



Statistical Signal Detection as a Routine Pharmacovigilance Practice: Effects of Periodicity and Resignalling Criteria on Quality and Workload

Magnus Lerch¹ · Peter Nowicki¹ · Katrin Manlik² · Gabriela Wirsching²

Published online: 21 September 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Introduction The goal of signal detection in pharmacovigilance (PV) is to detect unknown causal associations between medicines and unexpected events. Statistical methods serve to detect signals and supplement traditional PV methods. Statistical signal detection (SSD) requires decisions about various settings that influence the quality and efficiency of SSD, as shown in several studies. To our knowledge, the effects of SSD periodicity and resignalling criteria on the quality and workload of routine SSD have not been published before.

Objective To analyse the effects of different periodicities and resignalling criteria on signal detection quality and signal validation workload, and to test the impact of changing the signal threshold for number of cases.

Methods We calculated signals of disproportionate reporting (SDRs) using thresholds of number of cases ($N \geq 3$), proportional reporting ratio ≥ 2 and $\text{Chi}^2 \geq 4$. We retrospectively simulated recurrent SDR calculation and validation with varying periodicity (quarterly vs. monthly), resignalling criteria, and $N \geq 3$ vs. $N \geq 5$.

Results Changing the periodicity from quarterly to monthly increased the workload by 46.6 % (0 % signal loss). More restrictive resignalling criteria reduced the workload between 36.3 % (0 % signal loss) and 74.1 % (50 % signal loss). For $N \geq 3$, the most efficient monthly SSD resignalling criterion reduced the workload by 36.3 % and detected all true signals earlier than quarterly SSD.

$N \geq 5$ reduced the workload between 13.8 and 21.4 % (0 % signal loss).

Conclusions In real-life PV practice, signal detection and validation are recurrent periodic activities. Some true signals are only discovered upon resignalling. Our results demonstrate resignalling criteria with high signal detection quality and high efficiency. We found potential earlier detection of true signals using monthly SSD. Additional studies about resignalling should be performed to complement our findings.

Key Points

Resignalling plays an important role in real-life signal detection and validation practice.

Some true signals are only identified upon resignalling.

Resignalling criteria should be selected carefully to balance early detection of true signals with limiting the workload for the validation of false-positive signals.

✉ Magnus Lerch
magnus.lerch@lenolution.com

¹ Lenolution GmbH, Berlin, Germany

² Global Pharmacovigilance, Bayer Pharma AG, Berlin, Germany

1 Introduction

1.1 Statistical Signal Detection

In the context of pharmacovigilance (PV), a signal has been defined as

“Information that arises from one or multiple sources (including observations and experiments), which suggests a new potentially causal association, or a new aspect of a known association, between an intervention and an event or set of related events, either adverse or beneficial, that is judged to be of sufficient likelihood to justify vericatory action” [1].

Following this definition, the goal of signal detection is to identify new potentially causal associations, or new aspects of known associations. Traditional methods of signal detection include the review of scientific literature, individual case safety reports (ICSRs, cases) including “index cases”, the review of case series, Designated Medical Events and Targeted Medical Events, and the reviews performed during the creation of periodic aggregate reports such as Periodic Safety Update Reports, Periodic Benefit–Risk Evaluation Reports, Development Safety Update Reports, Annual Safety Reports, Periodic Adverse Drug Experience Reports and Investigational New Drug Safety Reports [1].

To support and enhance the traditional methods, statistical signal detection (SSD), e.g. using disproportionality algorithms (DAs) such as the proportional reporting ratio (PRR), has been added to the PV tool set to allow the computer-supported screening of large safety databases [2]. The goal of SSD is to detect signals earlier than traditional PV methods, and thereby gain time to start signal analysis and risk mitigation actions earlier. Studies have shown that this goal is indeed realistic, at least for a large proportion of signals [3, 4]. Agreement exists about the general role of SSD: it should be used to support, but not to replace, traditional PV methods [1, 5].

It is important to note that the initial result of SSD often is not called “signal” but e.g. “Signal of Disproportionate Reporting” (SDR) which requires the review by a safety expert who puts such an SDR into clinical context and decides if the SDR is a signal or not [6]. The definition of signal validation, although not explicitly mentioning SDR, contains the importance of review of the initial signal detection result:

“Signal validation is the process of evaluating the data supporting the detected signal in order to verify that the available documentation contains sufficient evidence demonstrating the existence of a new potentially causal association or a new aspect of a known association, and therefore justifies further analysis” [5].

This means that SSD results need to be reviewed, and this puts the burden of additional work on organisations using SSD. When an organisation plans to implement SSD as a periodic routine PV practice, it needs to decide on a

range of possible settings, e.g. the safety databases to be screened, the periodicity of SSD calculations (“signal runs”), one or several signal detection algorithms, and signal thresholds. All of these parameters influence the quality and efficiency of SSD. Quality is considered the capability of SSD to detect true signals, or to detect true signals earlier than traditional PV methods, respectively. Efficiency can be viewed as the periodic or cumulative workload put onto an organisation to perform the timely review, or signal validation [5], of all SDRs generated by the SSD system in relation to the ultimate goal of SSD: earlier detected true signals. This workload can be measured as the total number of SDRs to be reviewed, or the total number of SDRs per true signal detected early [3, 4].

In short, the real-life practice of signal detection and management has to manage two potentially conflicting goals: (1) quality: identify all true signals as early as possible, and (2) workload: let safety experts focus their precious time on the evaluation of true signals by reducing the number of false-positive signals and thereby the work needed to review them.

Several studies have compared safety databases, signal detection algorithms and thresholds regarding quality indicators such as sensitivity, specificity, and time to detect true signals [7–13], while others explicitly focused on the interrelated effects on quality and workload for variations of signal thresholds and demonstrated ways to decrease the workload efficiently while limiting the risk of missing true signals [3, 4, 14].

1.2 Resignalling

The methods used in many studies (e.g. [3, 4, 14]) analyse SSD as if an SDR for a specific product-event combination (PEC) appears only once when it reaches the defined signal threshold the first time, and at this point in time requires work for signal validation, but the same PEC does never appear as an SDR again, hence, each SDR-PEC only requires signal validation once.

Although this might be enough for SDRs that are validated as being “signals” upon first appearance, many true signals might be validated as “no signal” upon first appearance, but safety information coming in later might cross the “tipping point” where prior “no signals” turn into “signals” requiring vericatory actions. Signal validation always is a point-in-time decision that might change over time, and especially SDRs validated as “no signals” should be reevaluated, once new relevant safety information becomes available. Hence, detecting new aspects of prior “no signals” and of known causal associations, e.g. increased frequency, increased severity, greater specificity for selected patient populations or “striking” information

about causality in new ICSRs, requires methods to bring SDRs to the attention of safety experts again. One effective method is to allow SDRs to reappear in the SSD system. We call this reappearance “resignalling”.

To our knowledge, there are no published recommendations yet for the design and use of, what we call, “resignalling criteria”, i.e. conditions that need to be fulfilled for a prior generated SDR to reappear. However, we have found a few mentions of such criteria [1, 15] and have three main options to approach resignalling:

- Full resignalling:
For each SSD run, all PECs that reach the signal threshold are generated as SDRs, i.e. there is no check if they have been generated (and validated) before in prior SSD runs.
- Selective resignalling:
For each SSD run, all PECs that reach the signal threshold are checked against prior SSD runs. If they generated an SDR before, it is checked if they fulfill predefined resignalling criteria (see Sect. 2.3). Only if they reach the signal threshold plus fulfill the resignalling criteria, are they generated again as SDRs requiring validation.
Different options exist for the “baseline” the resignalling criteria are tested against. The baseline can be:
 - The properties (counts, frequencies, statistical scores) an SDR had a predefined period ago, e.g. “in the previous week” [1], “26 weeks ago” or “52 weeks ago” [15];
 - The properties an SDR had when it was validated by a safety expert the last time.
 The rationale for the latter option is: At this prior time, a safety expert reviewed the SDR including its scores, the available ICSRs and further safety information, e.g. scientific literature, and made a point-in-time validation decision (“signal” or “no signal”). Then, only relevant changes compared to this baseline, e.g. new ICSRs, an increased PEC frequency, need to trigger the reappearance of the same PEC for expert review.
- No resignalling:
A PEC is only generated as an SDR once and is never generated again thereafter.

1.3 The Need for Evidence About Expected Quality and Workload

Following the “Guideline on good pharmacovigilance practices (GVP): Module IX Signal management” (GVP Module IX), which recommends signal detection to be performed at least monthly [5], we considered what the

most suitable periodicity of SSD could be, while traditional signal detection methods were already performed as ongoing activities. As we were not aware of published information about the effects of SSD periodicity on signal detection quality and signal validation workload, we decided to perform our own study and to compare two SSD periodicities, quarterly and monthly. The study needed to provide evidence about quality and workload, and specifically answer the following questions: If, compared to quarterly periodicity, SSD is performed monthly,

- Would any true signal be missed?
- To what extent would the signal validation workload increase?
- Are there options to maintain or decrease the workload, while still detecting all true signals?

2 Methods

2.1 Data, Algorithms, Thresholds, and Periodicity

As data, we used spontaneous (including literature) reports extracted from the company safety database, the characteristics of which are published in Wisniewski et al. [15]. SDRs were calculated using the PRR in combination with Chi^2 (with Yates correction) with a signal threshold of number of cases (N) ≥ 3 , $\text{PRR} \geq 2$ and $\text{Chi}^2 \geq 4$, as described by Evans et al. [2]. In addition, we calculated the 95 % confidence interval for PRR as explained in the “Guideline on the use of statistical signal detection methods in the EudraVigilance data analysis system” [6].

Counts and statistics were calculated for suspect products for each PEC on the level of medicinal product name and MedDRA^{®1} Preferred Term. Both serious and non-serious events were included. For some products, lists of PECs, contained in the company core safety information and/or for which other signal detection and monitoring methods were implemented, were used to filter out corresponding SDRs. The same filter was applied to both initial and resignalled SDRs from all signal runs in this study to avoid any time-related bias.

In our study, we compared quarterly vs. monthly SSD. Inspired by Slattery et al. [4], we also evaluated the change

¹ MedDRA[®] (the Medical Dictionary for Regulatory Activities) terminology is the international medical terminology developed under the auspices of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). The MedDRA[®] trademark is owned by the International Federation of Pharmaceutical Manufacturers and Associations on behalf of ICH.

in the signal threshold for number of cases from $N \geq 3$ to $N \geq 5$.

2.2 Reference Set of True Signals

To evaluate the signal detection quality, we used a reference set of true signals. For this study, we defined “true signal” as a PEC for which an SDR had been generated in the productive SSD system, and which had undergone signal validation by a safety expert who assessed this SDR to be a signal requiring verificatory action. We included only those SDRs that were validated as signals the first time within the year preceding the study, i.e. the same 1-year period we used in our simulation. The reference set contained eight true signals, including four that appeared as new SDRs the first time and four that had appeared before and were validated as “no signal” at first, but appeared again upon resignalling and were validated as “signal” then.

2.3 Resignalling Criteria

As mentioned earlier, we are not aware of any published recommendation for the design and use of resignalling criteria. However, we found examples of how other organisations combined signal thresholds with time-dependent criteria:

- “at least 3 reports of the [PEC] with 1 report received in the previous week” [1].
- “EB05 ≥ 1.8 , and/or [positive] trend flag. A trend flag is [positive] if either of the following are true:
 - “An EB05 based on current data is $>EB95$ for the [PEC] 52 weeks ago”;
 - “A 50 % increase in EBGm score when current data are compared with the EBGm score 26 weeks ago” [15].

Recently, Candore et al. also reported on these criteria [13].

- “EB05 > 2 for non-serious unlisted adverse events; any event whose reporting rate has increased significantly compared to 6 months previously” [15].

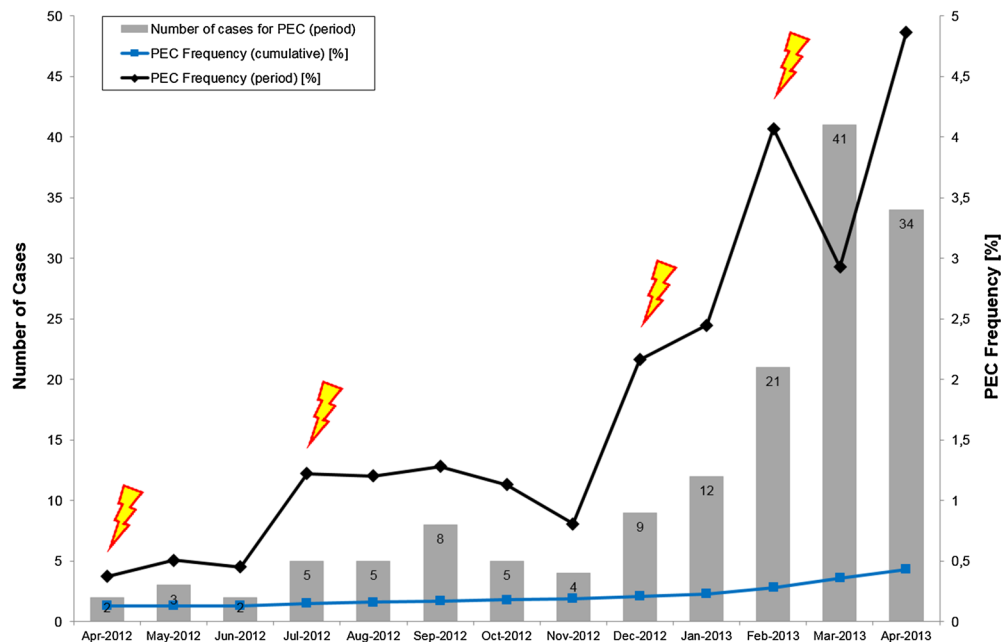
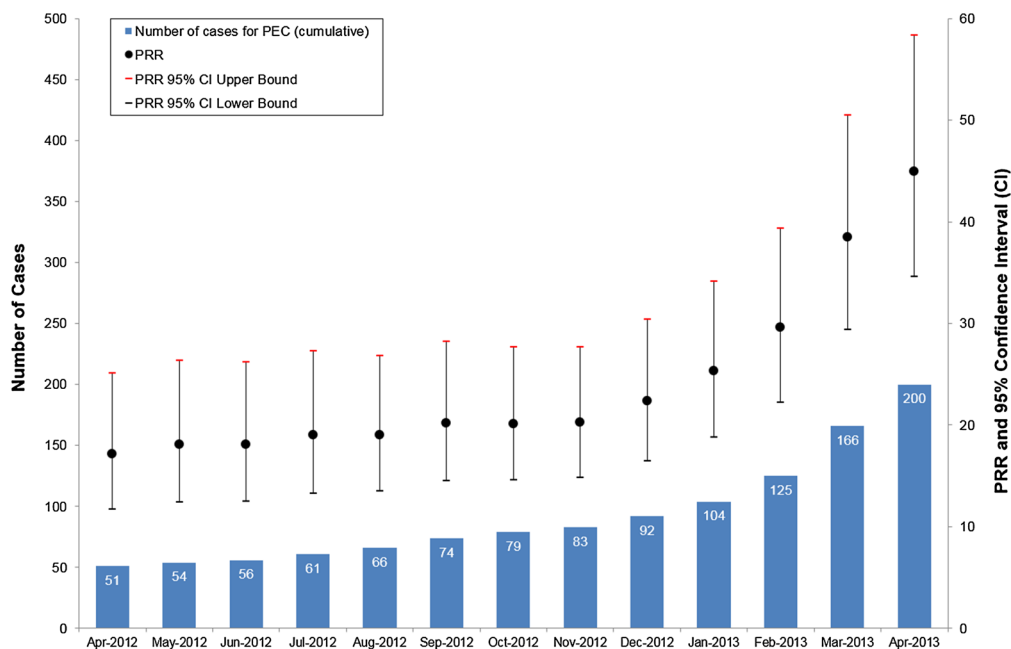
(EBGM: Empirical Bayes Geometric Mean; EB05, EB95: lower and upper bound of the 90 % confidence interval for EBGm)

Building on these ideas, our previous experience with SSD and resignalling, and using common sense about what type of relevant new information could change the signal validation decision and therefore should be brought to the attention of the safety expert again, we designed a range of different resignalling criteria for this study as follows:

If an SDR had been validated before (as “signal” or “no signal”), a new SDR for the same PEC is generated, if the PEC reaches the signal threshold and if:

- $N_{\text{new}} \geq N_{\text{validated}} + 1$ (“+1 new case”):
At least one more case containing the PEC was added to the safety database compared to the number of cases at the time the SDR was validated previously.
- $N_{\text{new}} > N_{\text{validated}} + 50\%$ (“>50 % new cases”):
The current number of cases for the PEC exceeds the number of cases at the time the SDR was validated previously by more than 50 %.
- Frequency (cumulative)_{new} $>$ Frequency (cumulative)_{validated} + 50 % (“>50 % increased cumulative frequency”):
The current cumulative PEC frequency (= cumulative number of cases for PEC/cumulative number of cases for product) exceeds the cumulative PEC frequency at the time the SDR was validated previously by more than 50 %.
- Frequency (period)_{new} $>$ Frequency (period)_{validated} + 50 % (“>50 % increased period frequency”):
The current period-based PEC frequency (= number of cases for PEC in the current SSD period/number of cases for product in the current SSD period) exceeds the period-based PEC frequency at the time the SDR was validated previously by more than 50 %.
- $PRR_{\text{new}} > PRR_{\text{validated}} + 50\%$ (“>50% increased PRR”):
The current PRR value exceeds the PRR value at the time the SDR was validated previously by more than 50 %.
- $PRR025_{\text{new}} > PRR975_{\text{validated}}$ (“PRR CI shift”):
The current lower bound of the 95 % PRR confidence interval (CI) (PRR025) exceeds the upper bound of the 95 % PRR CI (PRR975) at the time the SDR was validated previously.
- No resignalling:
For a given PEC, an SDR is only generated when it reaches the signal threshold the first time and does never reappear as SDR thereafter.

Figure 1 illustrates, for one PEC, the monthly counts, frequencies and PRR scores used for the different resignalling criteria from the initial appearance as SDR in the baseline SSD run over the course of the 1-year study period. The figure also shows for each resignalling criterion in which month(s) the PEC reappears as SDR. As an example, the flashes in Fig. 1 indicate when the PEC is detected as SDR the first time and reappears again as SDR for the resignalling criterion “>50 % increased period frequency”.



Resignalling criteria	Apr-2012	May-2012	Jun-2012	Jul-2012	Aug-2012	Sep-2012	Oct-2012	Nov-2012	Dec-2012	Jan-2013	Feb-2013	Mar-2013	Apr-2013
First signal	x												
Resignalling: + 1 new case	x	x	x	x	x	x	x	x	x	x	x	x	x
Resignalling: > 50 % increased period frequency	x			x					x		x		
Resignalling: > 50 % new cases	x						x				x		x
Resignalling: > 50 % increased cumulative frequency	x								x			x	
Resignalling: > 50 % increased PRR	x										x		x
Resignalling: $PRR_{025_{new}} > PRR_{975_{validated}}$	x											x	
No resignalling	x												

Fig. 1 Top and middle case counts, frequencies and proportional reporting ratio (PRR) with 95 % confidence interval (CI) for a selected product-event combination (PEC) from the baseline statistical signal detection (SSD) run over the course of the study period. Middle as example, the flashes indicate when the PEC is detected as signal of

disproportionate reporting (SDR) the first time and reappears again as SDR for the resignalling criterion “>50 % increased period frequency” (signal threshold: $N \geq 3$, $PRR \geq 2$ and $\chi^2 \geq 4$). Bottom an x in the table denotes in which month(s) the PEC is detected as SDR for each resignalling criterion

2.4 Outcome Measures

With this study, we aimed to gather estimates about SSD quality and workload.

To assess quality, we used three outcome measures:

- (a) Cumulative total number of true signals detected:
The total number per resigalling criterion includes both the signals identified by the respective resigalling criterion and the new signals that reached the signal threshold the first time during the study period.
- (b) Percentage of change of (a) named “signal loss” compared with the reference set of SDRs validated as signals².
- (c) Time of first detection compared with the reference set of SDRs validated as signals and expressed as “earlier”, “same time” or “later”.

Others have used alternative quality measures: Alvarez et al. and Slattery et al. computed the time gain of SSD for detecting signals, which led to updates of the European Union Summary of Product Characteristics compared with detection by traditional PV methods [3, 4]; Hochberg et al. compared SSD results against a “highly inclusive reference event database” and calculated the number of SDRs needed to detect new true signals [12]; Candore et al. calculated the sensitivity and precision of SSD against a reference set of PECs listed in the Summary of Product Characteristics and company core safety information [13]. Being aware of those alternatives, we decided to use a reference set of true signals that was created prospectively during the real-life process of validating SDRs generated by the SSD system. The advantage: We could be sure that these signals were based on SSD and knew exactly when they were first validated as signals (“index dates”), i.e. we did not need to retrospectively identify the signal detection sources (SSD vs. non-SSD) and index dates for the PECs in our reference set, an approach that could have been prone to bias owing to the retrospective approach. With this prospective SSD-based reference set, we had the ideal benchmark for the, now retrospective, SSD simulation, which covered the same time period the true signals were identified. As the number of true signals in the reference set was small (eight PECs), we chose to use simple measures rather than measures better suited for larger sample sizes, such as sensitivity, specificity or precision.

To assess workload, we used two outcome measures:

- (a) Cumulative total number of SDRs:
The total number per resigalling criterion includes both the resigalling SDRs identified by the respective resigalling criterion and the new SDRs that reached the signal threshold the first time during the study period.
- (b) Percentage of change of (a) compared with the total number of SDRs for the initial SSD settings (quarterly SSD runs, signal threshold for number of cases ($N \geq 3$, resigalling criterion “+1 new case”).

2.5 Simulation

We were interested how changes in the settings of the SSD system affect the quality and workload of SSD. As we wanted to get answers quickly, it would not have been useful to apply the changes to the productive SSD system because it would have taken a long time to get results for the various possible combinations of settings. Furthermore, we wanted to compare the SSD quality with a reference set of SDRs that had already been assessed as being signals during signal validation in real life and we needed to perform SSD runs for the same time period the validations had been done.

To suit our needs, we decided to perform a simulation: In June 2013, we simulated the recurrent SDR calculation and signal validation for a retrospective 1-year period while varying the settings for periodicity, resigalling criteria and signal threshold for number of cases. The simulation involved the following steps:

- SDRs were calculated for an initial quarterly (Q2 2012) and an initial monthly (April 2012) SSD run. These runs were used to simulate the cumulative history of SSD runs up to this point where all SDRs generated in this “seed run” would have been reviewed by safety experts during signal validation. All subsequent SSD runs could then use this signal validation history and only bring up new SDRs for PECs that had not been identified as SDRs before plus SDRs for PECs, which had been generated and validated as SDRs before, but reached the signal threshold again and fulfilled the respective resigalling criteria.
- After that, SDRs were calculated for a period of 12 months for quarterly (Q3 2012 to Q2 2013, i.e. four quarterly runs) and monthly (May 2012 to April 2013, i.e. 12 monthly runs) SSD periodicity. New SDRs and resigalling SDRs resulting from each of these runs were assumed to have been validated before the next run, hence, their counts, frequencies and statistical scores created a new baseline for the subsequent SSD runs to compare the resigalling criteria against.

² Also possible, but beyond the scope of this study, would be “signal gains”, i.e. SDRs identified and validated as signals in addition to the SDRs which actually had gone through signal validation, had been validated as signals and thereby made it into the reference set we used in this study.

- The SDRs resulting from all SSD runs were summed up to get the cumulative total number of SDRs, which was used as our measure for workload.
- Regarding quality, the SDRs resulting from any of the SSD runs were checked against our reference set of true signals.

In the simulation described above, we computed and analysed the number of true signals detected. As we also wanted to compare the times of first detection of these true signals, we repeated the simulation using data for the same study period but from a database snapshot that had been updated since the first simulation. Using the data from this new simulation, the index date for each true signal was defined as the time of first detection with the initial SSD system settings, i.e. quarterly SSD runs, signal threshold $N \geq 3$, $PRR \geq 2$ and $\text{Chi}^2 \geq 4$, and resignalling criterion “+1 new case”. Then, the time of first detection for each true signal and for each SSD setting was calculated, compared against the index date, and the values of interest assigned, i.e. “earlier”, “same time” or “later”, respectively.

3 Results

3.1 Monthly vs. Quarterly Periodicity

Our initial question was: If we leave the other SSD settings the same (i.e. signal threshold: $N \geq 3$, $PRR \geq 2$ and $\text{Chi}^2 \geq 4$; resignalling criterion: “+1 new case”), how would the workload for signal validation increase if we move from quarterly to monthly SSD runs?

Figure 2 shows that the number of SDRs per monthly SSD run is well below the ones for quarterly SSD runs.

However, looking at the “running total”, i.e. the accumulating number of SDRs from month to month, or quarter to quarter, respectively, we immediately see a workload increase of monthly compared with quarterly SSD (see Fig. 3).

Looking at the numbers (see Table 1), changing the periodicity from quarterly to monthly SSD, while keeping the resignalling criterion “+1 new case”, increased the workload (cumulative total number of SDRs) by 46.6 % with 0 % signal loss.

3.2 Comparing Resignalling Criteria

Comparing the effects of all resignalling criteria while keeping the signal threshold for number of cases $N \geq 3$, we got the following results:

- For quarterly SSD runs, the workload decrease ranges from 49.1 % (12.5 % signal loss) to 75.0 % (50.0 % signal loss).
- For monthly SSD runs, the workload change ranges from a 46.6 % increase (0.0 % signal loss) to a 74.1 % decrease (50.0 % signal loss).
- For $N \geq 3$, the most efficient resignalling criterion with 0.0 % signal loss is “>50 % increased period frequency” using monthly SSD runs with a workload decrease of 36.3 %.

As a visual summary, Fig. 4 shows the huge variations in workload for the different resignalling criteria. Compared with the baseline settings of the SSD system (6085 SDRs), when switching from quarterly to monthly SSD runs, the workload change ranges from 8923 SDRs (+46.6 %) to 1578 SDRs (−74.1 %). Considering also our prerequisite of not missing any true signal, we see the least workload for resignalling criterion “>50 % period frequency increase” (3878 SDRs, −36.3 %).

3.3 Threshold $N \geq 5$ vs. $N \geq 3$

Applying a signal threshold of $N \geq 5$, results were as follows:

- For quarterly SSD runs, the workload decrease ranges from 13.8 % (0.0 % signal loss) to 83.3 % (37.5 % signal loss).
- For monthly SSD runs, the workload change ranges from a 32.0 % increase (0.0 % signal loss) to a 82.8 % decrease (37.5 % signal loss).
- For $N \geq 5$, the most efficient resignalling criterion with 0.0 % signal loss is, just like for $N \geq 3$, “>50 % increased period frequency” using monthly SSD runs with a workload decrease of 49.9 %.

If we compare $N \geq 5$ with $N \geq 3$ for the initial SSD settings (quarterly SSD runs, “+1 new case”), $N \geq 5$ decreased the workload by 13.8 % with 0.0 % signal loss.

For the most efficient resignalling criterion for $N \geq 3$ (monthly SSD runs, 3878 SDRs, 36.3 % workload decrease, 0.0 % signal loss), a switch to $N \geq 5$ (3047 SDRs) would additionally decrease the workload by 21.4 % with 0.0 % signal loss.

3.4 Completeness and Timeliness of Detecting True Signals

Table 2 presents the results of our second simulation run where we calculated and compared the dates of the first detection of true signals (see Sect. 2.5).

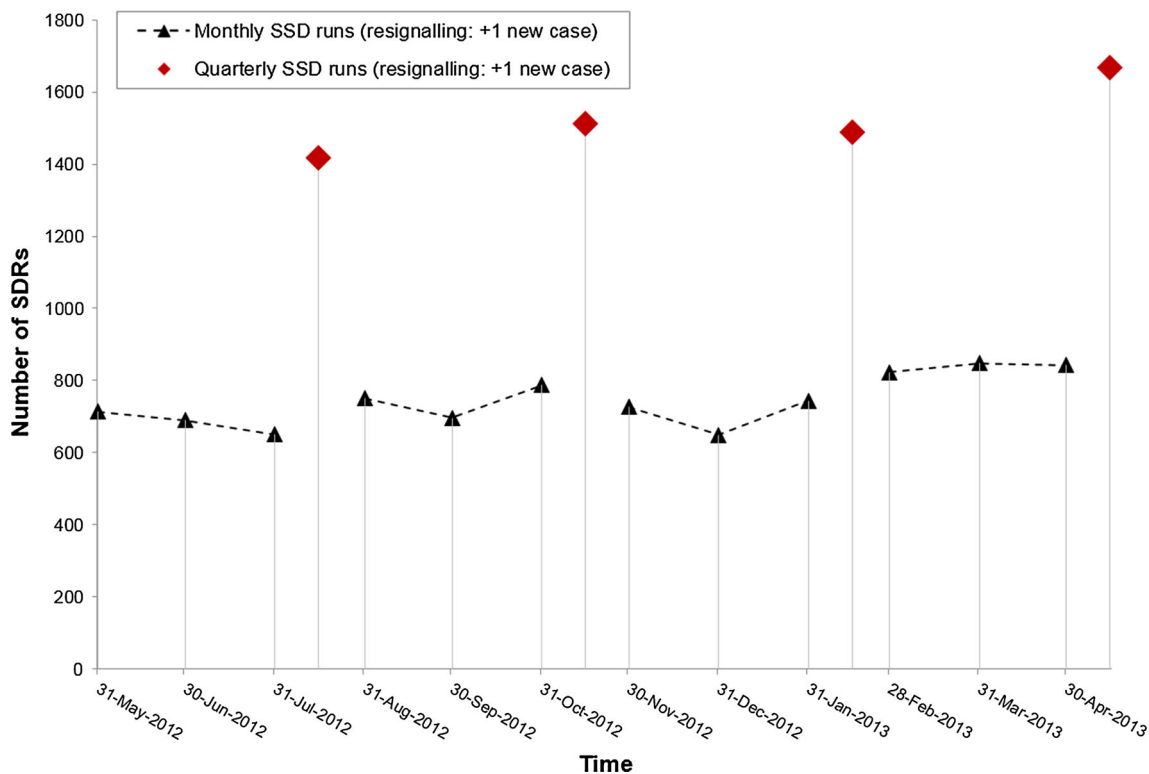


Fig. 2 Number of signals of disproportionate reporting (SDRs) for quarterly vs. monthly statistical signal detection (SSD) runs using proportional reporting ratio (PRR) and resignalling criterion “+1 new case” per SSD run over the 1-year study period (signal threshold:

$N \geq 3$, $PRR \geq 2$, $\text{Chi}^2 \geq 4$). Note: The display of the quarterly SSD results has been shifted by 15 days into the following month for visualisation purposes

Comparing monthly with quarterly SSD runs for each resignalling criterion, we see a consistent pattern of earlier detection of true signals with monthly SSD.

Independent of settings for periodicity (quarterly vs. monthly) and signal threshold ($N \geq 3$ vs. $N \geq 5$), the resignalling criteria seem to form a “quality hierarchy” when we sort them from low to high signal loss (0.0–57.1 %), while, from top to bottom, before losing signals, they first tend to detect signals later:

1. “+1 new case”
2. “>50 % increased period frequency”
3. “>50 % new cases”
4. “>50 % increased cumulative frequency” same quality as “>50 % increased PRR”
5. “ $\text{PRR}_{025_{\text{new}}} > \text{PRR}_{975_{\text{validated}}}$ ” same quality as “Only new SDRs, no resignalling”

“+1 new case” is the only criterion with 0.0 % signal loss for monthly and quarterly SSD runs with both $N \geq 3$ and $N \geq 5$.

The two best overall SSD settings regarding 0.0 % signal loss and earlier detection of all true signals are monthly SSD runs with signal threshold $N \geq 3$ and with

either “+1 new case” or “>50 % increased period frequency”.

4 Discussion

4.1 Monthly vs. Quarterly Periodicity

Our main question was: how much more SDRs have to be reviewed during signal validation when SSD is performed monthly compared with quarterly. If a SSD system would, for any PEC, only generate an SDR the very first time it reaches the signal threshold, and never again afterwards, then the change of periodicity would, in general, not increase the total number of SDRs.

An exception, we found, are SDRs that are only visible in monthly SSD runs and are not detected in quarterly runs. The likely cause for this phenomenon is this: In our study, the PRR is based on the cumulative counts of ICSRs up to specific “as of” dates (end of the month, end of the quarter). Using these counts, the “PRR is an estimate of the probability that a spontaneous report containing a product (X) will mention an adverse event (Y) divided by the

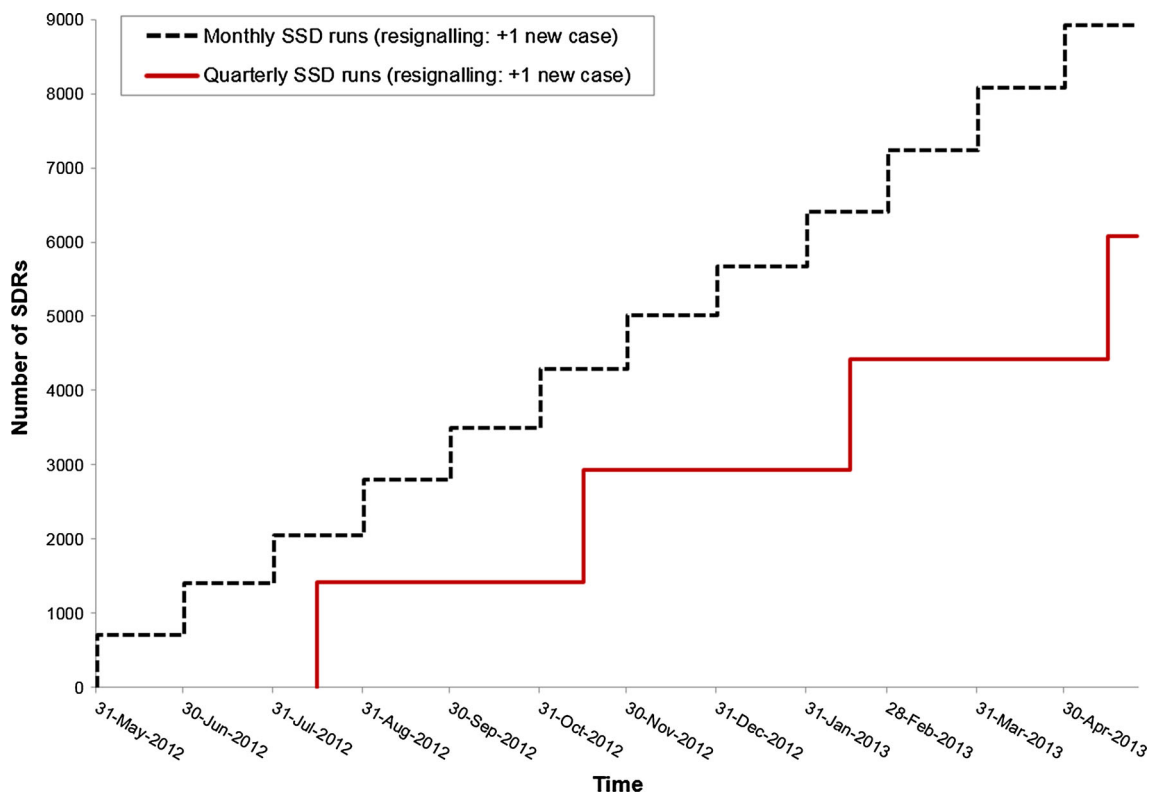


Fig. 3 Running total of number of signals of disproportionate reporting (SDRs) for quarterly vs. monthly statistical signal detection (SSD) runs using proportional reporting ratio (PRR) and resignalling criterion “+1 new case” over the 1-year study period (signal

threshold: $N \geq 3$, $PRR \geq 2$, $\text{Chi}^2 \geq 4$). Note: The display of the quarterly SSD results has been shifted by 15 days into the following month for visualisation purposes

probability that a report not containing X will mention Y ” [4]. However, the content of a safety database is dynamic, and ICSRs are added and updated all the time. Hence, the ICSR counts by product, by adverse event and in total change over time, and, therefore, the PRR changes dynamically, too. The same is true for the Chi^2 . As a consequence, PRR and Chi^2 can reach the signal threshold in a monthly SSD run, but may not reach the threshold in a quarterly SSD run. These exceptions account for only a small proportion of the total number of SDRs: We found 56 additional SDRs for the signal threshold $N \geq 3$, and 34 for $N \geq 5$. These additional SDRs are contained in the monthly runs for all resignalling criteria, so account for a portion of the difference between monthly and quarterly SSD runs. To analyse any potential “signal gain”, i.e. additional SDRs validated as “signals”, among such SDRs could be subject of a future study.

Interestingly, although the reference set of true signals contained eight PECs, which all were found in the first simulation (see Table 1), now, only seven were identified with the baseline SSD settings in the second simulation (see Table 2), i.e. one true signal was lost. The

reason: One case containing the PEC of this true signal had been identified as a duplicate after the first simulation, and was excluded in the second simulation. With one case less, this PEC did not reach the signal threshold in any of the quarterly SSD runs, including the baseline SSD settings, and, hence, was excluded from the results in Table 2.

However, this true signal was still found as a new SDR with monthly SSD runs in the second simulation, specifically, earlier than it was identified with quarterly SSD runs in the data used in the first simulation. If we would want to acknowledge this fact in Table 2, the number of signals found “Earlier” for all resignalling criteria for monthly SSD runs would increase by one, thus, further amplifying the effect of earlier detection with monthly compared with quarterly SSD.

Despite the difference of 7 vs. 8 signals, and considering the known root cause for this observation, our second simulation (Table 2) verifies the results of our first simulation (Table 1), as the difference in “# signals found” is consistently one signal less in Table 2 for each tested SSD setting.

Table 1 Cumulative total number of SDRs generated in the study period, percentage of difference compared with the baseline, number of true signals found and percentage of signal loss compared with the

baseline. SDRs were generated using PRR with a signal threshold $PRR \geq 2$ and $\text{Chi}^2 \geq 4$, using quarterly vs. monthly SSD runs, signal threshold $N \geq 3$ vs. $N \geq 5$, and different resignalling criteria

Resignalling criteria	Quarterly SSD runs				Monthly SSD runs			
	# SDRs	Difference (%)	# signals found	Signal loss (%)	# SDRs	Difference (%)	# signals found	Signal loss (%)
$N \geq 3$								
+1 new case ^a	6085		8		8923	+46.6	8	0.0
>50 % increased period frequency	3096	-49.1	7	12.5	3878	-36.3	8	0.0
>50 % new cases	2140	-64.8	6	25.0	2355	-61.3	6	25.0
>50 % increased cumulative frequency	1790	-70.6	5	37.5	1956	-67.9	5	37.5
>50 % increased PRR	1681	-72.4	5	37.5	1795	-70.5	5	37.5
$PRR_{025_{\text{new}}} > PRR_{975_{\text{validated}}}$	1551	-74.5	4	50.0	1611	-73.5	4	50.0
Only new SDRs, no resignalling	1522	-75.0	4	50.0	1578	-74.1	4	50.0
$N \geq 5$								
+1 new case	5245	-13.8	8	0.0	8031	+32.0	8	0.0
>50 % increased period frequency	2341	-61.5	7	12.5	3047	-49.9	8	0.0
>50 % new cases	1438	-76.4	6	25.0	1597	-73.8	6	25.0
>50 % increased cumulative frequency	1179	-80.6	6	25.0	1296	-78.7	6	25.0
>50 % increased PRR	1099	-81.9	6	25.0	1168	-80.8	6	25.0
$PRR_{025_{\text{new}}} > PRR_{975_{\text{validated}}}$	1043	-82.9	5	37.5	1080	-82.3	5	37.5
Only new SDRs, no resignalling	1014	-83.3	5	37.5	1048	-82.8	5	37.5

SDR signal of disproportionate reporting, PRR proportional reporting ratio, SSD statistical signal detection, N number of cases, # total number

^a This setting is used as the baseline for comparison: Quarterly SSD runs, $N \geq 3$, resignalling criterion “+1 new case”

4.2 Effects of Resignalling

Beyond these exceptions, the more important question is, if the SSD system allows SDRs to reappear, what we call “resignalling”. If resignalling is disabled and only first-time new SDRs are generated, then obviously the total number of SDRs and the total workload required for signal validation is much less than in a system with resignalling enabled. In our study, new SDRs accounted for only 1014 (16.7 %; quarterly runs, $N \geq 5$) to 1578 (25.9 %; monthly runs, $N \geq 3$) SDRs compared with the initial system settings (6085 SDRs).

We recommend to enable resignalling in the SSD system and consider it an important feature for detecting new aspects of known associations and for revalidating SDRs in the light of newly available safety information. As evidence of this importance, when we created the reference list of true signals, we found that four of eight SDRs validated as “signal” had initially been assessed as “no signal” and turned from “no signal” to “signal” upon resignalling. That means, without resignalling, 50 % of these true signals would have been missed. That does not

necessarily mean they would have been missed completely, as traditional signal detection methods make up for this gap, detecting many signals even earlier or at about the same time as SSD [3].

Regarding resignalling, the question was not “if”, but “to what extent” it is useful. To answer this question, besides comparing quarterly and monthly SSD periodicity, we simulated the effect of the different resignalling criteria. The simulation provided important answers:

- SSD can be performed with a monthly periodicity to align with recommendations in GVP Module IX [5] with options, compared with quarterly periodicity, to decrease the signal validation workload and, at the same time, improve the signal detection quality.
- Switching from the initially used resignalling criterion “+1 new case” to “>50 % increased period frequency” decreased the workload by 36.3 % and still detected all true signals.
- As an additional benefit, with monthly periodicity and the new resignalling criterion, all true signals were detected as SDRs earlier.

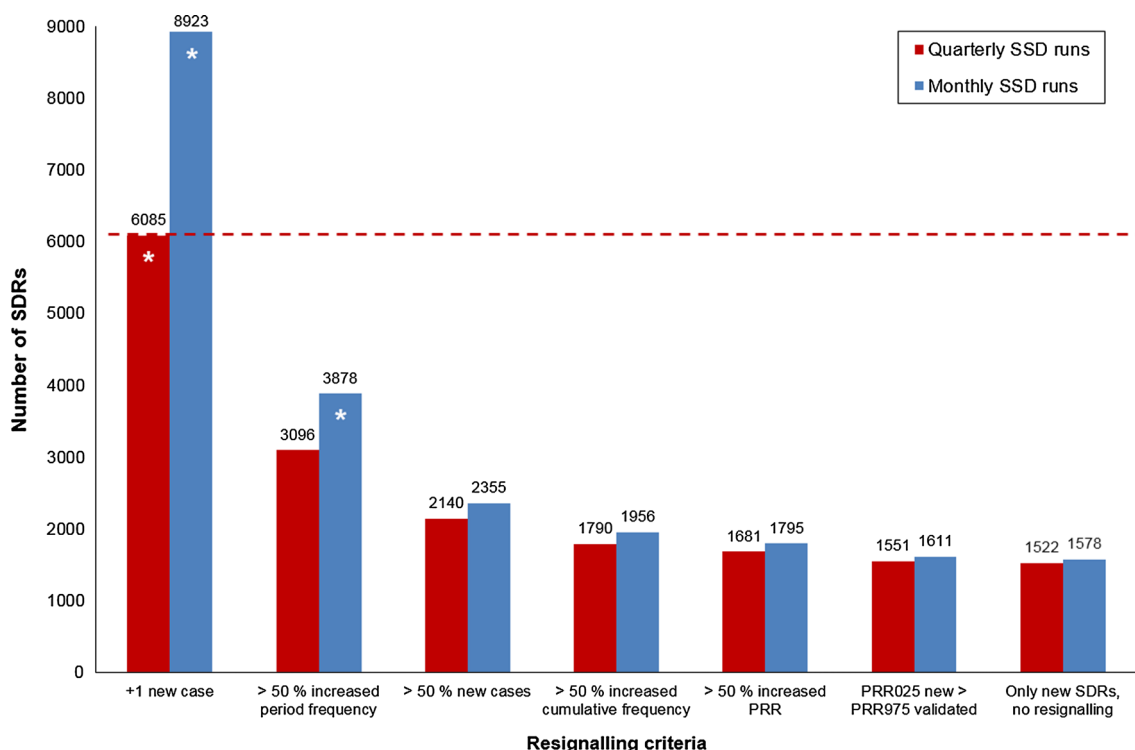


Fig. 4 Cumulative total number of signals of disproportionate reporting (SDRs) detected for quarterly vs. monthly statistical signal detection (SSD) runs using proportional reporting ratio (PRR) and

different resignalling criteria (signal threshold: $N \geq 3$, $PRR \geq 2$, $\chi^2 \geq 4$). An *asterisk* on the bar signifies results where all true signals were detected

4.3 Threshold $N \geq 5$ vs. $N \geq 3$

As one of our objectives, we also wanted to test results by Slattery et al. who also focussed on SSD quality and workload and found, for the EudraVigilance Data Analysis System, that a threshold of $N \geq 5$ compared with $N \geq 3$ “gave a reduction of 25 % in false-positive signals in return for a loss of 12 % in true signals detected early” [4]. Despite differences regarding the safety databases (EudraVigilance vs. company database), signal thresholds (lower bound of the 95 % confidence interval for $PRR > 1$ vs. $PRR \geq 2$ and $\chi^2 \geq 4$), and different signal reference sets and quality measures, our results support their findings:

- A threshold of $N \geq 5$ compared with $N \geq 3$ with quarterly SSD runs and the initial resignalling criterion “+1 new case” reduced the workload by 13.8 % with 0.0 % signal loss.
- For the most efficient resignalling criterion (“>50 % increased period frequency, monthly SSD runs, 3878 SDRs, 0.0 % signal loss), a switch from $N \geq 3$ to $N \geq 5$ (3047 SDRs) additionally reduced the workload by 21.4 % with 0.0 % signal loss, but one true signal (14.3 %) was detected later compared with $N \geq 3$.

4.4 Insights from Simulations

Resignalling criteria should be selected carefully to balance a high signal detection quality (early detection of true signals) with a high efficiency (minimum workload for false-positive signals). Simulations, like the one we described in this paper, are helpful to gather evidence about the effects of different settings before applying any changes to the productive SSD system.

With this study, we obtained evidence that supports performing SSD with a monthly periodicity and using a resignalling criterion that (a) would not compromise on quality, the early detection of true signals, and (b) would decrease the workload for false-positive signals compared with other SSD settings. Furthermore, we found a time gain for earlier detection of true signals when using monthly instead of quarterly SSD, if appropriate resignalling criteria are used.

Finally, numerous factors influence the composition of a safety database and its changes over time, e.g. addition of case reports for new or newly acquired medicinal products including new product classes, or case reports from emerging sources such as social media. Safety database changes, in turn, might affect the quality and efficiency of SSD. Hence, whenever substantial database changes with a

Table 2 Cumulative total number of true signals detected as SDRs in the study period either earlier, at the same time, or later compared with the baseline, total of true signals found and percentage of signal loss compared with the baseline. SDRs were generated using PRR

with signal threshold $PRR \geq 2$ and $\text{Chi}^2 \geq 4$, using quarterly vs. monthly SSD runs, signal threshold $N \geq 3$ vs. $N \geq 5$, and different resignalling criteria

Resignalling criteria	Quarterly SSD runs				Signal loss (%)	Monthly SSD runs					
	# signals found					# signals found	Signal loss (%)				
	Earlier	Same time	Later	Total			Earlier	Same time	Later	Total	
$N \geq 3$											
+1 new case ^a		7		7	0.0	7			7	0.0	
>50 % increased period frequency		6		6	14.3	7			7	0.0	
>50 % new cases		3		2	5	28.6	3		2	5	28.6
>50 % increased cumulative frequency		3		1	4	42.9	3		1	4	42.9
>50 % increased PRR		3		1	4	42.9	3		1	4	42.9
$PRR_{025_{\text{new}}} > PRR_{975_{\text{validated}}}$		3			3	57.1	3			3	57.1
Only new SDRs, no resignalling		3			3	57.1	3			3	57.1
$N \geq 5$											
+1 new case		6		1	7	0.0	6		1	7	0.0
>50 % increased period frequency		5		1	6	14.3	6		1	7	0.0
>50 % new cases		3		2	5	28.6	3		2	5	28.6
>50 % increased cumulative frequency		3		2	5	28.6	3		2	5	28.6
>50 % increased PRR		3		2	5	28.6	3		2	5	28.6
$PRR_{025_{\text{new}}} > PRR_{975_{\text{validated}}}$		3		1	4	42.9	3		1	4	42.9
Only new SDRs, no resignalling		3		1	4	42.9	3		1	4	42.9

SDR signal of disproportionate reporting, PRR proportional reporting ratio, SSD statistical signal detection, N number of cases, # total number

^a This setting is used as the baseline for comparison: Quarterly SSD runs, $N \geq 3$, resignalling criterion “+1 new case”

potential impact on SSD are occurring, reevaluation of the SSD performance is warranted.

5 Conclusions

In real-life routine PV practice, signal detection and validation are recurrent periodic activities within the overall signal management process. Resignalling is an important feature of this process, as some true signals are only discovered upon resignalling. The benefit of enabling resignalling is the detection of true signals which only are validated as signals upon resignalling. Within the study period, we found that 50 % true signals were only identified during resignalling, hence, disabling resignalling would mean a high risk of missing true signals. Based on these findings, we suggest that resignalling should be enabled in any complete SSD system. Enabling resignalling, however, comes with an organisational “cost” of an increased workload for signal validation. Fortunately, we were able to identify resignalling criteria that limit this workload increase while still finding all true signals.

Our reference set of true signals was rather small and specific to our study objective and methods. To derive general recommendations about resignalling and to complement our findings, it would be worthwhile to study the effects of resignalling in other safety databases, with other signal detection algorithms, signal thresholds and bigger signal reference sets.

Acknowledgments We thank Marian Klug for his part in the study and constructive comments. We also thank Roman Haack for his constructive comments. The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of Lenolition GmbH or Bayer Pharma AG.

Compliance with Ethical Standards

Funding This study was conducted during a project at Bayer Pharma AG that involved consultancy work carried out by Magnus Lerch and Peter Nowicki. Magnus Lerch and Peter Nowicki have received compensation for consultancy work from Bayer Pharma AG. No sources of funding were used to assist with the preparation of the manuscript.

Conflicts of interest Magnus Lerch has held shares in Bayer AG in the past. Magnus Lerch, Peter Nowicki, Katrin Manlik and Gabriela

Wirsching have no conflicts of interest that are directly relevant to the content of this study. Katrin Manlik and Gabriela Wirsching are full-time employees of Bayer Pharma AG.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. CIOMS. Practical aspects of signal detection in pharmacovigilance: report of CIOMS Working Group VIII. Geneva: CIOMS; 2010.
2. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.* 2001;10:483–6.
3. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. *Drug Saf.* 2010;33:475–87.
4. Slattery J, Alvarez Y, Hidalgo A. Choosing thresholds for statistical signal detection with the proportional reporting ratio. *Drug Saf.* 2013;36:687–92.
5. European Medicines Agency, Heads of Medicines Agencies. Guideline on good pharmacovigilance practices (GVP): Module IX Signal management. 2012. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129138.pdf. Accessed 10 Sept 2015.
6. EudraVigilance Expert Working Group. Guideline on the use of statistical signal detection methods in the EudraVigilance data analysis system (EMA/106464/2006 rev.1). http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/11/WC500011434.pdf. Accessed 10 Sept 2015.
7. Lindquist M, Ståhl M, Bate A, Edwards IR, Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Saf.* 2000;23:533–42.
8. van Puijenbroek EP, Bate A, Leufkens HGM, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf.* 2002;11:3–10.
9. Roux E, Thiessard F, Fourrier A, Bégaud B, Tubert-Bitter P. Evaluation of statistical association measures for the automatic signal generation in pharmacovigilance. *IEEE Trans Inf Technol Biomed.* 2005;9:518–27.
10. Almenoff JS, LaCroix KK, Yuen NA, Fram D, Dumouchel W. Comparative performance of two quantitative safety signalling methods: implications for use in a pharmacovigilance department. *Drug Saf.* 2006;29:875–87.
11. Hammond IW, Gibbs TG, Seifert HA, Rich DS. Database size and power to detect safety signals in pharmacovigilance. *Expert Opin Drug Saf.* 2007;6:713–21.
12. Hochberg AM, Hauben M, Pearson RK, O'Hara DJ, Reisinger SJ, Goldsmith DI, et al. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. *Drug Saf.* 2009;32:509–25.
13. 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.
14. Bailey S, Singh A, Azadian R, Huber P, Blum M. Prospective data mining of six products in the US FDA Adverse Event Reporting System: disposition of events identified and impact on product safety profiles. *Drug Saf.* 2010;33:139–46.
15. Wisniewski AFZ, Juhlin K, Laursen MV, Macia MM, Manlik K, Pinkston VK, et al. Characterisation of databases (DBs) used for signal detection (SD): results of a survey of IMI PROTECT work package (WP) 3 participants. In: 28th international conference on pharmacoepidemiology and therapeutic risk management, 23–26 Aug 2012, Barcelona. <http://www.imi-protect.eu/documents/WisniewskietalCharacterisationofdatabasesusedforsignaldetectionposterfinalICPE2012.pdf>. Accessed 10 Sept 2015.