ORIGINAL RESEARCH ARTICLE

A Comparison of Signals of Designated Medical Events and Non-designated Medical Events: Results from a Scoping Review

Daniele Sartori[1,2] · Jeffrey K. Aronson[2] · Nils Erlanson[1] · G. Niklas Norén[1] · Igho J. Onakpoya[2]

Abstract

Introduction and objective The European Medicines Agency (EMA) maintains a list of designated medical events (DMEs), events that are inherently serious and are prioritized for signal detection, irrespective of statistical criteria. We have analysed the results of our previously published scoping review to determine whether DME signals differ from those of other adverse events in terms of time to communication and characteristics of supporting reports of suspected adverse drug reactions.

Methods For all signals, we obtained the launch year of medicinal products from textbooks or regulatory agencies, extracted the year of the first report in VigiBase and calculated the interval between the first report and communication (time to communication, TTC). We further retrieved the average completeness (via vigiGrade) of the reports in each case series in the years before the communication. We categorised as DME signals those concerning an event in the EMA's list. We described the two groups of signals using medians and interquartile ranges (IQR) and compared them using the Brunner–Munzel test, calculating 95% confidence intervals (95% CI) and P values.

Results Of 4520 signals, 919 concerned DMEs and 3601 concerned non-DMEs. Signals of DMEs were supported by a median of 15 reports (IQR 6–38 reports) with a completeness score of 0.52 (IQR 0.43–0.62) and signals of non-DMEs by 20 reports (IQR 6–84 reports) with a completeness score of 0.46 (IQR 0.38–0.56). The probability that a random DME signal was supported by fewer reports than non-DME signals was 0.56 (95% CI 0.54–0.58, $P < 0.001$) and that of one having lower average completeness was 0.39 (95% CI 0.36–0.41, $P < 0.001$). The median TTCs of DME and non-DME signals did not differ (10 years), but the TTC was as low as 2 years when signals (irrespective of classification) were supported by reports whose average completeness was > 0.80.

Conclusions Signals of designated medical events were supported by fewer reports and higher completeness scores than signals of other adverse events. Although statistically significant, the differences in effect sizes between the two groups were small. This suggests that listing certain adverse events as DMEs is not having the expected effect of encouraging a focus on reports of the types of suspected adverse reactions that deserve special attention. Further enhancing the completeness of the reports of suspected adverse drug reactions supporting signals of designated medical events might shorten their time to communication and reduce the number of reports required to support them.

1 Introduction

According to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), a serious adverse drug reaction (ADR) is "any untoward medical occurrence [sc. attributed to a medication] that at any dose: results in death, is life-threatening, requires inpatient hospitalization or results in prolongation of existing hospitalization, results in persistent or significant disability/incapacity, is a congenital anomaly or birth defect, [or] is a medically important event or reaction" [1]. In many countries, any serious ADR must be reported within 15 days to the relevant regulatory agency. Two lists of medical events, currently in use internationally, include items that should be regarded as serious and requiring intervention: important medical events (IMEs) [2] and designated medical events (DMEs) [3]. The inclusion criteria for the list of

✉ Daniele Sartori
daniele.sartori@who-umc.org

1 Uppsala Monitoring Centre, Uppsala, Sweden

2 Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

IMEs are based on the ICH definition of a serious ADR. As of January 2023, the IME's list constitutes a supplement of 7525 Medical Dictionary for Regulatory Activities preferred terms (MedDRA PTs) that were deemed useful for analysing aggregated data, classifying and assessing cases in routine pharmacovigilance activities. A total of 62 of the PTs therein make up the list of DMEs, that is, a collection of PTs regarded as 'inherently serious' and considered to be 'often medicine-related'. Crucially, the purpose of the list of DMEs is to prioritize adverse events in signal detection, and it is described by the EMA as a 'safety net that ensures signals are not missed' [3].

Systematic reviews of withdrawals of marketing authorisations because of fatal ADRs have shown that the interval between the first report of a death and withdrawal of marketing authorisation in any country did not substantially change over the years between 1950 and 2013. The authors suggested that the delays may have been explained in part by the need for subsequent studies after the first indications of drug-attributed deaths [4].

In a previously published scoping review of signals of ADRs and signals of disproportionate reporting (SDRs), we identified over 10,000 signals/SDRs. The median time interval between the first report in VigiBase, the WHO's global database of reports of suspected ADRs, and the year in which a signal was communicated was 9 years [5].

We are unaware of prior research into the characteristics and timing of signals of DMEs compared with signals of other ADRs. Having not explored these aspects in our previous study, we therefore sought to fill this knowledge gap by characterizing types of evidence and timeliness of communication (i.e. the written transmission or exchange of information pertaining to signals, such as minutes of committee meetings), comparing and contrasting these two types of signals.

2 Methods

2.1 Data Sources

We used the dataset from our published scoping review of the evidence underpinning signals [5], which included studies of signals/SDRs communicated by stakeholders in pharmacovigilance between 1986 and 2020. Each study in the dataset had had a level of evidence attributed according to the Oxford Centre for Evidence-Based Medicine (OCEBM) classification tool; study designs were thus ranked from 1 (available evidence of the highest quality for decision-making) to 4 (lowest quality), following the row "What are the rare harms?". We retained the postulated subtypes of evidence from the scoping review to ensure granularity. The highest level of evidence applied when multiple studies supported a signal (e.g. if a signal was supported by a meta-analysis of randomised controlled trials and by reports of ADRs, the signal was categorised as OCEBM 1). To avoid undue biases in analysing possible delays in communicating signals/SDRs and other variables (see Sect. 2.2), we excluded signals/SDRs detected from studies whose aims were to develop, validate, or evaluate novel methods for signal detection, as these studies do not urge prompt regulatory or verificatory action. For the full list of excluded studies, see Supplementary Materials 1. We mapped the signals/SDRs from the scoping review to the MedDRA dictionary and anatomical therapeutic classification (ATC); for details see Supplementary Materials 1.

We obtained the list of DMEs from the European Medicines Agency (EMA), as published on 15 June 2020, to categorize DME signals that concerned at least one of the MedDRA (v. 23.1) PTs, whether the event of interest was composite or not. Conversely, we classified signals as 'non-DME' if all the events mapped to terms outside the EMA's list.

2.2 Variables

We extracted data from VigiBase using Structured Query Language together with Python. We queried a deduplicated [6] and frozen version of VigiBase (lock point: 30 August 2020), to extract a range of characteristics of reports containing the medicinal products (standardized to WHODrug, B3/C3 format, 01/09/2020) and adverse events involved in DME and non-DME signals, setting involvement of the medicinal

products as either 'suspected' or 'interacting'. Thus, any characteristics of DME signals we retrieved referred only to reports that included PTs belonging to the list of DMEs, and the same applied to non-DMEs. For each DME or non-DME signal/SDR, we obtained the following information:

(1) The first year in which a report was entered into the database (E2b fields: FirstDateDatabase or Receive-Date, whichever was earlier) and the year in which three such reports became available in the database.
(2) The first year in which the reporting of a medicinal product-event combination became disproportionate, using the information component (IC [7]) for medicinal product-event combinations and the omega interaction measure (Ω [8], for drug–drug-event combinations), defined as a positive IC_{025} or Ω_{025}. The omega interaction measure is based on a model with additive risk for the occurrence of adverse events during concomitant use of non-interacting drugs.
(3) The number of reports up to the year before that of communication.
(4) A breakdown by type of report, namely: spontaneous, from a study, from prescription event monitoring or special monitoring, unknown type, or 'other' (i.e. literature reports whose type, whether spontaneous or from a study, could not be ascertained at submission to a database or from follow-ups).
(5) The average vigiGrade completeness score of the case series [9]; a vigiGrade completeness score of 1 is assigned to a report with detailed information on time to onset, patient age and sex, indication for treatment, dosage, outcome, type of report, type of notifier and country of origin, plus some free text description; the score falls by a pre-specified multiplicative factor for each piece of information that is missing.
(6) The number of positive dechallenges and/or rechallenges at the medicinal product-event level.

We established the earliest known launch year for each medicinal product, cross-referencing the websites of 27 regulatory agencies and hand searching textbooks (for a complete list, see Supplementary Materials 1). From the dataset of the scoping review, we obtained the earliest year in which a stakeholder, irrespective of country of origin, first communicated a signal/SDR.

2.3 Data Analysis

We summarised data using proportions, medians and interquartile ranges (IQR) and plotted the data using boxplots or stacked bar charts, with years of communication or of launch as independent variables and the characteristics presented above as dependent variables. For boxplots involving

completeness scores, we categorized as 'well-documented' values strictly above 0.80 (as per [9]); as 'below average or average' those from 0.00 to 0.46 inclusive, based on the average completeness in the whole of VigiBase as of August 2020, and as 'above average' values between the two categories.

To calculate the delays in communication (time to communication, TTC), we subtracted the year of the first report, irrespective of country of origin, from the year of communication. When we had sufficient data, we computed the difference between (1) the year of communication and the year in which at least three reports had accumulated, (2) the year of communication and that in which a signal became disproportionate and (3) the year of communication and the launch year.

Pilot analyses showed that the data were not normally distributed and heteroskedastic, violating core assumptions of some statistical tests (e.g. the Wilcoxon rank-sum test). We therefore required a non-parametric test that made no distributional assumptions and chose the Brunner–Munzel test [10, 11]. This test determines whether there is stochastic equality between two groups, by comparing the entire empirical distribution functions of a variable in the two groups and accounting for their means, variances and other distributional properties. Essentially, it checks whether the probability that a randomly chosen observation from one group is higher (or lower) than a randomly chosen observation from the other group is equal to the probability of the reverse, across the whole range of values of the variable (i.e. $P(X<Y) + 0.5 \times P(X = Y)$). We ran two-tailed Brunner–Munzel tests across time periods or levels of evidence, comparing DME versus non-DME signals or (non-)DME versus (non-)DME signals. We computed P values and 95% confidence intervals (CI) [10]. In summary, when applied to this study, sample estimates exceeding 0.50 (or 50%) suggest that a random observation in the DME group has a higher probability of being lower than a random observation in the non-DME group. Sample estimates below 0.50 suggest a lower probability of being lower. For a result to be statistically significant, the sample estimates had to fall within the 95% CI and have a P value < 0.05. A hypothetical comparison of a variable X across two groups, 1 and 2, using the Brunner–Munzel test, which produces a statistically significant sample estimate of 0.62 suggests an estimated probability of 0.62 of observing lower values of X in group 1 than in group 2.

Data for calculations of statistical measures were managed in Microsoft Excel, whereas calculations and figures were made in R Statistical Software (v. 4.2.0) [12].

3 Results

We obtained 10,861 signals/SDRs from the dataset included in the primary analysis of the scoping review. Of the 4520 signals/SDRs remaining after exclusion of methods studies, 919 (20%) concerned DMEs and 3601 (80%) non-DMEs. 195 (4.3%) signals/SDRs were on drug–drug interactions, 37 (19%) of which were DMEs and 158 (81%) non-DMEs. A total of 3937 (87%) had at least one report in VigiBase in the year before communication, 3639 (80%) had at least three reports in VigiBase and 2448 (54%) were disproportionate as of the year of communication. The full results of our analyses are available in Supplementary Materials 1. Supplementary Materials 2 includes the dates of launch of the medicinal products.

3.1 Descriptive Analyses

We observed different median numbers of reports for each signal between DME and non-DME signals in the dataset; each DME signal was supported by 15 reports (IQR 6–38 reports), the others by 20 (IQR 6–84 reports). There were also differences in average completeness scores: each DME report had a median of 0.52 (IQR 0.43–0.62) and each non-DME report a median of 0.46 (IQR 0.38–0.56). The median numbers of dechallenges, rechallenges, countries of origin and report types were equal across the two categories.

3.1.1 Timing of Communications

DMEs and non-DMEs had the same medians across all the three measures of the timings we prespecified, i.e. the time from first report to communication, or TTC (9 years), from the year in which there were three reports to that of communication (7 years) and from the year in which a signal became disproportionate in VigiBase and that of communication (6 years).

There was an apparent increase in the median TTC over time (Fig. 1). In fact, for DME signals the TTC nearly doubled from a median of 5 years (IQR 2–14 years) during 1986–2005 to a median of 9 years (IQR 5–14 years) during 2006–2020. For non-DME signals, the medians were 4 years (IQR 2–9 years) for the first period, and 10 years (IQR 5–17 years) for the second. The same held when using the year in which there were at least 3 reports in VigiBase and that in which signals/SDRs became disproportionate as independent variables (Supplementary Materials 1).

DME signals supported by 'well-documented' reports (completeness score > 0.8) had a median TTC of 2 years (IQR 2–4; Fig. 2). This was about five times shorter than the median TTC for DME signals supported by reports of 'above average' completeness (9 years, IQR 5–14 years, vigiGrade score 0.47–0.80) or 'below average or average' completeness (10 years, IQR 4–15 years, vigiGrade score 0.00–0.46). Similar median TTC values applied to non-DME signals (3, 8 and 10 median years, for the respective classifications of completeness).
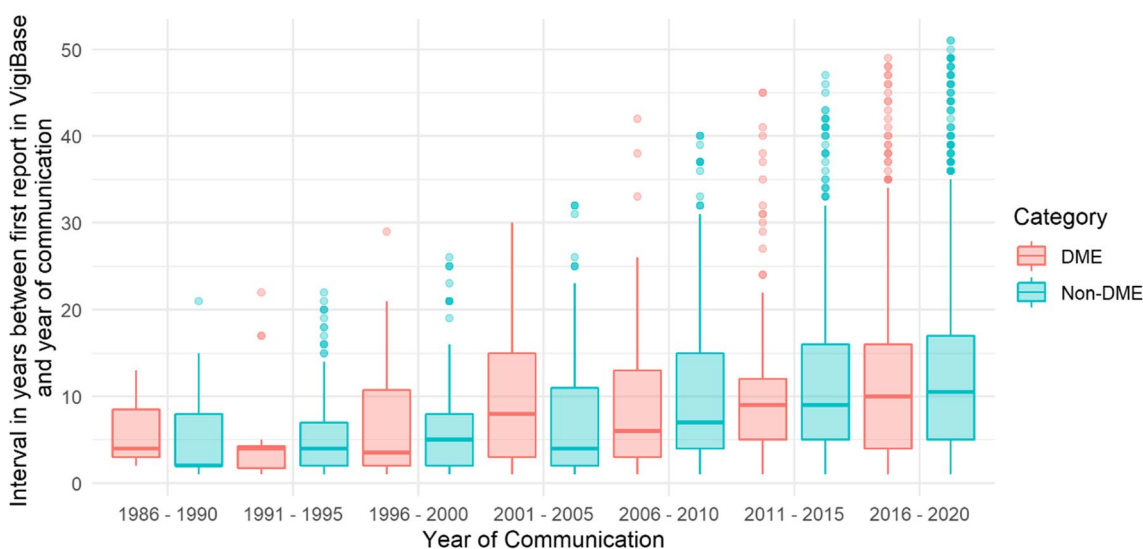
Fig. 1 Box plots of 3937 signals/signals of disproportionate reporting, categorized as DME (designated medical events, red) or non-DME (blue), with at least one report in VigiBase before communication and a positive, non-zero, time to communication. On the x-axis, the years of communication in 5-year periods, on the y-axis the delay in years in communicating signals, with interquartile ranges (whiskers); median values are indicated by horizontal lines within the boxes

3.1.2 Interval from Launch to Communication

The median launch to communication interval was 15 years for DMEs (IQR 6–30 years) and 14 years for non-DMEs (IQR 6–28 years). There was an apparent increase in the proportion of DME or non-DME signals concerning medicinal products that were launched 10 years (inclusive) before communication; for the period 1986–2005, the proportion of signals whose medicinal products were launched 10 years (inclusive) before communication was 49%, while in 2006–2020 it rose to 64%.

3.1.3 Levels of Evidence

The 4520 signals/SDRs fell predominantly under OCEBM level 4 (3983, 88%) and its subtypes; 2203 (49%) signals were supported by clinical assessments of reports of ADRs, of which 479 were classified as DME signals (22% of 2203) and 1724 (78%) as non-DME signals. Of all OCEBM 4 signals, 1443 (32%) were supported by analyses of disproportionality, with 307 (21% of 1443) categorised as DME signals and 1136 (79%) non-DME signals.

In Fig. 3, we show the time intervals to communication, from first report in VigiBase and from launch year. The results suggest higher medians of either intervals for signals/SDRs belonging to OCEBM 4, with minor differences between signals of DMEs and non-DMEs.

3.2 Statistical Analysis

The full results of the statistical analysis are available in Supplementary Materials 1. We report the main findings from the comparison of DME and non-DME signals using the Brunner–Munzel test in Table 1. There were statistically significant differences in average completeness, numbers of reports, and numbers of dechallenges and rechallenges. The only time interval that was statistically significant was that between the year in which a signal became disproportionate in VigiBase and the year of communication.

In Table 2, we report additional comparisons across DME signals, relevant to the descriptive analyses, and in further support of the apparent trends shown in the figures above. In comparisons across OCEBM level and average completeness of the case series, we noted large effect sizes in TTC, numbers of reports and differences in intervals between launch and communication. We obtained similar
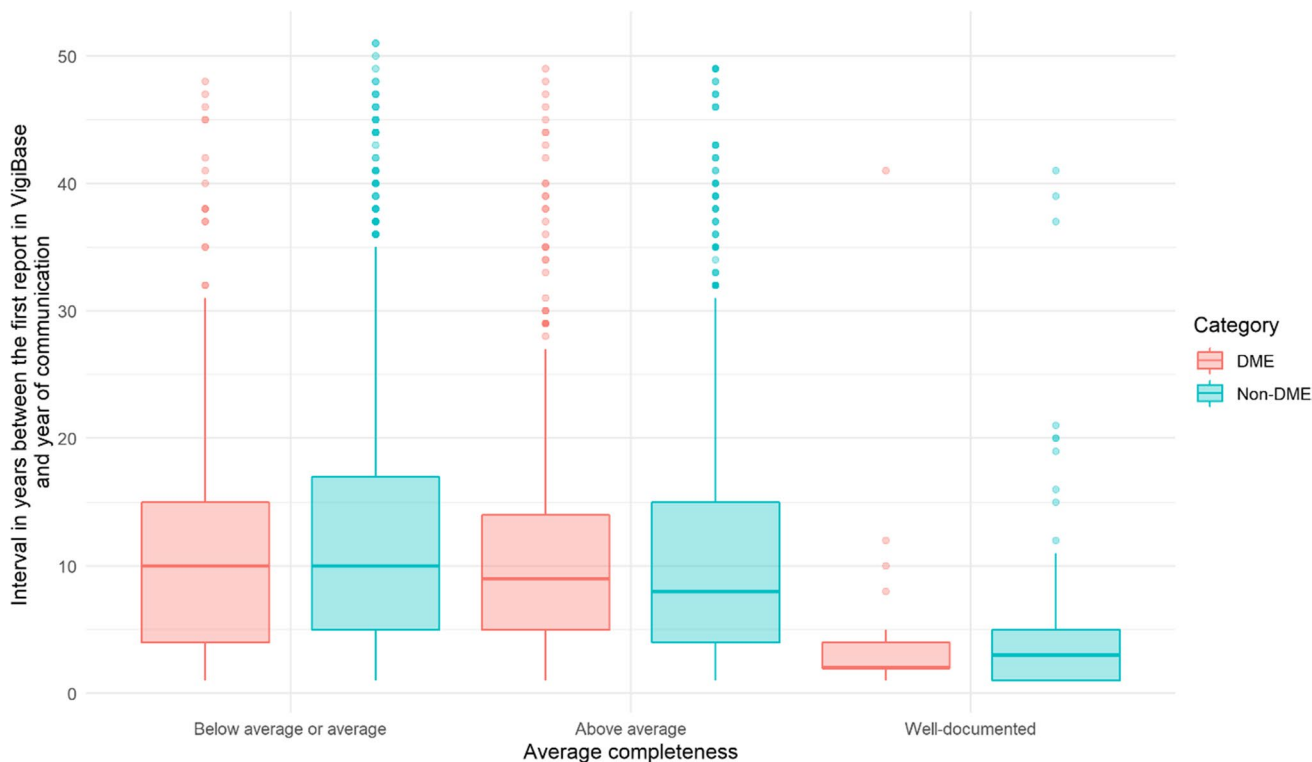
Fig. 2 Time to communication versus average completeness for 3937 signals/signals of disproportionate reporting with at least one report in VigiBase in the years leading up to that of communication. 'Well documented' was defined as an average completeness of strictly above 0.80; 'below average or average' refers to an average completeness score between 0.00 and 0.46 based on the average completeness in VigiBase as of August 2020; median values are indicated by horizontal lines within the boxes
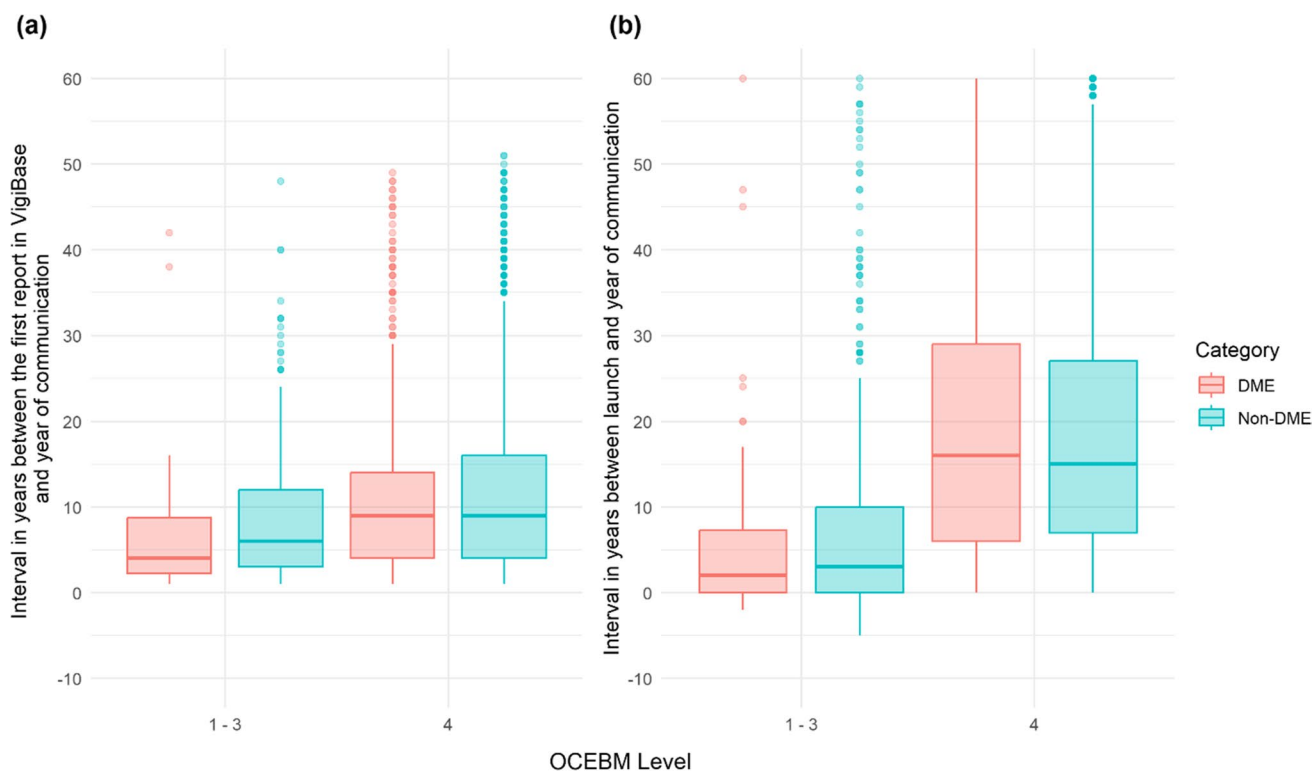
Fig. 3 Box plots of **a** time to communication for 3858/3937 signals/ signals of disproportionate reporting (SDR) by Oxford Centre for Evidence-Based Medicine (OCEBM) level, excluding studies with unclear design. **b** Box plots of the intervals in years between launch and communication of 4426/4513 signals/SDRs by OCEBM level, excluding studies with unclear design. OCEBM levels 1 through 3 were aggregated. The *y*-axis in (**b**) was truncated at 60 years (max, 119 years). Median values are indicated by horizontal lines within the boxes

findings for non-DME signals, all of which are reported in Supplementary Materials 1.

3.3 ATC and MedDRA System Organ Classes of the Signals

The results of ATC and MedDRA System Organ Class coding are in Supplementary Materials 1. There were no unexpected imbalances in the proportions of DME or non-DME signals across either of these standardized terminologies.

4 Discussion

4.1 Summary of Key Results

This analysis highlights statistically significant differences in the characteristics of the case series of DME and non-DME signals as they appeared in VigiBase up to the year of communication. These were: the numbers of reports and their average completeness and the counts of positive

dechallenges and rechallenges. Furthermore, except for a statistically significant difference in the interval in years between the first indication of disproportionality and the year of communication, the timing of communication did not differ between the two groups of signals. Finally, we found statistically significant patterns shared between DME and non-DME signals, such as the strong association between TTC and both completeness score and level of evidence, and how the time after launch of medicinal products relative to communication increased over the last 15 years of the study period.

4.2 Statistical Significance and Practical Relevance

While the comparisons of DME and non-DME signals were statistically significant, they were accompanied by small effect sizes in Brunner–Munzel estimates for numbers of reports, dechallenges, rechallenges and average completeness scores. This was especially surprising in relation to the difference in numbers of reports, as one might expect a DME signal in some cases to be based on as few as one report (i.e. 'between-the-eyes' adverse reactions [13]). However, DMEs tended to be supported by a number of reports exceeding by

Table 1 Results of the Brunner–Munzel test comparisons between designated medical events (DME) and non-DME signals, with respect to the characteristics of the case reports in VigiBase and the time intervals considered in the study

Characteristics	Brunner–Munzel Test	
	Sample estimate (95% CI)	P value
Number of reports	0.56 (0.54, 0.58)*	< 0.001
Number of countries	0.52 (0.50, 0.54)	0.022
Number of dechallenges	0.53 (0.51, 0.55)*	0.0024
Number of rechallenges	0.56 (0.54, 0.57)*	< 0.001
Average completeness	0.39 (0.36, 0.41)*	< 0.001
Intervals in years to communication		
From first report (TTC)[a]	0.51 (0.49, 0.53)	0.24
From 3 reports[b]	0.51 (0.49, 0.53)	0.31
From disproportionality[c]	0.54 (0.51, 0.57)*	0.007
From launch[d]	0.49 (0.47, 0.51)	0.29

We considered as statistically significant sample estimates within the 95% confidence interval that did not cross 0.50, with P values strictly less than 0.05, and marked these results with an asterisk (*). Notes: In this case, one may note the point estimate of 0.51 and the 95% CI crossing 0.50; no inferences can be made from this, other than that the point estimate does not reach statistical significance. For a statistically significant example, we may consider the Brunner–Munzel sample estimate for average completeness, i.e. 0.39 (95% CI 0.36–0.41, P < 0.001). This indicates that if one were to browse the average completeness in the group of reports of DMEs and the average completeness in the group of reports of non-DMEs, a randomly selected DME report would have a 0.39 estimated probability that its average completeness was lower than that of a randomly selected non-DME report

CI confidence interval, *TTC* time to communication

[a]Refers to 3937 observations where the difference in years was strictly above 0

[b]Refers to 3639 observations where the difference in years was ≥ 0

[c]Refers to 2448 observations where the difference in years was ≥ 0

[d]Refers to 4513 observations out of 4520, for which we could retrieve a launch year

several times the (canonical) minimum of three required for signal detection [14]. Less stringent criteria for signal detection when fatal events are involved have been previously advocated [15], and the same may be extended to DMEs. The EMA states that member states use the categorisation of a range of adverse events as DMEs to focus on reports of suspected adverse reactions that deserve special attention. However, we could not find evidence of such an effect in our analysis, and prioritization of such signals may require further attention. An important consideration is that the size of case series in VigiBase may have been larger than those on which the communicated signals were based. In view of this, it may be helpful for pharmacovigilance stakeholders to consult global databases of case reports when a signal of DME is detected to ensure more data are available for its clinical assessment.

4.3 Relationship Between Strength of Evidence and TTC

Irrespective of categorization into DME or non-DME signals, we found not only statistical significance but also larger effect sizes in the association between the TTC and the strength of evidence. Whether in the form of higher quality evidence (i.e. OCEBM 1–3) or high average completeness of the information in an underlying case series (i.e. 'well documented'), the strength of evidence appeared to be linked to an up to fivefold shorter TTC (Fig. 2 and Table 2).

4.3.1 Relationship Between OCEBM Level and TTC

A possible contributor to the observed relationship between OCEBM level and TTC may be that evidence of higher quality (OCEBM 1–3) tends to be collected and appraised in pre-approval stages, as evidenced, in part, by the negative intervals between launch and communication (Fig. 3b). Conversely, evidence of lower quality (OCEBM 4) begins to accrue later, during post-marketing; in this phase, signals are detected mainly through reports of ADRs and are continuously prioritized as per good vigilance practices through analyses of patient exposure and estimates of frequencies of ADRs [16]. Limitations inherent to the systems for collecting reports of ADR, such as under-reporting or low completeness of the reports, may have further contributed to the relationships observed in Table 2. Nevertheless, the types and frequencies of ADRs detected through pre- and post-marketing are different, the latter phase being primarily concerned with rare ADRs.

4.3.2 Association Between Completeness of Information in a Case Series and TTC

Well-documented reports have been associated with 'certain', 'probable' or 'possible' outcomes of causality assessments or with reports flagged as serious by international standards [17, 18]. We should stress that completeness of reports is accounted for in methods for signal detection, such as vigiRank [19], which increase the rate of detected signals compared with disproportionality analysis [20], but does not necessarily have a bearing on the timeliness of signal detection from disproportionality analyses [21]. Rather, completeness of reports has been regarded as useful in performing clinical reviews [14], which also constitute the main type of evidence underpinning signals [5]. We reiterate that although the difference in median completeness score between DME and non-DME signals was statistically significant, the effect size was small and we did not record a difference in TTC for DME and non-DME signals supported by well-documented reports. These results may call for improved international collaboration between regulators and reporters, with the aim

Table 2 Brunner–Munzel test results for the comparisons over completeness score, level of evidence and communication year for signals of designated medical events

Groups of DME signals for comparison (Group 1 versus group 2)		Variable	Sample effect size estimate (95% CI)	P value
Well documented	Other categories of completeness	TTC	0.84 (0.79, 0.89)*	< 0.001
Well documented	Other categories of completeness	Number of reports	0.88 (0.84, 0.91)*	< 0.001
OCEBM 4	OCEBM 1–3	TTC	0.35 (0.24, 0.45)*	0.005
OCEBM 4	OCEBM 1–3	Interval between launch year and communication year	0.19 (0.13, 0.26)*	< 0.001
Communication year 1986–2005	Communication year 2006–2020	TTC	0.61 (0.54, 0.67)*	0.0012
Communication year 1986–2005	Communication year 2006–2020	Interval between launch year and communication year	0.57 (0.52, 0.63)*	0.005

For 'well documented' and 'other completeness categories', see Methods. We considered as statistically significant sample estimates within the 95% confidence interval that did not cross 0.50, with P values strictly less than 0.05, and marked these results with an asterisk (*)

CI confidence interval, *TTC* time to communication, *OCEBM* Oxford Centre for Evidence-Based Medicine level

of increasing the completeness of information in reports of suspected ADRs, as means of facilitating clinical reviews and expediting the TTCs of both DME and non-DME signals. The matter of completeness becomes more relevant when one appreciates that the volume of reports in databases has increased substantially over the past 30 years [22] and may continue to do so as developing countries progress towards more mature pharmacovigilance systems [23, 24]. Since a high degree of completeness is not always achievable [25–27] and completeness may vary across settings or attitudes of health carers and patients towards reporting [28], any intervention geared towards increasing the completeness of reports would probably be a complex one [29].

4.4 Increase in TTC Over Time

Our observations support prior research showing a growing proportion of signals supported by medicinal products launched over 10 years before communication. Early reviews of signals of the Pharmacovigilance Risk Assessment Committee suggested that the concerned medicinal products were on the market for a median of 12 years and 42% of them for less than 10 years [30]. In 2010, the median time on market of medicinal products involved in regulatory actions in the USA was 11 years [31]. Thus, on one hand, the increase in TTC may reflect evolving pharmacovigilance systems, able to manage signals concerning medicinal products that have been on the market for several decades, as noted in [30], namely: improved monitoring, completion of long-term observational studies to evaluate suspected harms or changes in patterns of use of medicinal products. On the other hand, it is worth considering that some adverse effects may be only detected with enough length of exposure; indeed, medicinal

products that require longer durations of exposure have been found to be associated with larger numbers of amendments to product information [32]. In addition, the amount of post-approval exposure data (rather than pre-approval) predicts changes to the sections of untoward effects, and warnings and precautions, in European summaries of products characteristics (SmPCs) [33]. Taking these insights together, the increase in TTC may be conditional on the time needed to accrue sufficient data in the postmarketing phase, a time that may have been longer for some classes of medicinal products.

4.5 Strengths and Limitations

We compared large sets of DME and non-DME signals, relying on systematically collected data covering roughly 30 years. We used a heteroskedasticity-robust statistical method to compare the two groups of signals, ensuring intuitive interpretability of the results, excluding signals that may have distorted calculation of the intervals we had postulated. We are not aware of similar published work. These findings may provide a way forward for regulators and researchers in prioritising and communicating signals of rare events that are typically associated with medicines.

As this study concerned any reported signals/SDRs, irrespective of regulatory requirements for action or verificatory studies, our findings are relevant to the communication of signals alone. In other words, any differences or lack thereof we have identified may not necessarily solely concern signals that have significant effects on public health [16].

We used the list of DMEs rather than the list of important medical events (IMEs), since the latter includes far

more events that may cause a report to be marked as 'serious' by international standards [34]. Both lists, however, presuppose seriousness; we chose to use DMEs, as they are regarded as often drug related. Relatedly, we did not quantify proportions of serious reports in either group of signals, so we cannot conclude whether DME signals were supported mostly (or not) by reports marked as non-serious.

Findings about TTC should be considered carefully. Although we have manually verified dates of receipt of reports (in VigiBase or at the national centres level) that were discrepant with launch years, we could not control for potential data entry errors in VigiBase. In addition, data retrieval was based on the definition of the events in the original publications; in the case of composite events, we retrieved all relevant MedDRA preferred terms. Consequently, frequently reported events may have biased the TTC of some of the most recently communicated signals/SDRs.

In our search for launch years, we have encountered minor mistakes in the available data sources (and have reported them to the data holders). For medicinal products launched in countries that no longer exist (e.g. Eastern Germany since 1978), some dates may have been replaced by default values by regulatory agencies, but we did not encounter enough examples (four in all) to warrant concern. Furthermore, we did not systematically evaluate any discrepancies between the sources we used to obtain launch years and the published literature, so there may be instances in which some medicinal products may have been launched earlier than recorded. More important is the effect of censoring, which may not have allowed sufficient time for an ADR to be recognized by reporters for medicinal products launched in recent years.

The method used to compute the completeness of a case series measures technical completeness but not clinical utility. In other words, formally complete case reports may still not necessarily contain sufficient information to produce a clinically sound judgment on a possible causal relationship between a medicinal product and an adverse event. It may well be that signals that were communicated rapidly contained a higher degree of clinically relevant information, which we could not measure.

5 Conclusions

We found that DME and non-DME signals differed in number of reports, completeness score and counts of positive dechallenges and rechallenges. The differences in the effect sizes were small, albeit statistically significant. The threshold in the median number of reports supporting DME signals by far exceeded the minimum number of reports required for detecting any signals. As such, the list of DMEs may not be attaining its intended purpose of prioritizing signals that

concern suspected ADRs deemed to be often medication-related. The stronger the evidence, either as completeness of the case reports or in the form of higher quality of evidence, the shorter the TTC in both cases. Because clinical reviews of reports of ADRs are the main type of evidence that supports signals, we suggest that improved quality of reports may come with better prioritisation of communication of DME and non-DME signals alike.

Declarations

Conflicts of Interest J.K.A. has written papers on adverse drug reactions in peer-reviewed journals and has received royalties from textbooks that he has edited or co-edited; he has often acted as an expert witness in cases involving adverse drug reactions, most often in coroners' courts. D.S. and G.N.N. are employed by the UMC, which manages VigiBase and holds the archives of the signal document used to retrieve part of the included studies. Part of the reviewed materials was published in peer-reviewed journals, drug bulletins and abstracts or posters by researchers presently or formerly belonging to this foundation. Specifically, D.S. has authored signals published on drug bulletins, as abstracts or posters, some of which were included in this review, and received funding for doctoral studies from the UMC. N.E. (formerly UMC) has no competing interests. I.J.O. has no competing interests.

Ethics Approval This research relied on previously collected, anonymised, non-National Health Service data. As such, ethics approval was not required.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material The data that support the findings of this study are not publicly available. Access to the data is restricted based on the conditions for access within the WHO Programme for International Drug Monitoring. Subject to these conditions, data is available from the authors on reasonable request.

Code Availability The code used to retrieve the dates, counts and characteristics of the reports from VigiBase and that used to compute Brunner–Munzel statistics, will be made available on reasonable request.

Authors' Contributions D.S. designed the study, with support from I.J.O., J.K.A. N.E. and G.N.N. joined the research team later; the first carried out data retrieval from VigiBase, the other provided essential suggestions for data analysis. This manuscript was written by D.S. with support from I.J.O., N.E., G.N.N. and J.K.A. All authors read and approved the final manuscript.

References

1. Council for International Organizations of Medical Sciences. Glossary of ICH terms and definitions. Compiled by CIOMS from the International Council for Harmonization (ICH)'s Quality Safety Efficacy Miscellaneous Guidelines. Version 3. 20 April 2023. Available from: https://cioms.ch/publications/product/glossary-of-ich-terms-and-definitions. Accessed 20 Jun 2023.

2. European Medicines Agency. Inclusion/exclusion criteria for the "Important Medical Events" list. 2021. https://www.ema.europa.eu/en/documents/other/inclusion-exclusion-criteria-important-medical-events-list-meddra_en.pdf. Accessed 1 Jan 2023.

3. European Medicines Agency. Signal management—designated medical events. 2023. https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/signal-management#designated-medical-events-section. Accessed 1 Jan 2023.

4. Onakpoya IJ, Heneghan CJ, Aronson JK. Delays in the post-marketing withdrawal of drugs to which deaths have been attributed: a systematic investigation and analysis. BMC Med. 2015;13(1):26. https://doi.org/10.1186/s12916-014-0262-7.

5. Sartori D, Aronson JK, Norén GN, Onakpoya IJ. Signals of adverse drug reactions communicated by pharmacovigilance stakeholders: a scoping review of the global literature. Drug Saf. 2023;46(2):109–20. https://doi.org/10.1007/s40264-022-01258-0.

6. Norén GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. Data Min Knowl Discov. 2007;14(3):305–28. https://doi.org/10.1007/s10618-006-0052-8.

7. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. Stat Methods Med Res. 2013;22(1):57–69. https://doi.org/10.1177/0962280211403604.

8. Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug–drug interaction surveillance. Stat Med. 2008;27(16):3057–70. https://doi.org/10.1002/sim.3247.

9. Bergvall T, Norén GN, Lindquist M. vigiGrade: a tool to identify well-documented individual case reports and highlight systematic data quality issues. Drug Saf. 2014;37(1):65–77. https://doi.org/10.1007/s40264-013-0131-x.

10. Brunner E, Munzel U. The nonparametric Behrens-Fisher Problem: asymptotic theory and a small-sample approximation. Biom J. 2000;42(1):17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1%3c17::AID-BIMJ17%3e3.0.CO;2-U.

11. Neubert K, Brunner E. A studentized permutation test for the non-parametric Behrens-Fisher problem. Comput Stat Data Anal. 2007;51(10):5192–204. https://doi.org/10.1016/j.csda.2006.05.024.

12. R Core Team. R: A Language and environment for statistical computing. 2022. https://www.R-project.org/. Accessed 1 Apr 2022.

13. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. BMJ. 2006;333(7581):1267–9. https://doi.org/10.1136/bmj.39036.666389.94.

14. Edwards IR, Lindquist M, Wiholm BE, Napke E. Quality criteria for early signals of possible adverse drug reactions. Lancet. 1990;336(8708):156–8. https://doi.org/10.1016/0140-6736(90)91669-2.

15. Onakpoya IJ, Heneghan CJ, Aronson JK. Delays in the post-marketing withdrawal of drugs to which deaths have been attributed: a systematic investigation and analysis. BMC Med. 2015;13:26. https://doi.org/10.1186/s12916-014-0262-7.

16. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module IX – Signal management (Rev 1). 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-signal-management-rev-1_en.pdf. Accessed 1 Jan 2023.

17. Oh IS, Baek YH, Kim HJ, Lee M, Shin JY. Differential completeness of spontaneous adverse event reports among hospitals/clinics, pharmacies, consumers, and pharmaceutical companies in South Korea. PLoS ONE. 2019;14(2): e0212336. https://doi.org/10.1371/journal.pone.0212336.

18. Durrieu G, Jacquot J, Mège M, Bondon-Guitton E, Rousseau V, Montastruc F, et al. Completeness of spontaneous adverse drug reaction reports sent by general practitioners to a regional pharmacovigilance centre: a descriptive study. Drug Saf. 2016;39(12):1189–95. https://doi.org/10.1007/s40264-016-0463-4.

19. 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(8):617–28. https://doi.org/10.1007/s40264-014-0204-5.

20. Caster O, Sandberg L, Bergvall T, Watson S, Norén GN. vigiRank for statistical signal detection in pharmacovigilance: First results from prospective real-world use. Pharmacoepidemiol Drug Saf. 2017;26(8):1006–10. https://doi.org/10.1002/pds.4247.

21. Lee I, Jokinen JD, Crawford SY, Calip GS, Kilpatrick RD, Lee TA. Exploring completeness of adverse event reports as a tool for signal detection in pharmacovigilance. Ther Innov Regul Sci. 2021;55(1):142–51. https://doi.org/10.1007/s43441-020-00199-z.

22. Lindquist M. VigiBase, the WHO Global ICSR Database System: Basic Facts. Drug Inf J. 2008;42(5):409–19. https://doi.org/10.1177/009286150804200501.

23. Kiguba R, Olsson S, Waitt C. Pharmacovigilance in low- and middle-income countries: A review with particular focus on Africa. Br J Clin Pharmacol. 2023;89(2):491–509. https://doi.org/10.1111/bcp.15193.

24. Barry A, Olsson S, Minzi O, Bienvenu E, Makonnen E, Kamuhabwa A, et al. Comparative assessment of the National Pharmacovigilance Systems in East Africa: Ethiopia, Kenya. Rwanda and Tanzania Drug Saf. 2020;43(4):339–50. https://doi.org/10.1007/s40264-019-00898-z.

25. Niu R, Xiang Y, Wu T, Zhang Z, Chen Y, Feng B. The quality of spontaneous adverse drug reaction reports from the pharmacovigilance centre in western China. Expert Opin Drug Saf.

2019;18(1):51–8. https://doi.org/10.1080/14740338.2019.15598 12.

26. Plessis L, Gómez A, García N, Cereza G, Figueras A. Lack of essential information in spontaneous reports of adverse drug reactions in Catalonia—a restraint to the potentiality for signal detection. Eur J Clin Pharmacol. 2017;73(6):751–8. https://doi.org/10.1007/s00228-017-2223-5.

27. Crépin S, Villeneuve C, Merle L. Quality of serious adverse events reporting to academic sponsors of clinical trials: far from optimal. Pharmacoepidemiol Drug Saf. 2016;25(6):719–24. https://doi.org/10.1002/pds.3982.

28. Nicol C, Moulis F, Bondon-Guitton E, Durrieu G, Montastruc J-L, Bagheri H. Does spontaneous adverse drug reactions' reporting differ between different reporters? A study in Toulouse Pharmacovigilance Centre. Therapies. 2019;74(5):521–5. https://doi.org/10.1016/j.therap.2019.01.008.

29. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337: a1655. https://doi.org/10.1136/bmj.a1655.

30. Pacurariu AC, Coloma PM, van Haren A, Genov G, Sturkenboom MCJM, Straus SMJM. A description of signals during the first 18 months of the ema pharmacovigilance risk assessment committee. Drug Saf. 2014;37(12):1059–66. https://doi.org/10.1007/s40264-014-0240-1.

31. Lester J, Neyarapally GA, Lipowski E, Graham CF, Hall M, Dal Pan G. Evaluation of FDA safety-related drug label changes in 2010. Pharmacoepidemiol Drug Saf. 2013;22(3):302–5. https://doi.org/10.1002/pds.3395.

32. Pinnow E, Amr S, Bentzen SM, Brajovic S, Hungerford L, St George DM, et al. Postmarket safety outcomes for new molecular entity (NME) drugs approved by the Food and Drug Administration between 2002 and 2014. Clin Pharmacol Ther. 2018;104(2):390–400. https://doi.org/10.1002/cpt.944.

33. Pacurariu AC, Hoeve CE, Arlett P, Genov G, Slattery J, Sturkenboom M, et al. Is patient exposure preapproval and postapproval a determinant of the timing and frequency of occurrence of safety issues? Pharmacoepidemiol Drug Saf. 2018;27(2):168–73. https://doi.org/10.1002/pds.4359.

34. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. Postapproval safety data management: definitions and standards for expedited reporting E2d. 2003. https://database.ich.org/sites/default/files/E2D_Guideline.pdf. Accessed 1 Jan 2023.