9 October 2017
EMA/209012/2015 - Track-change version following public consultation

Guideline on good pharmacovigilance practices (GVP)

Module IX Addendum I – Methodological ~~Aspects~~aspects of ~~Signal Detection~~signal detection from ~~Spontaneous Reports~~spontaneous reports of ~~Suspected Adverse Reactions~~suspected adverse reactions

Draft finalised by the Agency in collaboration with Member States	30 June 2016
Draft agreed by the European Risk Management Facilitation Group (ERMS FG)	18 July 2016
Draft adopted by Executive Director	4 August 2016
Released for public consultation	8 August 2016
End of consultation (deadline for comments)	14 October 2016
Revised draft finalised by the Agency in collaboration with Member States	27 September 2017
Revised draft agreed by the EU Network Pharmacovigilance Oversight Group (EU-POG)	4 October 2017
Revised draft adopted by Executive Director as final*	9 October 2017
Date for coming into effect of final version*	22 November 2017

Note: This guidance extends and updates some of the information given in the Guideline on the Use of Statistical Signal Detection Methods in the EudraVigilance Data Analysis System (EMEA/106464/2006 rev. 1) and supersedes the previous advice in the areas addressed by this new guidance.

This track-change version identifies the majority of changes introduced to the public consultation version of this document as the Agency's response to the comments received from the public consultation. This track-change version is published for transparency purposes and must not be taken or quoted as the final version.

* For this reason, the timetable above, and in particular the date of coming into effect, apply only the clean version published as final.

For the final version of this module and any future updates, please see the GVP webpage of the Agency's website.

See websites for contact details

European Medicines Agency www.ema.europa.eu
Heads of Medicines Agencies www.hma.eu

The European Medicines Agency is an agency of the European Union

Table of contents

IX. Add I.1. Introduction

Monitoring of databases of spontaneously reported suspected adverse reactions (in the format of individual case safety reports (ICSRs), see GVP Module VI) is an established method of signal detection. The monitoring process is facilitated by statistical summaries of the information received for each "drug-event" combination over defined time periods. To limit the chances of failing to detect a signal and to ensure that the processes in place are controlled and predictable in terms of resources required, it is recommended that these summaries are produced in a routine periodic fashion. For the same reasons, when possible, the criteria for selecting "drug-event" combinations (DECs) for further investigation should be objectively defined. The aim of this Addendum to GVP Module IX on signal management is to describe components of an effective system for routine scanning of accumulating data focusing on components that have been proved to be effective. It does not give details of particular implementations of such system because these may be influenced by a number of factors that differ between databases. For those interested in the specific implementation developed for use in EudraVigilance other guidance is available (see Screening for Adverse Drug Reactions in EudraVigilance[1]). In common with other GVP documents, the information given herein is guidance on good practice to assist in ensuring compliance with Commission Implementing Regulation (EU) No 520/2012[2]. Other methods may also satisfy this requirement.

This Addendum lists some of the methodological aspects that should be considered in detecting potential signals. The proposed approach complements the classical disproportionality analysis with additional data summaries, based on both statistical and clinical considerations. Although disproportionality methods have been demonstrated to detect many adverse reactions before other currently used methods of signal detection, this is not true for all types of adverse reactions. Hence a comprehensive and efficient routine signal detection system will seek to integrate a number of different methods to prioritise DECs for further evaluation.

The specific details of implementation of the methods proposed may vary depending on, for example, the nature of the medicinal products in the dataset or the rate at which new ICSRs are received. The approaches to signal detection discussed herein have been tested in a number of large and medium sized reporting databases[3] with some variations in performance (see IX. Add I.2.1.2.) noted between databases. Thus, a general principle is that any system of signal detection should be monitored not only for overall effectiveness but for the effectiveness of its components (e.g. statistical methods and specific group analyses).

The decision based on the assessment of the data summaries described herein is whether more detailed review of ICSRs should be undertaken. Such review may then prompt a search for additional data from other pharmacovigilance data sources. The decision process may rely on factors beyond the data summaries, for instance if the suspected adverse reaction is a specific incidence of a class of events already listed in the summary of product characteristics (SmPC). So far as possible the decision process should be formally pre-specified and validated. In each case it should be fully documented.

IX. Add I.1.1. Abbreviations ~~2. Statistical methods~~

ADR	Adverse drug reaction

[1] ~~See www.ema.europa.eu, available as of Q4 2016.~~ See http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000587.jsp.
[2] Commission Implementing Regulation (EU) No 520/2012 Article 19 and 23.
[3] Wisniewski A, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. Good signal detection practices: evidence from IMI-PROTECT. Drug Saf. 2016; 39: 469–490.

DEC	Drug-Event combination
HLT	High-level term (in MedDRA)
ICSR	Individual case safety report
PT	Preferred term (in MedDRA)
SDA	Signal detection algorithm
SDR	Signal of disproportionate reporting
SMQ	Standardised MedDRA query
SOC	System organ class (in MedDRA)

IX. Add I.2. Statistical methods

When the accrual to the ~~dataset~~database is too large to allow individual scrutiny of all incoming ICSRs, it is useful to calculate summary statistics on (subsets of) the data that can help to focus attention on groups of ICSRs containing an adverse reaction. Generally such statistics are used to look for high proportions of a specific adverse event with a given medicinal product, compared to the reporting of this event for all other medicinal products (disproportionate reporting). Sudden temporal changes in frequency of reporting for a given medicinal product may also indicate a change in quality or use of the product with adverse consequences (which could include a reduction in efficacy).

IX. Add I.2.1. Disproportionate reporting

~~*IX. Add I.2.1.1. Components of the statistical signal detection system based on disproportionate reporting*~~

Disproportionality statistics take the form of a ratio of the proportion of spontaneous ICSRs of ~~an~~a specific adverse event with a specific medicinal product to the proportion that would be expected if no association existed between the product and the event. The calculation of the expected value is based on ICSRs that do not contain the specific product and it is assumed that these ICSRs contain a diverse selection of products most of which will not be associated with the ~~adverse~~ event. Hence the reporting proportions for events in these ICSRs will reflect the background incidence of the event in patients receiving any medicine. There are a number of different ways to calculate such statistics and this choice is the first step involved in designing a statistical signal detection system.

When an adverse event is caused by a medicine, it is reasonable to assume that it will be reported more often (above the reporting rate associated with the background incidence~~)~~), and hence ~~this~~the reporting ratio will tend to be greater than one. Thus high values of the ratio for a given DEC suggest further investigation may be appropriate. In practice a formal set of rules, or signal detection algorithm (SDA) is adopted. This usually takes the form of specified thresholds that the ratio or other statistics must exceed, but more complex conditions may also be used. When these rules are satisfied for a given DEC, it is called a signal of disproportionate reporting (SDR). Then a decision needs to be made regarding whether further investigation is required.

A further decision needs to be taken as to whether the statistics are to be calculated across the whole database or if modifications based on subgrouping variables would be of value. While the decision is motivated by theoretical consideration, the specific choice of whether to use subgroups and, if so, which to use, should be based on empirical assessment of signal detection performance. In particular

the impact on the false positive rate should be considered. Whether the database is sufficiently large to avoid very low case counts within subgroups may also be a factor in this decision.

IX. Add I.2.1.~~2~~1. Considerations related to performance of ~~statistical~~ signal detection systems

The performance of signal detection systems, including the SDA, can be quantified using three parameters that reflect the intended objective of the system. Desirable properties are:

1. high sensitivity (the proportion of adverse reactions for which the system produces SDRs);

2. high positive predictive value or precision (the proportion of SDRs that relate to adverse reactions);

3. short time to generate ~~SDR~~SDRs (that can be assessed from a chosen time origin, possibly the granting of a marketing authorisation to the first occurrence of an SDR for an adverse reaction).

Estimates of these performance parameters depend on the particular reference set[4] of known adverse reactions selected for their evaluation and are also not fixed because spontaneous reports accumulate over time. They are thus best used as relative measures for comparing competing methods of signal detection within the same spontaneous reporting system at the same point in time.

The following factors may affect the performance of signal detection systems:

- MedDRA hierarchy

A precondition for automated screening of DECs for adverse reactions is the availability of schemes for classifying adverse events and medicinal products. The nature and granularity of these schemes affects the performance of the screening. MedDRA (see GVP Annex IV), used for reporting suspected adverse reactions for regulatory purposes, ~~classifies~~ provides terms for adverse events and classifies them in a multi-axial hierarchical structure and a choice must be made whether to screen at one level of granularity (e.g. SOC, HLT, PT) or several and whether to include all terms or only a subset. Screening at the second finest level of granularity, i.e. Preferred Term (PT), has been shown to be a good choice in terms of sensitivity and positive predictive value[5].

Finally, focus of statistical signal detection on ~~to~~ adverse events considered ~~most~~ clinically most important avoids time spent in assessments that are less likely to benefit patient and public health. A subset of MedDRA terms judged to be important medical events (IMEs[6]) is thus considered a useful tool in statistical signal detection when filtering results for medical review.

The remarks above relate to routine signal detection and not to targeted monitoring of potential risks associated with specific products where ad hoc use of other levels of MedDRA terms may be appropriate. In addition, although no formally defined MedDRA term subgroups (e.g. HLT, SMQ) have proven better for signal detection than the PTs, some of them are effectively synonymous. The definition of a synonym in this context is the pragmatic one, i.e. that two PTs are considered synonyms if it is reasonable to suppose that a primary reporter of a suspected adverse reaction, presented with a single patient and without a specialist evaluation, would not necessarily be able to decide which term to use. It may also be appropriate to combine such terms when they relate to identified areas of interest.

- Thresholds

[4] ~~Further guidance to be finalised in a separate document in Q4 2016.~~ 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: 577–587.

[5] Hill R, Hopstadius J, Lerch M, Noren ~~G.N.~~GN An attempt to expedite signal detection by grouping related adverse reaction terms. Drug Saf. 2012; 35:1194~~–~~-1195.

[6] http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/q_and_a/q_and_a_detail_000166.jsp&mid=WC0b01ac0580a68f78.com

The SDA applied to the summary statistics for each DEC usually takes the form of a set of threshold values such that SDRs occur only if each statistic exceeds its corresponding threshold. Very low thresholds will result in large, and potentially unmanageable, numbers of SDRs to investigate with a higher probability of being false. This will also reduce the resources available for assessment of true SDRs. Too high thresholds will result in identification of adverse reactions being delayed or even entirely prevented. Thus the appropriate choice of thresholds is fundamental to the success of the statistical signal detection system.

This has also been confirmed by studies comparing different disproportionality methods and different sets of threshold showing that the former can achieve similar overall performance by choice of appropriate SDA. Therefore, in contrast to the choice of disproportionality statistic, it is the choice of SDA to define a SDR that will strongly influence signal detection performance[7].[4].

Thresholds for disproportionality methods are usually based on two separate indicators, one reflecting the disproportionality statistic itself and another based on the number of ICSRs received. For the formerFor a reason discussed later, limiting false positives is better handled by raising the threshold for the number of ICSRs than that for the disproportionality statistic. For the disproportionality statistic, in practice, rather than the point estimate, a formal lower confidence bound is often used. The rationale for its use is that when the statistic is based on few ICSRs, it falls further below the point estimate and makes an SDR less likely. Hence, this is an intuitive way of incorporating into the signal detection process the degree of confidence about the reliability of the data. It has also been shown that a threshold based on the lower confidence bound performed better alone than with an additional threshold for the absolute value of the disproportionality statistic itself.[5]4.

In addition, it has been shown that a correlation exists between the value of a disproportionality statistic and the relative risk of an adverse reaction when exposed to the medicinal product estimated in epidemiological studies[8], therefore setting any threshold on the lower confidence bound of the disproportionate statistic above 1 might lead to missing an adverse reaction for which the risk ratio is not great enoughhigh.

Finally, there appears to be a reduction in positive predictive value with a medicinal product's time on the market, hence it might be more efficient to vary the amount of effort to invest in signal detection over the life-cycle of the product. This might involve the use of differing thresholds to define an SDR depending on the time of the product on the market.[5]4.

- Periodicity of monitoring

A one-month interval between consecutive data summaries has been investigated in validation studies for signal detection methods. More frequent monitoring has also been used, for instance for medicinal products under additional monitoring or during intensive vaccination programmes. The appropriate frequency of monitoring may vary with the accumulation of knowledge of the risk profile of a specific active substance/medicinal product (see GVP Module IX.C.2.)).

- Spontaneous ICSR databases

The performance has also been shown to depend on the nature of the spontaneous ICSR database and this appears to be related to the mixrange of medicinal products included in the database.[4].

[7] 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: 577–587.

[8] Maciá-Martínez M,Abajo FJ De, Roberts G, Macia M, Slattery J, Thakrar B, Wisniewski AFZ. An empirical approach to explore the relationship between measures of disproportionate reporting and relative risks from analytical studies. Drug Saf. 2016; 39: 29-43.

An important inference from these considerations is that organisations doing signal detection should assess the performance of a signal detection system directly on the database to which it will be applied. This will allow the ability to detect new adverse reactions and the work load involved to be predicted and controlled by appropriate changes to the SDA. As databases evolve in terms of numbers of ICSRs included and their mix of medicinal products, periodic reassessment of performance should be undertaken.

- Subgroup analysis and stratification

Spontaneous ICSR databases cover a range of medicinal products with different indications and that are used across a broad range of patient populations. Also, ICSR reporting patterns vary over time and between different geographical regions. Many quantitative signal detection algorithms disregard this diversity which may result in an SDR either being masked or ~~a false~~in an association being incorrectly flagged as a signal.

Stratification and subgroup analysis are generally used in epidemiology to reduce bias due to confounding and may also have advantages in statistical signal detection. By subgroup signal detection here is meant analyses carried out to detect ~~SDRs~~ADRs within specific ICSR subgroups~~.~~ e.g. by indication, age group, region or time period. Stratification involves combining results from within different ~~subgroup~~subgroups to obtain an adjusted result for the whole dataset.

The comparison of stratified versus subgroup analysis has shown that the ~~latter~~subgroup analysis consistently performed better ~~than the former~~. Moreover, subgroup analysis has also shown to provide clear benefits in both sensitivity and precision over crude analyses for large international databases[9]. However, such benefits may not be obtained in small databases.

Subgrouping variables that showed the most promising results included age and reporting region/country, but it is likely that choice of variables for subgroup analyses varies according to the database.

IX. Add I.2.2. Increased ICSR reporting frequency

Most routine signal detection is aimed at identifying unknown, potentially causal associations between medicinal products and adverse events that are assumed ~~likely~~ to ~~result in a~~be constant ~~or slowly changing reporting rate~~ over time. However, some causal associations of medicinal products with events of interest in the context of pharmacovigilance may show a marked temporal variation. Examples are manufacturing quality issues, a developing culture of abuse, evolving antimicrobial resistance or changes in the use of the product and, in particular, new off-label use. One way of detecting signals associated with such events, that may add value to simple disproportionality methods, is to monitor changes in the frequency of overall reporting for the products.

However, changes of reporting frequency are also expected that do not reflect ~~the~~new safety issues of the medicinal products. These may result from rapid increases in use when the product is first marketed or new indications are authorised, ~~publicity associated with unfounded safety concerns,~~ sudden changes in exposure (e.g. seasonal use of vaccines), publicity associated with unfounded safety concerns, reporting promoted by patient support schemes not clearly labelled as studies, clusters of ICSRs reported in the scientific literature ~~reports~~ or duplicated ICSR reports.

There are several options for detecting temporal changes in reporting frequency. The simplest method examines the changes in the number of ICSRs received per product over a fixed time period as an

[9] Seabroke S, Candore G, Juhlin K, Quarcoo N, Wisniewski A, Arani R, et al. Performance of stratified and subgrouped disproportionality analyses in spontaneous databases. Drug Saf. 2016; 39: 355-364.

absolute count. Statistical tests compare recent counts with the latest count, testing for significant increases. Similar methods can be used at the DEC level and, for these, relative values compared to the total ICSR count for the product may be considered as an alternative to absolute counts. The method disregards however quantitative changes in exposure, which would impact on the frequency of adverse reactions.

Another option is to consider changes in the disproportionality statistics over time. This approach ~~would be~~is less susceptible to increase in number of ICSRs triggered by effects related to the product rather than a specific adverse event~~. For~~ - for example general publicity about the product, stimulated reporting (see GVP Annex I) or changes in exposure~~;~~ - however, results will still be influenced by the background distribution in the rest of the database and not only by changes in reporting frequency for the specific medicinal product. In addition, results might be less reactive to transient temporal variations since the focus is on changes in statistics based on the cumulative count, not in comparing recent counts with the latest count. This problem will be more pronounced when large numbers of cases have accumulated, as proportional changes will then be smaller.

Limited work has been performed to assess the effectiveness of these methods even if theoretically they seem appealing. Thus these methods might be implemented with ongoing quality control measures to ensure acceptable performance.

IX. Add I.3. Methods aimed at specific groups of adverse events

IX. Add I.3.1. Designated medical events

Some medical events are known to result on most occasions from exposure to medicines. Thus, when such events are reported, the prior probability of a causal relationship to one of the ~~medicines~~medicinal products listed in the ~~report~~ICSR is high. Hence the ICSRs will evoke concerns even before an SDR is observed. A list of these terms, complemented by important and serious events that should not be missed, should then be created and can serve as a safety net in signal detection. It is recommended that these designated medical events (DME) are drawn to the attention of signal detection assessors irrespective of any other statistical methods used and that they are prioritised for clinical review. Elements of the DME list are generally a relatively small subset of the IME list.

The list of DME should also be ~~occasionally~~periodically reviewed ~~and revised~~ based on experience gained and performance.

IX. Add I.3.2. Serious events

The seriousness of events described in spontaneous ICSRs does not obviously relate to the probability that they are causally related with the medicine ~~related~~. However, it may impact the patient and public health importance should they later prove to be related. This reason is a rationale for prioritising assessment of serious events. Complementary to the creation of a list of DMEs and in addition to the use of lists of IMEs, a simple approach to such prioritisation is to highlight new ICSRs in which a death is reported and to give separate counts of those ICSRs for each DEC. It should be appreciated that this may be a rather imprecise criterion and prioritising all ICSRs with reported death may result in many false positive signals. Hence it is considered that further methodological research may be required in this area.

IX. Add I.4. Methods aimed at specific patient populations

When ICSR databases are sufficiently large, some ~~classes~~group of patients may be identified that merit separate attention in signal detection due to known or suspected systematic differences in their responses to medicines. Two such groups that can be differentiated in most databases are the youngest and oldest patients.

A caveat relevant to analyses restricted to any ~~subgroup~~subset of spontaneous ICSRs is that homogeneity of adverse events may be increased resulting in greater potential for masking of signals. For example, analyses within a group of patients who are the main recipients of a class of medicines may not highlight effects related to the entire class. A possible solution is to monitor specific patient ~~populations~~groups in parallel ~~with~~to analyses of the total ~~dataset~~population.

IX. Add I.4.1. Paediatric ~~populations~~population

Often a single paediatric group is chosen below a selected age threshold. Although childhood is a period of rapid change and no threshold is likely to define a homogenous group, this succeeds in defining a population with marked developmental, physiological and psychological differences from adults~~.~~ (see GVP P.IV).

Separate presentation of suspected adverse reactions that ~~occur~~are detected in the paediatric population and use of both clinical and statistical methods ~~seems~~seem to be justified to improve the detection of signals ~~in~~for the paediatric population. ~~In line with the general population,~~ ~~statistical~~Statistical disproportionality tools should be applied to ICSRs ~~relating to the use of medicines in~~reported for children to increase the ability to detect signals in the paediatric population from spontaneous ICSR databases. Within-group disproportionality statistics that are significantly higher than those in the non-paediatric group should be highlighted for additional consideration[10]. Additionally, given the lower number of ICSRs usually received for the paediatric population compared to the ~~rest of the~~adult population, it is recommended to use a lower thresholds based on the number of ICSRs received.

An additional aid to focusing on paediatric safety issues can be provided by a list of adverse events (a targeted medical events list) that tend to have more serious outcomes in children than adults[~~11~~]. This list should be used to reduce missed signals that are more clinically relevant in the paediatric population, otherwise not flagged by other methods. More extensive discussion of pharmacovigilance in the paediatric population ~~will be~~are available in ~~the revised Guideline~~GVP P.IV. on ~~Conduct of Pharmacovigilance for Medicines Used by the Paediatric Population~~[~~12~~]paediatric pharmacovigilance. The age threshold for paediatric signal detection should be chosen to align with the upper age limit from this guideline.

IX. Add I.4.2. Geriatric ~~populations~~population

Specific signal detection measures aimed at older recipients of medicines are a reasonable precaution given the high frequency of concomitant use of multiple medicines and the possibility of impaired physiological elimination mechanisms~~.~~ (see GVP P.V).

The age threshold at which such measured should be implemented has not been clearly established. Although the proportion of patients for whom suspected adverse reactions are reported increases with

[10] Blake KV, Saint-Raymond A, Zaccaria C, Domergue F, Pelle B, Slattery J. Enhanced paediatric pharmacovigilance at the European Medicines Agency: a novel query applied to adverse drug reaction reports. Pediatr Drugs. 2016; 18: 55-63.

[~~11~~] ~~Further guidance to be finalised in a separate document in Q4 2016.~~

[~~12~~] ~~Currently under review; to be finalised in 2016-2017.~~

age, some research has suggested that this can be explained by more common use of medicines[13]. Thus it may be better to choose a threshold based on increased exposure rather than possible increased susceptibility. Alternatively, a consistent approach is to use the same age group in routine signal detection as selected for other pharmacovigilance activities~~. In this respect refer to GVP P IV: Geriatric population.~~ (see GVP P. V.).

For routine signal detection processes it is recommended that ICSRs from patients above the chosen age threshold should be clearly identified and that, as for the paediatric population, within-group disproportionality statistics that are significantly higher than those ~~in~~within the non-geriatric group should be highlighted for additional consideration.

IX. Add I.5. Methods aimed at ~~underlying causal processes~~specific circumstances of medicines use

In addition to the description of the clinical manifestation of the suspected adverse reaction, ICSRs may include information on the ~~potential causal mechanisms for the reaction. Such information may relate to the~~ circumstances of medicine use which could have contributed to the occurrence of the adverse reaction, e.g. abuse, misuse, overdose, medication error or occupational exposure~~.~~ (see GVP Annex I).

~~*IX. Add I.5.1. Abuse, misuse, overdose, medication error or occupational exposure*~~

Although the coding of these circumstances is enabled as Preferred Terms in MedDRA (see GVP Annex IV), they are qualitatively different from the clinical circumstances which are the focus of disproportionality-based signal detection. Firstly, they are manifestly related to at least one medicinal product identified in the ICSR. With suspected adverse reactions in normal circumstances of use this relationship is a matter of clinical judgement. Secondly, the circumstances described by each of these terms differ depending on the product concerned. Hence between-medicine comparisons of reporting frequency of ICSRs with MedDRA-codes describing these circumstances are both unnecessary and potentially misleading.

However, knowledge of these circumstances can appreciably alter the assessment of causality when reviewing a potential signal. Thus, it is recommended that the numbers of ICSRs with the respective MedDRA codes should be displayed for each DEC in signal detection listings.

[13] Begaud B, Martin K, Fourrier ~~a~~A, Haramburu F. Does age increase the risk of adverse drug reactions? Br J Clin Pharmacol. 2002; 54: 550–552.