The validity and reliability of a signal impact assessment tool†

Leàn Rolfes1,2*, Judith Kolfschoten3, Florence van Hunsel1,2 and Eugène van Puijenbroek1

1Netherlands Pharmacovigilance Centre Lareb, Hertogenbosch, The Netherlands
2Department of Pharmacy: Pharmacotherapy and Pharmaceutical Care, University of Groningen, Groningen, The Netherlands
3Department of Pharmacy: Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands

ABSTRACT

Background The Netherlands Pharmacovigilance Centre Lareb uses an operating Signal Impact Assessment Tool (SIAT) as aid in signal selection. SIAT prioritized signals into one of four categories: strong/moderate signal strength and similarly health impact. Although the SIAT has been used for many years, validity and reliability was never explored.

Purpose The aim of this study is to test the validity and reliability of the operating and weight-adjusted SIAT.

Method For validity testing, judgments of a Delphi panel of three pharmacovigilance experts were used as a ‘gold standard’. First, the panel judged the weighting of the items included in the SIAT. Then, during two phases, the panel rated 40 signals in one of the four categories. Two researchers scored the signals using the SIAT. Panel judgments were compared with scores for the operating and weight-adjusted SIAT. Inter- and intra-observer variability was also tested. The Cohen’s Kappa coefficient (k) was calculated. At least substantial agreement (k > 0.6) was considered to be necessary for an acceptable reliability.

Results Validity did not meet predefined criteria: operating and weight-adjusted tool, respectively, k-phase1 = 0.83 and 0.83, k-phase2 = 0.18 and 0.36. Differences were found for signal strength and health impact. Inter- and intra-observer variabilities were good, k of 0.78 and 0.72, respectively.

Conclusions The SIAT was found to have an insufficient validity and proper reliability. Although SIAT scores should not be decisive in the decision making process, items included can be used as an aid to decide which signals deserve further investigation. Copyright © 2016 John Wiley & Sons, Ltd.

INTRODUCTION

Spontaneous reporting systems in pharmacovigilance primary aim at the timely detection of unknown adverse drug reactions (ADRs) or unknown information about known ADRs, also called ‘signal detection’. A signal contains information on a possible causal relation between an adverse event and a drug, the relation being previously unknown or incompletely documented. The number of reports that is required to generate a signal is dependent on the seriousness of the event and the quality of the information. In practice, a signal is something that would be considered important and might impact on patient management or the balance of benefits and risks. Signals should be evaluated as soon as possible, particularly when they are a potential threat to public health or require for other reasons prompt regulatory action. Unfortunately, there is limited information available on the way signals are prioritized for further action including signal validation. Having a method or tool in place to prioritize signals is important for a pharmacovigilance center to ensure that the most important signals, based on strength and health impact, get more immediate action and to make the decision process which signal is most important less based on personal experience and preferences.

This issue of signal prioritization has also been a point of the attention in other countries. In the UK, a signal impact analysis for a regulatory setting was developed and implemented. This tool has a systematic approach to drug safety issue prioritization that may help to reduce the subjectivity of reliance on individual...
judgments. At the Netherlands Pharmacovigilance Centre Lareb, a similar signal impact assessment tool (SIAT) was developed for the procedures in place of the Dutch pharmacovigilance centre. This tool has been used for more than 10 years in order to facilitate the process of signal selection. The SIAT was drafted according to findings from previous research on the determinants important for signal selection. At the Netherlands Pharmacovigilance Centre Lareb, signal detection is carried out mainly by means of a case-by-case analysis. Each incoming ADR report undergoes an individual systematic review. All ADR reports with potential signal value are reviewed by trained assessors during a weekly meeting. Potential signals undergo a more detailed analysis. The SIAT is used in order to assess the ‘signal strength’ and ‘health impact’ of potential signals. The inputs and scores of the SIAT are shown in Table 1 for ‘signal strength’ and Table 2 for ‘health impact’. By assessing the signal strength and health impact, SIAT prioritized signals into one of four categories: strong/moderate signal strength and similarly health impact. ‘Poor’ signals will not make it through the first selection round. Because of this, signals that are scored using the SIAT are prioritized as ‘moderate’ of ‘strong’. At Lareb, all ADR reports are assessed case-by-case, and signals are mainly found in this manner, although a computer-assisted database screening tool is in place. Lareb works in close collaboration with the Dutch Medicine Evaluation Board (MEB) and until recently has informed the MEB periodically about the most relevant drug safety concerns in a so-called ‘quarterly-report’. Since 2015, reports on safety concerns are sent to the MEB on a continuous basis.

The validity (the degree to which the tool measures the construct it purports to measure) and the reliability (the degree to which the measurement if free from measurement error) have never been explored because this is a time-consuming process, and because the tool served its purpose in everyday pharmacovigilance work, there has never been a need to validate it. However, for the purpose of study, the need to test the validity and reliability of the SIAT on a dataset of The Netherlands Pharmacovigilance Centre arose. The aim of this study is to test the validity and reliability of the SIAT because the tool served its purpose in everyday pharmacovigilance work, there has never been a need to validate it. However, for the purpose of study, the need to test the validity and reliability of the SIAT on a dataset of The Netherlands Pharmacovigilance Centre arose. The aim of this study is to test the validity and reliability of the SIAT. This paper will describe the results as well as the process of testing such a tool.

Table 1. Inputs and scoring for signal strength

| Part I scoring for strength of the cases |
| Score | Features of the cases |
| 4 | Very well-documented cases. Information about latency time, demand and rechallenge, and indication and concomitant medication. Taking into account the confounding. |
| 2.5 | Well-documented cases |
| 1 | Moderately documented cases |

| Part II scoring for reported odd ratio Lareb |
| Score | 4 | Lower 95%CI > 2 |
| 3 | Lower 95%CI > 1 |
| 2 | Less than three reports, no ROR can be calculated ROR > 1, lower 95%CI < 1 |
| 1 | ROR < 1, lower 95%CI < 1 |

| Part III scoring for reported odd ratio WHO* |
| Score | 4 | Lower 95%CI > 2 |
| 3 | Lower 95%CI > 1 |
| 2 | Less than three reports, no ROR can be calculated ROR > 1, lower 95%CI < 1 |
| 1 | ROR < 1, lower 95%CI < 1 |

| Part IV scoring for literature |
| Score | Literature |
| 4 | Clinical or epidemiologic study, described clinical relevant CYP interaction, plausible mechanism |
| 3 | Several case report or case series. Shorty mentioned in case reports of other SRS. Described not (yet) clinical relevant CYP interaction. Possible mechanism |
| 2 | Briefly described in literature, a few case reports |
| 1 | No literature available or contrary ADR described in product information, information about CYP-interaction that invalidates causality |

*The database of the World Health Organization Collaborating Centre for International Drug Monitoring, the Uppsala Monitoring Centre (WHO-UMC).

Copyright © 2016 John Wiley & Sons, Ltd.
METHOD

Adjustments based on expert evaluation

A panel of three pharmacovigilance experts (two physicians and one pharmacist) was used in order to evaluate the SIAT. Twenty drug safety signals disseminated by Lareb between 2006 and 2007 were selected at random from the Lareb database. Associations described in the quarterly reports may present drug safety concerns, for example, new ADRs, interactions, but also overviews of ADRs of drugs that were marketed recently in The Netherlands.8,11 For this study, all overviews were excluded.

The panel of experts tested the face validity of the SIAT and the weighting of the several items included in signal strength and health impact. In the operating SIAT, all items have the same weighting. It was explored if this is a proper assumption in order to measure the signal strength and health impact of a signal. The panel of experts was asked to divide 100 points over items I–IV of signal strength and another 100 points over items V–VIII of health impact based on their importance for each individual signal. Average weighting was then determined for each individual input. The SIAT was adjusted based on the weightings, multiplying the scores with the weighting of the item.

Validity and reliability testing

Validity. The validity testing took place in two phases in which the judgments of the panel of pharmacovigilance experts were used as a ‘gold standard’. The Delphi method is a systematic, interactive forecasting method which relies on a panel of independent experts by answering questionnaires in two or more rounds.12 For the first phase of the validity testing, 20 drug safety signals disseminated by Lareb between 2006 and 2007 were selected at random from the Lareb database.

During the first Delphi round, the panel of experts was asked to individually score each signal for signal strength (strong or moderate) and health impact (major or minor). After the first round, signals without consensus were summarized and anonymously presented in a next round. Based on the summary, the panel of experts could adjust their answer given in the first round. Signals without consensus in the second round were discussed by the panel in a face-to-face final round to reach consensus. The 20 signals were also scored individually by two pharmacovigilance assessors using the SIAT. Signals without consensus were discussed between the assessors until consensus was researched.

The degree of agreement between the Delphi panel judgment and the SIAT score was explored by calculating the Cohen’s Kappa (k) coefficient.13 For the k, the following standards for strength of agreement can be used: slight (≤0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0).13 At least substantial agreement (k ≥ 0.61) was thought to be necessary for an acceptable validity. Agreement with the Delphi panel was calculated for the operating as well as weight adjusted SIAT.

In the second phase, the process was repeated with 20 new drug safety signals, disseminated by Lareb in 2006–2007 and 2013 and selected at random.

Reliability. In order to test the inter-observer variability, the 20 signals were scored individually by two pharmacovigilance assessors using the SIAT. To test the intra-observer variability, the pharmacovigilance assessors scored the signals again after 1 week. Given the construct of the SIAT, 1 week was considered to be a good time lag. The degree of agreement between the assessors was explored by calculating the k. At least substantial agreement (k > 0.6) was considered to be necessary for an acceptable reliability.

RESULTS

Adjustments based on expert evaluation

After face validity of the SIAT by the panel of experts, only few adjustments were made. The experts believed that SIAT already included all items relevant for assessment of signal strength and health impact. However, some items were merged. The weighting of the items is shown in Table 3. Cutoff point was arbitrary set to 250.

Validity and reliability testing

Validity. The first phase of this study showed almost perfect agreement between the judgments of the Delphi panel and the SIAT scores (k: 0.83 for operating as well as weight adjusted SIAT). Agreement was found for 16 signals. For two signals, the signal strength differed; the panel judged these signals as ‘moderate’, whereas the SIAT categorized it as ‘strong’. For two signals, the health impact differed; the panel judged these signals as ‘strong’, whereas the SIAT categorized it as ‘moderate’.

Copyright © 2016 John Wiley & Sons, Ltd.
In the process of signal selection, many drug-ADR associations will be identified as possible signals, but only some of them will turn out to be real and/or important. The SIAT provides a more standardized insight in the strength of a signal and the impact it may have on the patient’s health. Both aspects are considered to be important when it comes to signal selection. This study explored if the SIAT score provides a proper reflection of the signal strength and health impact by using judgments of a panel of pharmacovigilance experts. The results demonstrated an insufficient validity. In addition, we also explored the reliability of the SIAT which was found to be good.

Discussing the results with the panel of experts revealed that personal experiences and interpretation play a great role in their judgment. An example is the association of salbutamol inhalation and dental carries. One panel expert found that this signal had a ‘strong’ signal strength mainly based on the plausible mechanism. Another expert considered the signal strength ‘moderate’ because despite of the plausible mechanism, the reports scored relatively low on causality. Another example is the association of omeprazole and erectile dysfunction. One panel expert found that the proton pump inhibitor omeprazole has a high number of users in The Netherlands, meaning that there is a high group ‘at risk’ for this possible ADR, which can effect patients’ quality of life. However, because this reaction is not life-threatening and it is reversible, the expert considered that this signal had a ‘minor’ health impact. Another expert considered the health impact ‘major’ because it is used on a large scale in The Netherlands. For this reason, the experts believed that this reaction will have significant health consequences.

These examples demonstrate that the outcome of validity testing may be dependent on the signals that are included. Pharmacovigilance assessors all bring their own expertise based on education, for instance, whether they are medical doctors or pharmacists, and previous work experiences. These differences in expertise and experiences color the way they judge the importance of a signal.

Discussing signals between pharmacovigilance experts is very important, and the final decision on what action to take with a signal should be based on a mutual decision round with multiple experts with varying backgrounds. We believe that the SIAT can be used as an aid in the decision-making process. Its score should not be decisive, but it can be used to make sure all important items, considering that the signals are taken into account.

For this study, we also explored the weighting of the items included in the SIAT. The expert panel judged that there should be a difference in weighting between the items. The panel, for example, judged that the ‘value of the cases’ should have the highest contribution to the score for signal strength. The ‘disproportionality of the drug-ADR association in the database of Lareb, or the WHO on the other hand, should have the lowest contribution. We believe that the difference in weighting is an interesting point of this study, and this could be taken into account when items included in SIAT are used to prioritize signals.

| Table 3. Weighting of the items included in signal strength and health impact |
|--------------------------|--------------------------|
| Signal strength weighting | Health impact weighting  |
| I  | Strength of cases 45 | V  | Seriousness of the association 30 |
| II | Disproportionality in the database of Lareb 10 | VI | Type of drug 20 |
| III | Disproportionality in the database of WHO 10 | VII | Users in the population 25 |
| IV | Literature 35 | VIII | Nature of the indication 25 |

The health impact score is calculated as the sum of the four variables (V + VI + VII + VIII).

**Reliability.** Inter- and intra-observer variabilities were good, $k$ of 0.78 and 0.72, respectively. Most differences were related to items included in signal strength. In-depth evaluation revealed that these differences were caused by differences in scoring for the input ‘literature’. For intra-observer variability, differences were mainly caused by differences in scoring for the input ‘literature’ and in a lesser extent, for the input ‘strength of the cases’. Overall, the SIAT did not consequently score higher or lower for signal strength or health impact compared with the panel judgment.

**DISCUSSION**

In the process of signal selection, many drug-ADR associations will be identified as possible signals, but only some of them will turn out to be real and/or important. The SIAT provides a more standardized insight in the strength of a signal and the impact it may have on the patient’s health. Both aspects are considered to be important when it comes to signal selection. This study explored if the SIAT score provides a proper reflection of the signal strength and health impact by using judgments of a panel of pharmacovigilance experts. The results demonstrated an insufficient validity. In addition, we also explored the reliability of the SIAT which was found to be good.

Discussing the results with the panel of experts revealed that personal experiences and interpretation play a great role in their judgment. An example is the association of salbutamol inhalation and dental carries. One panel expert found that this signal had a ‘strong’ signal strength mainly based on the plausible mechanism. Another expert considered the signal strength ‘moderate’ because despite of the plausible mechanism, the reports scored relatively low on causality. Another example is the association of omeprazole and erectile dysfunction. One panel expert found that the proton pump inhibitor omeprazole has a high number of users in The Netherlands, meaning that there is a high group ‘at risk’ for this possible ADR, which can effect patients’ quality of life. However, because this reaction is not life-threatening and it is reversible, the expert considered that this signal had a ‘minor’ health impact. Another expert considered the health impact ‘major’ because it is used on a large scale in The Netherlands. For this reason, the experts believed that this reaction will have significant health consequences.

These examples demonstrate that the outcome of validity testing may be dependent on the signals that are included. Pharmacovigilance assessors all bring their own expertise based on education, for instance, whether they are medical doctors or pharmacists, and previous work experiences. These differences in expertise and experiences color the way they judge the importance of a signal.

Discussing signals between pharmacovigilance experts is very important, and the final decision on what action to take with a signal should be based on a mutual decision round with multiple experts with varying backgrounds. We believe that the SIAT can be used as an aid in the decision-making process. Its score should not be decisive, but it can be used to make sure all important items, considering that the signals are taken into account.

For this study, we also explored the weighting of the items included in the SIAT. The expert panel judged that there should be a difference in weighting between the items. The panel, for example, judged that the ‘value of the cases’ should have the highest contribution to the score for signal strength. The ‘disproportionality of the drug-ADR association in the database of Lareb, or the WHO on the other hand, should have the lowest contribution. We believe that the difference in weighting is an interesting point of this study, and this could be taken into account when items included in SIAT are used to prioritize signals.
There are, however, some items that are not included in the SIAT but may also be relevant in signal selection. One of them is media attention, because associations discussed by the media are mostly driven by a public safety concern. In order to respond to these concerns, these signals should be taken seriously.

Another item that may be important is the impact of an ADR on the patient’s daily life. The seriousness of the ADR, according to the CIOM’s definition, is included in the SIAT. The more serious the ADR, for example, permanent major disability, the higher the score for health impact. Non-serious ADRs may, however, have a great impact on the patient’s daily life. Because information about the impact of ADRs is not a standard field in the ADR reporting form, it was not possible to include this item in the SIAT. This is, however, worth further investigation.

In addition to validity and reliability testing of the SIAT, we took some efforts for improvement. In the original tool, all inputs have the same weighting. This study showed that the several items are found to be of different important according to the panel of experts. Despite the fact that this study found an insufficient validity, the process of validity and reliability testing provided us new insights. During the process of signal selection, personal experiences and interpretation play a great role. Because of this, other items that are important for signal selection may be overlooked. The SIAT can be used as a guide in this process. Even though we included a limited number of 40 signals, because of the real-life setting with signals from daily practice and views of professionals with experience in pharmacovigilance that were leading, this study provided us a proper reflection of the process of signal selection.

CONCLUSION

The SIAT was found to have an insufficient validity and proper reliability. The insufficient validity may be explained by the fact that personal experiences and interpretation played a great role in the panel judgments. We believe that final decision on what action to take with a signal should be based on a mutual decision round with multiple experts with varying backgrounds. Although SIAT scores should not be decisive in the decision-making process, items included in the SIAT can be used as an aid to decide which signals deserve further investigation.

CONFLICT OF INTEREST

The author declare no conflict of interest.

KEY POINTS

- Both, related to signal strength, and the impact it may have on the patient’s health are important in the process of signal selection.
- Items included in the SIAT should not be decisive but can be used as an aid in signal selection.
- The SIAT was found to have an insufficient validity and proper reliability.

REFERENCES