extent of replication described in this study is likely to be an underestimate of the true extent of replication as replication was conceptualized at a country/regional level. However, in some countries, different PV databases exist differentiated by, for example, marketed versus non-marketed products, prescription versus non-prescription, biologic versus non-biologic, etc., resulting in the ICSR being sent to multiple databases within a country.
- This study covers primary transmission only; a network transmission model which accounts for how the recipient HA processes and shares data would allow for an estimate to quantify dataflows and assess submission selectivity across the network model. It is likely that this data is generalizable across the pharma industry. Independent of the size of a pharmaceutical company, MAHs are sending replicate reports to more or less the same or at least the same average number of HAs.
- The variables in this study are not independent (e.g., there is likely some level of correlation between seriousness and listedness), and it would have been interesting to see which combination of factors leads to more replication. However, the design of this study required data to be submitted in aggregate by each MAH; therefore, these interactions could not be studied.

6 Conclusions

The main objective of the paper was to quantify the complexity of ICSR migratory patterns, as they result in the replication and divergence of information in stakeholders' information systems. This complexity is largely driven by the multitude of diverging reporting requirements across the globe and the numerous marketing agreements between MAHs.

Replication of ICSRs and the existence of fractured and incomplete safety information distributed across databases is demonstrated to be a real phenomenon. The average number of recipients per case version across the surveyed companies is approximately three recipients (four recipients for serious, unlisted cases), and there is additional variability in which actual HAs receive a case version. It was shown that a significant portion (12.4%) of case versions is submitted to ten or more HAs, making up about 40% of submissions.

A discussion on the concept of a globally distributed sharing platform in combination of retrieval rules corresponding to local regulatory requirements is perhaps best initiated and facilitated by an organization such as the ICH or a consortium of HAs.

Additional suggested research may include the exploration of secondary transmission from the initial recipients to additional parties. The value of which should be investigated.

In terms of potential solutions to alleviate the inefficiencies and inconsistencies resulting from the replication phenomenon, in-depth technical work is needed, for example, in the realm of ICSR tracing capabilities, (de)-centralized common database solutions, and other potential approaches.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40264-022-01251-7.

Acknowledgements The authors worked as part of a project team organized by TransCelerate Biopharma Inc., which is funded by 20 organizations dedicated to making research and development, including pharmacovigilance monitoring and reporting, more efficient and effective in the pharmaceutical industry. The authors also gratefully acknowledge the support of their member company teams for the considerable effort involved in collecting data for this project. This includes the advice, insight, and revised data they offered during frequent quality checks. Thanks as well to Neal Grabowski (Sanofi) for workstream leadership support and Clint Craun (TransCelerate) for project management support.

Declarations

Funding This work was supported by TransCelerate BioPharma Inc.

Conflicts of interest All authors, including John van Stekelenborg, Vijay Kara, Roman Haack, Ulrich Vogel, Anju Garg, Markus Krupp, Kate Gofman, Brian Dreyfus, Manfred Hauben, and Andrew Bate, declare no conflicts of interest. However, all authors affiliated with pharmaceutical companies are employees and, in some cases, stockholders of the companies. The views expressed in this article represent the authors' thoughts and are independent of their employers.

Ethics approval This declaration is not applicable as no human participants were involved and no protected/personal health information (PHI) was used. The data used for this study were metadata related to case transmissions, not safety case data itself.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material (data transparency) TransCelerate ensured collection, anonymization, and aggregation of member company safety data transmissions were performed and coordinated by a third-party project manager. The resulting datasets will remain confidential. Please see Electronic Supplementary Material for more information regarding data collection.

Code availability Not applicable.

Author contributions All authors contributed to the article's conception and design. All authors read and approved the final manuscript. Collection, anonymization, and aggregation of safety data were performed and coordinated by a third party project manager.

References

1. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18:427–36. https://doi.org/10.1002/pds.1742 (**Accessed May 5, 2022**).
2. Ghosh R, Kempf D, Pufko A, Martinez LFB, Davis CM, Sethi S. Automation opportunities in pharmacovigilance: an industry survey. Pharm Med. 2020;34:7–18. https://doi.org/10.1007/s40290-019-00320-0 (**Accessed May 5, 2022**).
3. Stergiopoulos S, Fehrle M, Caubel P, Tan L, Jebson L. Adverse drug reaction case safety practices in large biopharmaceutical organizations from 2007 to 2017: an industry survey. Pharm Med. 2019;33:499–510. https://doi.org/10.1007/s40290-019-00307-x (**Accessed May 5, 2022**).
4. Hauben M, Reich L, DeMicco J, Kim K. "Extreme duplication" in the US FDA adverse events reporting system database. Drug Saf. 2007;30:551–4. https://doi.org/10.2165/00002018-200730060-00009 (**Accessed May 5, 2022**).
5. Tregunno PM, Fink DB, Fernandez-Fernandez C, Lázaro-Bengoa E, Norén GN. Performance of probabilistic method to detect duplicate individual case safety reports. Drug Saf. 2014;37:249–58. https://doi.org/10.1007/s40264-014-0146-y (**Accessed May 5, 2022**).
6. Brown JS, Maro JC, Nguyen M, Ball R. Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system. J Am Med Inform Assoc. 2020;27(5):793–7. https://doi.org/10.1093/jamia/ocaa028 (**Accessed May 5, 2022**).
7. Streefland MB. Why are we still creating individual case safety reports? Clin Ther. 2022;40:1973–80. https://doi.org/10.1016/j.clinthera.2018.10.012 (**Accessed May5 2022**).
8. Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18 (**Accessed May 5, 2022**).

9. Levenshtein VI. Binary codes with correction of deletions and insertions of symbol 1. Probl. Transmission of Information. 1965;1(1):12–25. http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=ppi&paperid=69&option_lang=rus. Accessed May 6, 2022.

10. Cohen R, Elhadad M, Elhadad N. Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC Bioinform. 2013;14:10. https://doi.org/10.1186/1471-2105-14-10 (**Accessed May 6, 2022**).

Authors and Affiliations

John van Stekelenborg[1] · Vijay Kara[2] · Roman Haack[3] · Ulrich Vogel[4] · Anju Garg[5] · Markus Krupp[6] · Kate Gofman[7] · Brian Dreyfus[8] · Manfred Hauben[9,10] · Andrew Bate[2,11,12]

✉ John van Stekelenborg
jvanstek@its.jnj.com

1 Global Medical Safety, Janssen, the Pharmaceutical Companies of Johnson & Johnson, Horsham, PA, USA

2 Global Safety, GSK, Brentford, UK

3 Bayer AG, Pharmacovigilance Analytics, Berlin, Germany

4 Global Patient Safety and Pharmacovigilance, Boehringer Ingelheim International GmbH, Ingelheim am Rhein, Germany

5 Safety Analyses Innovation and Submission Readiness Lead, Global Pharmacovigilance, Sanofi, Bridgewater, NJ, USA

6 Global Patient Safety, Merck Healthcare KGaA, Darmstadt, Germany

7 CVRM Unit, Global Patient Safety Physician, CMO Patient Safety, AstraZeneca, Gaithersburg, MD, USA

8 Epidemiology Lead for Integrated Oncology, Bristol-Myers Squibb Company, Princeton, NJ, USA

9 Pfizer Inc, New York, NY, USA

10 Department of Medicine, NYU Langone Health, New York, NY, USA

11 Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

12 Department of Medicine at NYU Grossman School of Medicine, New York, NY, USA