



Exploring Completeness of Adverse Event Reports as a Tool for Signal Detection in Pharmacovigilance

Inyoung Lee, MS¹ · Jeremy D. Jokinen, PhD^{1,2} · Stephanie Y. Crawford, MPH, PhD^{1,3} · Gregory S. Calip, PharmD, MPH, PhD^{1,3,4} · Ryan D. Kilpatrick, PhD⁵ · Todd A. Lee, PharmD, PhD^{1,3}

Received: 7 April 2020 / Accepted: 14 July 2020 / Published online: 27 July 2020
© The Drug Information Association, Inc 2020

Abstract

Background Completeness of adverse event (AE) reports is an important component of quality for good pharmacovigilance practices. We aimed to evaluate the impact of incorporating a measure of completeness of AE reports on quantitative signal detection.

Methods An internal safety database from a global pharmaceutical company was used in the analysis. *vigiGrade*, an index score of completeness, was derived for each AE report. Data from various patient support programs (PSPs) were categorized based on average *vigiGrade* score per PSP. Performance of signal detection was compared between: (1) weighting and not weighting by *vigiGrade* score; and, (2) well documented and poorly documented PSPs using sensitivity, specificity, area under the receiver operating characteristics curve (AUC) and time-to-signal detection.

Results The ability to detect signals did not differ significantly when weighting by *vigiGrade* score [sensitivity (50% vs. 45%, $p=1$), specificity (82.8% vs. 82.8%, $p=1$), AUC (0.66 vs. 0.63, $p=0.051$) or time-to-signal detection (HR 0.81, $p=0.63$)] compared to not weighting. Well documented PSPs were better at detecting signals than poorly documented PSPs (AUC 0.66 vs. 0.52; $p=0.041$) but time-to-signal detection did not differ significantly (HR 1.54, $p=0.42$).

Conclusion Completeness of AE reports did not significantly impact the ability to detect signals when weighting by *vigiGrade* score or restricting the database based on the level of completeness. While the *vigiGrade* helps provide quality assessments of AE reports and prioritize cases for review, our findings indicate the tool might not be useful for quantitative signal detection when used by itself.

Keywords *vigiGrade* · Patient support program · Weighting · Completeness · Pharmacovigilance · Well documented

Introduction

Signal detection is a process to identify a potential causal association between drugs and adverse events (AE) that requires further evaluation [1]. Among many different sources of data for signal detection, spontaneous reporting systems (SRS) are the most widely used [2,3]. One of the major limitations of SRS is variability in the quality of the AE reports [2,4]. Completeness of AE report has been identified as a key parameter of quality for good pharmacovigilance practices [5,6]. Completeness may work as a good indicator of quality as AE reports that are missing key information will provide limited value to drug safety surveillance [7,8].

The reporter's knowledge, beliefs, and attitudes are factors associated with underreporting of AEs which negatively impacts quantitative signal detection [9]. The factors

✉ Inyoung Lee
ilee37@uic.edu

¹ Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street MC 871, Chicago, IL 60612, USA

² Bristol-Myers Squibb Company, New York, NY, USA

³ Center for Pharmacoepidemiology and Pharmacoeconomics Research, University of Illinois at Chicago, Chicago, IL, USA

⁴ Flatiron Health, Inc., New York, NY, USA

⁵ AbbVie Inc., North Chicago, IL, USA

associated with underreporting can also affect completeness of reports. For example, depending on the reporter's attitudes, one may decide not to report an AE or provide partial information regarding the AE. Conversely, reporters may provide a more complete report when they suspect or believe that the drug and AE are associated [10]. This suggests that completeness of reports could impact quantitative signal detection by having shared risk factors with underreporting.

Increasingly more AE reports are received from manufacturer-patient interaction via manufacturer-sponsored programs such as patient support programs (PSP) [11]. The PSPs aim to help patients in managing their conditions or medications and result in more contact between the manufacturer and the patients or healthcare professionals, leading to additional AE reports that would not have been received in the absence of PSPs [12]. Importantly, the completeness or the quality of the reports coming from PSPs may differ from reports from other sources. Klein et al. reported that PSPs had a higher proportion of "precautionary reports" which they defined as reports with minimal information compared to other non-specified sources [13]. Those authors found that the precautionary reports can negatively impact signal detection by leading to signals that are ultimately determined to be 'false-positives' and masking true signals. In contrast, Harinstein et al. found that reports from industry-sponsored programs, which included PSPs, had more complete data compared to reports from non-industry-sponsored programs [14]. The impact of data from PSPs on the overall quality of safety database is unclear and warrants further investigation.

The objective of this study was to assess the impact of the completeness of reports on signal detection. Specifically, we sought to compare the performance of signal detection from data when weighting by the completeness of reports versus not weighting. Furthermore, we compared the ability to detect signals between using data from well documented PSPs versus moderately to poorly documented PSPs.

Methods

Data Source

The database we used for this analysis consisted of reports of AE received within a major, global pharmaceutical company from 2015 to 2017. AE reports relevant to the manufacturer's products were collected globally from various sources: direct spontaneous reports from patients or healthcare professionals, clinical trials, non-interventional studies, scientific literature, pharmacoepidemiological studies, PSPs, market research programs, and internet sources. Information on patient demographics, drugs, reactions, outcomes,

reporting sources, therapies and indications were available in the safety database.

Identifying Cases from PSPs

Reports from PSPs were identified by searching for program names and program service providers' names in the safety database. Reports identified as being from PSPs but did not contain enough information to be matched to a specific PSP were excluded from analyses that required identification of specific PSP.

Reference Set

A reference set ("gold standard") of drug-AE pairs was established in order to evaluate the ability to detect true signals within the dataset. The reference set included a list of internally validated signals and signals for which regulatory authorities requested investigation between 2015 and 2018. All signals included in the reference set had undergone internal signal assessment, meaning that each of the signals was evaluated for a causal association between the drug and the AE. The causal assessments used a combination of resources including review of the cases within the internal safety database, expert opinion, existing literature, and plausible biological mechanism. The positive references were the group of AE-drug signals that were confirmed to be causally associated. The negative references consisted of AE-drug signals that were not confirmed to be causally associated. There were 20 positive reference signals and 58 negative reference signals for this analysis. Adverse events from the reference set were recorded in medical concepts, which could be a single or a combination of the following: Medical Dictionary for Regulatory Activities (MedDRA) Preferred Terms (PTs), Lowest Level Terms (LLTs), standardised MedDRA queries (SMQs) or company MedDRA queries (CMQs).

Signal Detection

The multi-item Gamma Poisson Shrinker (MGPS) method, a Bayesian approach, was utilized to detect signals. The MGPS lowers false positive signals by providing a robust estimate when the events are rare [2,15,16]. The Empirical Bayes geometric mean (EBGM) score was calculated using the MGPS method for drug-AE pairs in the reference sets within the safety database. A signal was determined to be detected for drug-AE pairs when the lowest fifth percentile of EBGM (EB05) was greater than 2, i.e., threshold of $EB05 > 2$ [15]. Signal detection was performed using SAS software, version 9.4. Specific SAS codes used for EBGM calculations were available from the publicly accessible GitHub repository [17].

Weighting by Completeness of Reports

The Uppsala Monitoring Centre (UMC) developed *vigiGrade*, an index score of completeness for spontaneous reports [6]. Calculating *vigiGrade* involves ten fields within the safety database that are important in assessing causality. The ten fields are: time-to-onset, indication, outcome, sex, age, dose, country, primary reporter, report type, and comments [6]. Different penalties are applied for each of the missing variables based on their importance in causality assessment following the criteria described in Bergvall et al. [6]. The *vigiGrade* score can range from 0.07 (reports missing information in all ten fields) to 1 (complete report) [6]. According to the *vigiGrade* score criteria, uninformative text fragments in the comment field, which consists of information in free text, should be excluded [6]. In order to approximate informativeness of the free text field within the database used in this study, the length of free text strings were measured and comments equal to or greater than 288 characters were considered informative. This was based on the assumption that an informative comment consists of at least three sentences with the average number of words in an English sentence being 17.5 and the average length of an English word being 4.5 [18–21]. Spaces between words and sentences were also accounted for in calculating the minimum number of characters needed for an informative comment.

Each case report was assigned a *vigiGrade* score based on the information available. All reports were frequency weighted by multiplying the *vigiGrade* score. Therefore, reports with perfect *vigiGrade* scores had a frequency weight of 1 and reports with the lowest *vigiGrade* scores had a frequency weight of 0.07. For example, if a case had a *vigiGrade* score of 0.5, it had a frequency weight of half a case. By mechanism, there became “less” reports with poorly documented reports compared to well documented reports. Total weights for reports of a particular drug-AE pair were summed to calculate the weighted frequency of drug-AE pairs. The purpose of frequency weighting the reports by *vigiGrade* was to give more weights to well documented reports as we expected that reporters may provide more complete reports when they suspected causality.

Signal Detection Performance

Using the gold standard reference sets of positive and negative signals, the performance of the signal detection method was compared when the reports were weighted by their *vigiGrade* score versus using the unweighted reports. The signal detection performance was evaluated using four measures: sensitivity, specificity, area under the receiver operating characteristics (ROC) curve (AUC), and time-to-signal detection.

Sensitivity is the ability to detect a signal in a positive reference drug-AE pair. Specificity is the ability to accurately classify the negative references as not triggering a drug-AE signal. Sensitivity and specificity can be calculated using the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) cases from comparing the gold standard (reference set) to test results (detected signals). Sensitivity was calculated as $TP/(TP + FN)$ and specificity as $TN/(FP + TN)$. McNemar’s test was used to test whether sensitivity or specificity was greater when weighting by *vigiGrade* compared to not weighting. While sensitivity and specificity were based on the threshold of $EB05 > 2$ for detecting signals, AUC is a general performance measure that does not rely on a single threshold for signal detection. ROC curves were plotted as sensitivity against (1-specificity) and AUC values were compared using chi-square tests.

In order to estimate time-to-signal detection, $EB05$ was calculated for each drug-AE pair in the positive references for each cumulative months between 2015 and 2017. Time-to-signal detection was measured as the number of months from the index month, T_0 , until the month the signal was detected or the end of study follow-up whichever occurred earlier. T_0 was defined as January 2015 or the month of first global drug approval date for the drug in the drug-AE pair, whichever occurred later.

Kaplan–Meier curves were plotted for time-to-signal detection using the *vigiGrade* weighted data and the unweighted data. The proportional hazards assumption was tested by graphically assessing the scaled Schoenfeld residuals and by including time dependent covariates in the Cox proportional hazards model to assess their statistical significance. A Cox proportional hazards model was used to calculate the hazard ratio adjusting for time since drug approval.

Sensitivity analyses were conducted using different thresholds for determining informativeness of the free text comments field when calculating *vigiGrade* score. Two separate scenarios on the length of the comment field were used. The first was to double the length of the base case such that comments less than 576 characters were penalized. The second analysis took the opposite extreme and only penalized comments where there was no information (i.e., the comment was missing).

Comparing Signal Detection Performance Across PSPs Based on *vigiGrade* Score

The mean *vigiGrade* scores were calculated for reports from non-PSP sources and from PSPs. Additionally, the mean *vigiGrade* scores were calculated for each of the PSPs that were included in the database. From the range of the mean *vigiGrade* scores by PSP, those with a mean score in the

highest tertile were classified as well documented PSPs; the middle tertile as moderately documented PSPs; and the lowest tertile as poorly documented PSPs. All performance measures (sensitivity, specificity, AUC and time-to-signal detection) were compared for well documented PSPs and moderately documented PSPs to poorly documented PSPs.

Association Between Average vigiGrade Score and True Signals

For each drug-AE pair in the reference set, the average vigiGrade score was calculated from all the reports containing the drug-AE pair of interest. The association between the average vigiGrade score and the true signals (i.e., all positive reference signals) was tested using the Wilcoxon-Mann-Whitney test.

Results

In the safety database, the vigiGrade score of AE reports ranged from 0.14 to 1. The median score was 0.56 and mode was 1 with 19.5% of the total cases having complete fields. The distribution of vigiGrade scores (Fig. 1) was spread throughout the range but as the vigiGrade was calculated by applying certain penalties based on missing fields, there were certain vigiGrade scores with high frequency. The field with the most missing data was time-to-onset (44.9%), followed by age (40.2%) and dose (31.2%). As penalties for missing time-to-onset, age and dose were 50%, 30% and 10% respectively, we saw peaks in vigiGrade scores at 0.9

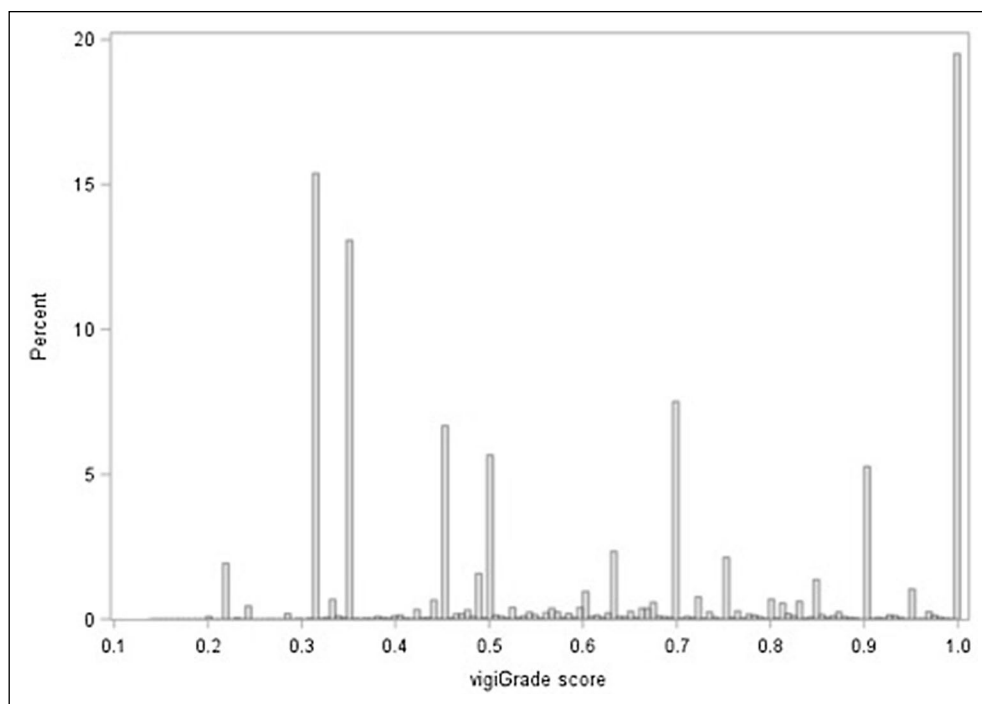
(missing dose), 0.7 (missing age), 0.5 (missing time-to-onset), 0.45 (missing time-to-onset and dose) 0.35 (missing time-to-onset and age) and 0.315 (missing time-to-onset, age and dose). The average vigiGrade score for reports containing drug-AE pairs in the positive reference set (“true signals”) was 0.63 compared to 0.68 for reports containing drug-AE pairs in the negative reference set (“false signals”). There was no association between the average vigiGrade score and the true signals (Wilcoxon-Mann-Whitney test $p=0.34$).

Throughout the 36-month study period, the mean vigiGrade score for cumulative months increased until 22 months and then decreased over time (Fig. 2).

Weighting the reports by vigiGrade scores did not result in significantly different sensitivity (50% vs. 45%; $p=1$), specificity (82.8% vs. 82.8%; $p=1$) or AUC (0.66 vs. 0.63; $p=0.051$) for signal detection compared to not weighting (Table 1). Varying the length of free text in assessing the informativeness in the comments field did not affect any of the results in sensitivity analyses.

In time-to-signal detection analysis, not applying weights to reports resulted in more signals detected over time than weighting by vigiGrade scores, as shown in Fig. 3 (55% vs. 50%). The incidence rate of signal detection was higher when not weighting by vigiGrade score (29 signals detected per 1000 data-months vs. 22.3 signals detected per 1000 data-months, $p=0.55$) compared to weighting but the difference was not statistically significant (Table 2). Weighting the reports with vigiGrade score had lower likelihood to detect signals at any given point in time compared with not weighting when adjusted for time since drug approval but

Fig. 1 Distribution of vigiGrade score



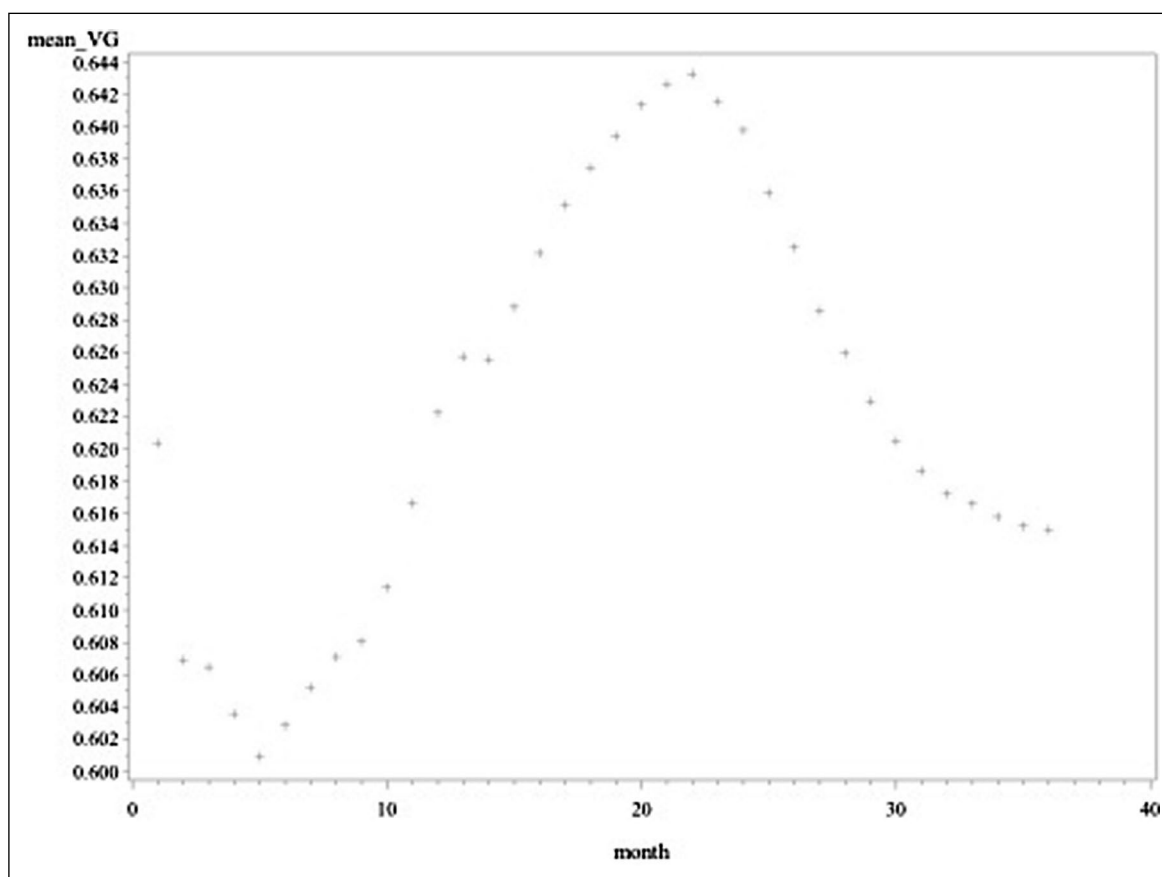


Fig. 2 Cumulative monthly vigiGrade score over time

Table 1 Sensitivity, specificity and area under the receiver operating characteristics (ROC) curve (AUC) for signal detection weighting by vigiGrade vs. not weighting by vigiGrade

	Sensitivity ^a	Specificity ^a	AUC ^b
Not weighting by vigiGrade	45%	82.8%	0.63
Weighting by vigiGrade (base case)	50% ($p=1$)	82.8% ($p=1$)	0.66 ($p=0.051$)
Weighting by vigiGrade (sensitivity analysis 1) ^c	50% ($p=1$)	82.8% ($p=1$)	0.66 ($p=0.085$)
Weighting by vigiGrade (sensitivity analysis 2) ^d	50% ($p=1$)	82.8% ($p=1$)	0.66 ($p=0.051$)

^aSensitivity and specificity were compared using McNemar's test. The reference group for all comparisons was 'Not weighting by vigiGrade'

^bAUC values were compared using chi-square test. The reference group for all comparisons was 'Not weighting by vigiGrade'

^cPenalized when length of comment was less than 576 when calculating vigiGrade

^dPenalized when comment was missing when calculating vigiGrade

the result was not statistically significant (HR 0.81, $p=0.63$). The results remained consistent in sensitivity analyses.

The mean vigiGrade score was higher for reports from PSPs compared to non-PSP sources (0.62 vs. 0.56; $p < 0.0001$) indicating that reports from PSPs were more complete than non-PSP sources. Among reports from PSPs, 62% were able to be identified as having come from a specific PSP and were used for the rest of the analyses. When the mean vigiGrade score was calculated for each

PSP, the scores ranged from 0.22 to 1. Dividing the mean vigiGrade scores by tertile, the PSPs with a mean vigiGrade score greater than 0.70 (highest tertile) were defined as well documented PSPs; the moderately documented PSPs had mean vigiGrade scores greater than 0.52 and less than or equal to 0.70; and those with mean vigiGrade score less than or equal to 0.52 (lowest tertile) were defined as poorly documented PSPs.

Fig. 3 Kaplan–Meier curve for signal detection weighting by vigiGrade vs. not weighting by vigiGrade

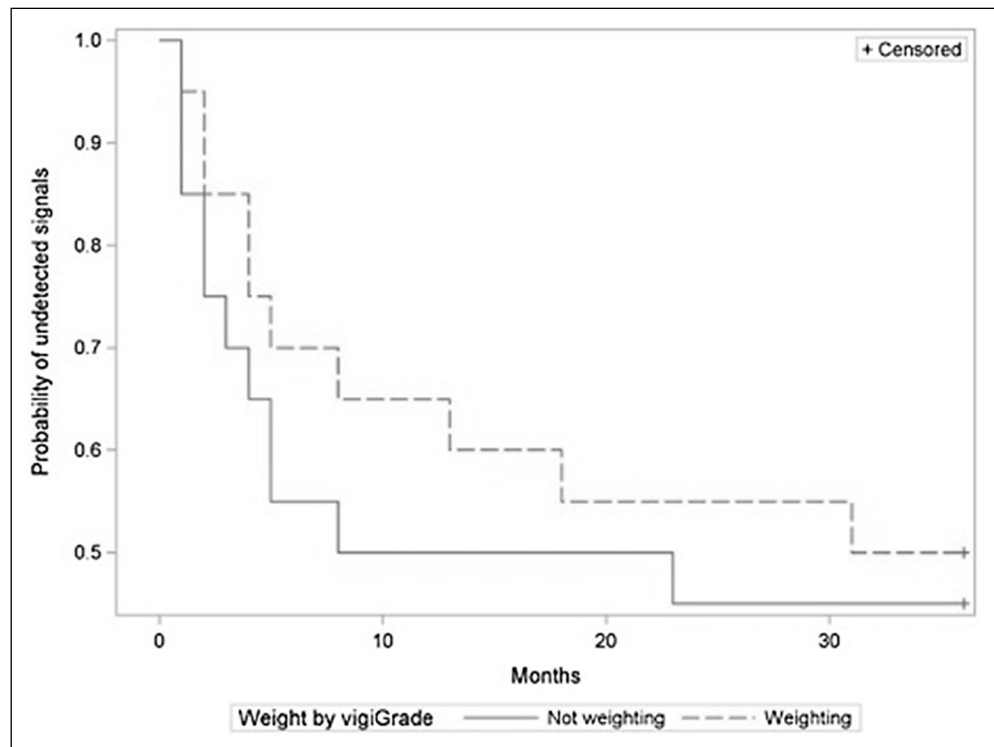


Table 2 Incidence rates and hazard ratios (HR) of signal detection weighting by vigiGrade vs. not weighting by vigiGrade

	N of signals detected ^a (%)	Data-months	Incidence rate (number of signals detected per 1000 data-months)	Crude HR (95% CI)	<i>p</i> -value	Adjusted HR (95% CI) ^b	<i>p</i> -value
Not weighting by vigiGrade	11 (55%)	379	29	1	Ref	Ref	Ref
Weighting by vigi-Grade (base)	10 (50%)	448	22.3	0.80 (95% CI 0.34–1.87)	0.60	0.81 (0.34–1.91)	0.63
Weighting by vigi-Grade (sensitivity analysis 1) ^c	11 (55%)	401	27.4	0.98 (95% CI 0.42–2.26)	0.96	0.99 (0.43–2.27)	0.97
Weighting by vigi-Grade (sensitivity analysis 2) ^d	10 (50%)	461	21.7	0.79 (95% CI 0.33–1.86)	0.59	0.80 (0.34–1.89)	0.61

^aTotal of N in this analysis is the total number of positive reference signals, which should equal to 20

^bHR adjusted for time since drug approval

^cPenalized when length of comment is less than 576 when calculating vigiGrade

^dPenalized when comment is missing when calculating vigiGrade

Well documented PSPs and poorly documented PSPs did not demonstrate significantly different sensitivity (40% vs. 30%; $p=0.25$) and specificity (93.1% vs. 87.9%; $p=0.29$) when $EB05 > 2$ threshold was used (Table 3). Based on significant difference in AUC of well documented PSPs and poorly documented PSPs (0.66 vs. 0.52; $p=0.041$), well documented PSPs had better ability at detecting signals compared to poorly documented PSPs. Moderately documented PSPs and poorly documented PSPs did not result in

significant difference in sensitivity (25% vs. 30%; $p=0.36$), specificity (86.2% vs. 87.9%; $p=0.44$) or AUC (0.54 vs. 0.52; $p=0.70$).

The Kaplan–Meier curve in Fig. 4 depicts that well documented PSPs detected 45% of positive references over time compared to moderately and poorly documented PSPs detecting 30% each. The incidence rate of detecting signals was higher for well documented PSPs than poorly documented PSPs (16.8 signals per 1000 data-months vs. 10.1 signals per

1000 data-months, $p=0.3343$) although the difference was not significant (Table 4). Using data from well documented PSPs did not lead to significantly higher likelihood of detecting signals at any given point in time compared to poorly documented PSPs when adjusted for time since drug approval (HR 1.54; $p=0.42$) indicating there was no significant difference in time-to-signal detection between the two groups. There was also no significant difference in the likelihood to detect signals over time between moderately documented PSPs and poorly documented PSPs (HR 1.15; $p=0.81$).

Discussion

Weighting the AE reports by their vigiGrade scores, an indicator of completeness of reports, did not result in significantly better performance in signal detection compared to

not weighting. The completeness of reports varied widely between PSPs (0.22–1). Well documented PSPs (PSPs with average vigiGrade score in the highest tertile) were better at detecting signals compared to poorly documented PSPs (PSPs with average vigiGrade score in the lowest tertile) based on AUC comparison.

The comparison of well documented PSPs and poorly documented PSPs in our study suggests that using more complete reports may lead to better signal detection in general. We found that reports from PSPs were more complete than non-PSP sources (mean vigiGrade score 0.62 vs. 0.56; $p < 0.0001$). Inferred from our findings in this study, we might expect that reports from PSPs would be better at detecting signals compared to non-PSP sources. On the contrary, a previous study found that signal detection performance was higher for reports from non-PSP sources compared to PSPs [22]. There are several potential

Table 3 Sensitivity, specificity and area under the receiver operating characteristics (ROC) curve (AUC) for signal detection for well documented PSPs; moderately documented PSPs; and poorly documented PSPs

	Sensitivity	<i>p</i> -value	Specificity (%)	<i>p</i> -value	AUC	<i>p</i> -value
Well documented PSPs ^a	40%	0.25	93.1	0.29	0.66	0.041
Moderately documented PSPs ^b	25%	0.36	86.2	0.44	0.54	0.70
Poorly documented PSPs ^c	30%	Ref	87.9	Ref	0.52	Ref

^aWell documented PSPs are defined as PSPs with average vigiGrade score less than or equal to the highest tertile

^bWell documented PSPs are defined as PSPs with average vigiGrade score in the middle tertile

^cPoorly documented PSPs are defined as PSPs with average vigiGrade score in the lowest tertile

Fig. 4 Kaplan–Meier curve for signal detection using data from well documented PSPs; moderately documented PSPs and poorly documented PSPs

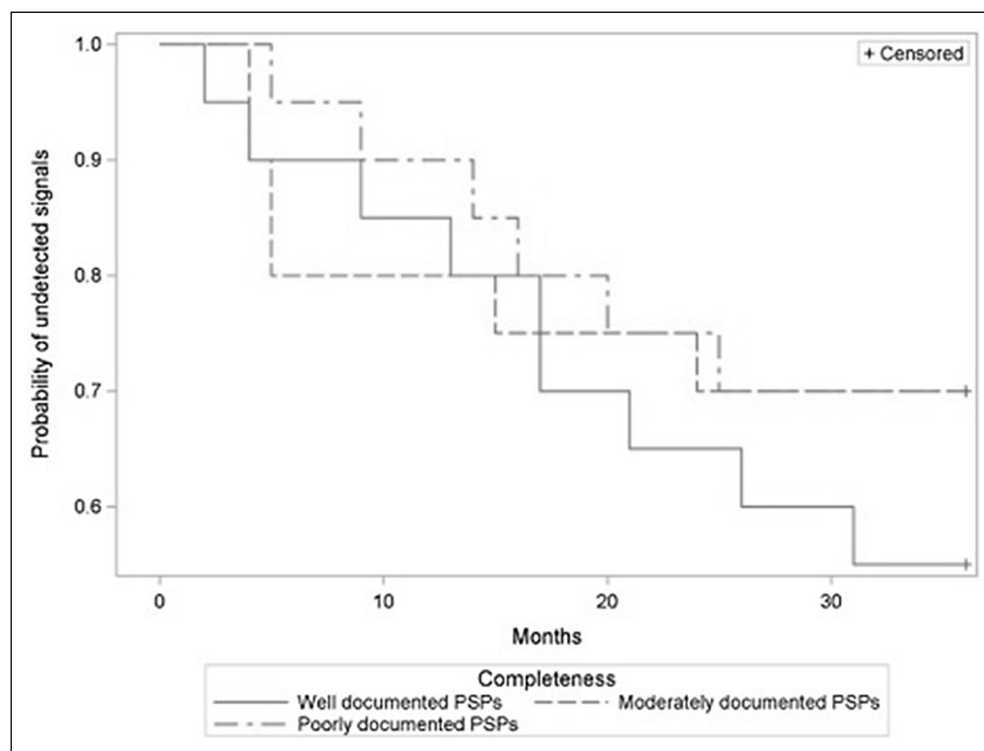


Table 4 Incidence rates and hazard ratios (HR) of signal detection using data from well documented PSPs and moderately documented PSPs vs. poorly documented PSPs

	N of signals detected ^a (%)	Data-months	Incidence rate (number of signals detected per 1000 data-months)	Crude HR (95% CI)	<i>p</i> -value	Adjusted HR (95% CI) ^b	<i>p</i> -value
Well documented PSPs ^c	9 (45%)	536	16.8	1.62 (0.58–4.55)	0.36	1.54 (0.55–4.32)	0.42
Moderately documented PSPs ^d	6 (30%)	561	10.7	1.07 (0.34–3.30)	0.91	1.15 (0.37–3.58)	0.81
Poorly documented PSPs ^e	6 (30%)	593	10.1	1	Ref	1	Ref

^aTotal of N in this analysis was the total number of positive reference signals, which was 20

^bHR adjusted for time since drug approval

^cWell documented PSPs are defined as PSPs with average vigiGrade score less than or equal to the highest tertile

^dModerately documented PSPs are defined as PSPs with average vigiGrade score in the middle tertile

^ePoorly documented PSPs are defined as PSPs with average vigiGrade score in the lowest tertile

explanations for this occurrence. One possibility is due to differences between the sources. Reports from PSPs may be more complete due to the system in place for PSP operations. These systems facilitate continued interaction between patients and PSP providers. The increased interaction could lead to easier follow-up of potential AEs compared to non-PSP sources and thus result in more complete reports. PSPs and non-PSP sources may also differ in the reported drugs and AEs leading to a different pool of safety data ultimately affecting the result of signal detection. Another possibility is that completeness of reports may not directly lead to better signal detection since the two key information elements required for quantitative signal detection are information on suspect drug and suspect AE. While the completeness of reports may be helpful for signal assessment after signals are detected, it may not impact signal detection as all reports have information on suspect drug and suspect AE which is enough to trigger a signal using methods such as MGPS. The performance of signal detection could have been influenced more on the reported drugs and AEs rather than how complete the reports are.

We found that the reports of true signals in this study were no more likely to be complete than reports of “false signals” (i.e., signals confirmed to be not causally associated) and the information forming vigiGrade were missing completely at random with regards to true safety signals. This could have been the underlying reason for no difference observed in signal detection when weighting versus not weighting by the vigiGrade score. From this, we can infer that reporters may not necessarily provide more information when they suspect an AE. It should be noted, however, that as we used an internal safety database within a pharmaceutical company, the reporters may be systematically different from the reporters in SRS databases or other pharmaceutical company’s databases.

There are several limitations in this study. First, we assessed uninformative free text by the length of comment which is not the most accurate means to measure informativeness and may lead to potential misclassification of informative comments. Although we conducted sensitivity analyses varying the length of comments to be considered as informative, misclassification is still a possibility that could affect the vigiGrade score. For part of the analysis, we were not able to utilize all reports from PSPs as 38% of reports were unable to be matched to a specific PSP. This can limit generalizability of the results. Another limitation is that there was a limited number of reference signals available leading to insufficient power to detect a significant difference when comparing signal detection performance. Additionally, the findings in this study were based on a single pharmaceutical company’s database. The completeness of reports could vary between companies and the result may not be generalizable to a database containing different drug-AE information. Finally, using a pharmaceutical company’s database meant that certain information could not be disclosed due to confidential nature. In this study, we were not able to report the number of AE reports within the company’s internal safety database. While we acknowledge this limitation, we believe our study provides a unique insight into utilizing a company’s safety database for signal detection, which enabled us to identify reports from PSPs, an information that is not available in the U.S. Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS).

vigiGrade is used at UMC, which manages the World Health Organization global AE database for pharmacovigilance activities, mainly for communicating with countries regarding data quality and is also incorporated in vigiRank, a predictive model that prioritizes signals [10,23]. In our study we found that utilizing vigiGrade by weighting or restricting

the database did not improve quantitative signal detection. The *vigiGrade* score was developed to highlight systematic data quality issues and the applied penalties may not be specific to benefit signal detection [6]. It may be beneficial to develop a scoring system specifically designed to enhance signal detection by selecting variables that are associated with true signals or applying different penalties to missing values.

While *vigiGrade* by itself may not be useful for signal detection, *vigiRank*, which incorporates *vigiGrade* into the model, performed better than a method of disproportionality analysis at signal detection [10]. This indicates *vigiGrade*'s potential use to assist signal detection as a component in a broader model. Alternatively, *vigiGrade* may serve better as a tool that is used to assess completeness of reports to categorize them as well documented or poorly documented. Well documented reports could serve as the starting point for signal assessment as they contain more information that is useful for causality assessment. Regardless of its use in quantitative signal detection, *vigiGrade* is a useful tool in Pharmacovigilance that provides valuable information by highlighting data quality issues. Within pharmaceutical company's pharmacovigilance departments, *vigiGrade* could be used as an internal quality control tool such as identifying if there is a field systematically missing from a certain PSP as well as a tool to assist prioritization of case review.

Conclusions

Weighting the AE reports by an indicator for completeness did not significantly improve signal detection compared to not weighting. In this study, reports from well documented PSPs were better at signal detection compared to poorly documented PSPs, but it is uncertain whether using more complete report in general leads to better signal detection as previous findings suggest otherwise. Further research is warranted in different databases to generalize the findings. We recommend *vigiGrade* to be used as a tool for quality control and prioritization of case review.

Author contributions

All authors contributed to the study conception and design. Material preparation and data analysis were performed by IL. The first draft of the manuscript was written by IL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

No source of funding was used to assist in the preparation of this study.

Compliance with Ethical Standards

Conflict of interest

Inyoung Lee was supported by the UIC-AbbVie Fellowship in Pharmacovigilance and Patient Safety. Jeremy D. Jokinen was employed by AbbVie, Inc. and owned stocks of the company while the submitted work was conducted; and currently is an employee and shareholder of Bristol-Myers Squibb Company. Gregory S. Calip received grants from AbbVie, Inc. and Pfizer, Inc. outside the submitted work; and at the time of publication, reports current employment with Flatiron Health, Inc., which is an independent subsidiary of the Roche group. Ryan D. Kilpatrick is an employee of AbbVie, Inc. and may own stocks of the company. All other authors have no conflicts of interest to disclose.

References

1. Hauben M, Aronson JK. Defining 'Signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. *Drug Saf.* 2009;32(2):99–110.
2. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf.* 2009;18(6):427–36.
3. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiol Drug Saf.* 2007;16(4):359–65.
4. Koutkias VG, Jaulent M-C. Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. *Drug Saf.* 2015;38(3):219–32.
5. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module I – Pharmacovigilance systems and their quality systems [Internet]. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-module-i-pharmacovigilance-systems-their-quality-systems_en.pdf. Accessed 5 Nov 2015.
6. 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.
7. Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology. Chapter 10 Postmarketing Spontaneous Pharmacovigilance Reporting System.* Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated; 2012.
8. Ribeiro A, Lima S, Zampieri M-E, Peinado M, Figueras A. Filling quality of the reports of adverse drug reactions received at the Pharmacovigilance Centre of São Paulo (Brazil): missing information hinders the analysis of suspected associations. *Expert Opin Drug Saf.* 2017;16(12):1329–34.
9. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. *Drug Saf.* 2009;32(1):19–311.
10. 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.
11. ISMP Quarter Watch: A Critique of a Key Drug Safety Reporting System [Internet]. 2015. <https://www.ismp.org/quarterwatch/drug-safety-reporting-system>. Accessed 9 Mar 2018.
12. Portnoff JM, Lewis DJ. The enigma of pharmacovigilance of patient support programs: a survey of marketing authorization holders in Europe. *Ther Innov Regul Sci.* 2017;51(4):486–93.
13. Klein K, Scholl JHG, De Bruin ML, van Puijtenbroek EP, Leufkens HGM, Stolk P. When more is less: an exploratory study of the

- precautionary reporting bias and its impact on safety signal detection. *Clin Pharmacol Ther.* 2018;103(2):296–303.
14. Harinstein L, Kalra D, Kortepeter CM, Muñoz MA, Wu E, Dal Pan GJ. Evaluation of postmarketing reports from industry-sponsored programs in drug safety surveillance. *Drug Saf.* 2019;42(5):649–55.
 15. Poluzzi E, Raschi E, Piccinni C, De Ponti F. Data mining techniques in pharmacovigilance: analysis of the publicly accessible FDA adverse event reporting system (AERS). In: *Data mining applications in engineering and medicine*. InTech; 2012.
 16. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther.* 2013. <https://doi.org/10.1038/clpt.2013.24>.
 17. Henderson M. Disproportionality measures [Internet]. GitHub. <https://github.com/statmike/Disproportionality-Measures-with-SAS>. Accessed 10 Oct 2019.
 18. Cutts M. *Oxford guide to plain english* [Internet]. OUP Oxford; 2013. <https://books.google.com/books?id=xd11AQAAQBAJ>.
 19. Palmer DD. A trainable rule-based algorithm for word segmentation. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*. Madrid, Spain: Association for Computational Linguistics; 1997. p. 321–8.
 20. Pierce JR. *An introduction to information theory: symbols, signals and noise* [Internet]. Dover Publications; 2012. (Dover Books on Mathematics). <https://books.google.com/books?id=eKvhiI2ogwEC>.
 21. Barnard G. Statistical calculation of word entropies for four Western languages. *IRE Trans Inf Theory.* 1955;1(1):49–53.
 22. Lee I, Lee TA, Crawford SY, Kilpatrick RD, Calip GS, Jokinen JD. Impact of adverse event reports from marketing authorization holder-sponsored patient support programs on the performance of signal detection in pharmacovigilance. *Expert Opin Drug Saf.* 2020. <https://doi.org/10.1080/14740338.2020.1792883>.
 23. UMC | *vigiMethods* [Internet]. <https://www.who-umc.org/vigibase/vigilyze/vigimethods/>. Accessed 3 Nov 2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.