European Medicines Agency
Post-authorisation Evaluation of Medicines for Human Use

London, 26 June 2008
Doc. Ref. EMEA/252216/2007

OVERVIEW OF COMMENTS RECEIVED ON DRAFT GUIDELINE ON THE USE OF STATISTICAL SIGNAL DETECTION METHODS IN THE EUDRAVIGILANCE DATA ANALYSIS SYSTEM (EMEA/106464/2006)

Table 1: Organisations that commented on the draft Guideline as released for consultation

	Name of Organisation or individual	Country
1	European Generic Medicines Association	N/A
2	Merck Sharp & Dohme (Europe) Inc.	N/A
3	ProSanos (Harrisburg, PA 17101)	USA
4	Roche Products Limited	
5	Lilly Deutschland GmbH	Germany
6	European Federation of Statisticians in the Pharmaceutical Industry	N/A
7	EFPIA	France
8	Novo Nordisk A/S	Denmark
9	Medical Governance & Pharmacovigilance	N/A
10	Pfizer Ltd	UK
11	GE Healthcare Ltd.	UK

7 Westferry Circus, Canary Wharf, London, E14 4HB, UK
Tel. (44-20) 74 18 84 00 Fax (44-20)
E-mail: mail@emea.europa.eu http://www.emea.europa.eu

Table 2: Discussion of comments

GENERAL COMMENTS - OVERVIEW
Appropriate and necessary cautions There are appropriate and necessary cautions in sections 1, 4(d), 6, 7 on the interpretation of the methods described. However, these omit some important additional caveats of the method proposed, for example, stronger recognition of the need to consider bias implications from the use of observational databases. If background information (e.g., severity of condition) will not be available from the database, such confounding factors will unknowingly influence the proposed signal detection procedure. It would also be valuable to include a discussion about the appropriateness of the assumptions required (e.g. a repeatable sampling mechanism and asymptotic arguments) in the context of SRS data. There are obviously significant problems in utilising inferential statistical methods in an exploratory analysis and we feel that the document needs to address the issues that this raises, particularly the possibility that statistical validity may wrongly be implied for purely speculative analyses. Some specific comments are made in the relevant sections, but, given the number of available choices (none of which are convincingly validated) for designing a data mining analysis, we would particularly warn of the dangers of 'retro-fitting' an analysis such that it fulfils pre-existing expectations. Outcome: The pharmacovigilance data mining on SRS databases is exploratory in essence on a dataset which is subject to important confounding (confounding by indication or due to the underlying disease). In addition, other factors will dramatically influence the outcome of the analysis (data quality, coding practices, information on medicinal products) therefore these elements as well as the statistical limitations of the data mining algorithms used in these exploratory analysis must be borne in mind when interpreting the results obtained with the data mining algorithm. Some of these considerations have been highlighted under section 4.5 interpretation of SDRs. The systematic medical assessment of the SDRs is aimed at assessing the causal relationship between the administration of the medicinal product and the occurrence of the reaction.
I would like to express some concern over the statement in Section 4.2 that "concomitant medication[s] are not normally taken into account in the [PRR] calculations". We studied the question of whether reporters of adverse events can accurately assign blame to the correct medicinal product by designating it as "suspect" rather than "concomitant". We found that the accuracy of this designation varies substantially from product to product. A paper on this topic has been submitted for publication. The Working Group may want to reconsider the role of medications designated as "concomitant". There is a possibility that their elimination could inhibit the detection of signals for highly unexpected adverse events. Outcome: The EudraVigilance Data Analysis System has been modified in such a way that concomitant medications may be taken into account in the computation of the PRR.

Analysis Data Set Considerations about "analysis data set" (i.e. number of cases, possible subset of data, types of data, data quality and completeness, duplications) are reported in different parts of the text. A specific paragraph at the beginning of Section 4 could be set up. Outcome: Considering that the dataset is changing constantly (which would require a periodic update of the guideline with limited added value) and considering that it would make the guideline much more difficult to read, no detailed description of the dataset has been included in the guideline. The data set can be accessed directly with the EudraVigilance Data Analysis System.
Interpretation It is not stated in the document that the presented techniques and methods (in case of a SDR) can or should also applied to other databases of ICSRs like the one from FDA or the company database and that signals may be put in perspective by output from different safety databases. Also no guidance is provided on interpretation of the results, obtained by the application of different methods in specific situations, in particular when the expected number of cases of drug combinations is small. Outcome: The scope of the guideline is the use of quantitative methods in the EudraVigilance Data Analysis System. The guideline does not apply to any other spontaneous reporting system databases or methods.
In general, the guideline is helpful and the concept behind it is sound. Despite some limitations, this is an excellent initiative and a good start to standardize signal detection methods. Outcome: N/A
The objective(s) of this guideline need to be specified i.e. what does it mean to achieve? For instance, it is not immediately clear who (EMEA, member state pharmacovigilance offices, or other party) is expected to conduct routine or ad hoc data mining runs and a subsequent review of any signals of disproportionate reporting (SDRs). According to Diagram 1 (page 18), the proposed SDR review process appears to be directed towards the competent authorities and involves the MAH at a late stage. Is this the correct intention? Outcome: Access policies to EudraVigilance are being developed. The guideline will be changed (and will make reference to these access policies) when these access policies will be implemented.
The guideline provides a detailed description on the application of a specific statistical method to the analysis of pharmacovigilance data. It recognizes the challenges in interpreting the results originating from statistical calculations of data that are generally poor in quality. The emphasis on comprehensive medical assessment on a statistical signal is rightly evident. However, it could be made clearer that a statistical association does not imply a causal relationship between the administration of the drug and the occurrence of the adverse event. The guideline should state very explicitly that 'positive' PRRs should not be interpreted as defined risk – it should be clear that PRRs are only used for signal detection purposes and that this guideline does not give guidance on signal evaluation: PRRs are simply indicators of a possible risk. In addition, it should consider whether comparisons of PRRs between different drugs might be problematic. Outcome: The fact that a statistical association does not imply a causal relationship between the administration of the drug and the occurrence of the adverse event was already mentioned in the guideline and has been highlighted.

We commend the Expert Working Group for producing such an eloquent and well thought out system for signal detection in the Eudravigilance database. As a result we only have minor and mainly aesthetic comments. Outcome: N/A
A description of the EudraVigilance database and how it is used for signal detection purposes would help place this guideline into context. For example: - are 'as reported terms' or only MedDRA coded terms accessible? - is there a routine MedDRA code refresh to all data? - is this based on the 'as reported terms?etc. Outcome: The disproportionality analysis is conducted on the MedDRA preferred terms as they were transmitted to EudraVigilance by the sender (or the primary receiver of the case). MedDRA is the medical terminology implemented in EudraVigilance. All the information concerning the use of MedDRA in EudraVigilance, the periodicity of the updates and the coding practices may be found on the EudraVigilance website and the MedDRA MSSO websites http://eudravigilance.emea.europa.eu/human/meddra01.asp and http://www.meddramsso.com/MSSOWeb/index.htm.
Generally, the system is proposing to fully adopt the methodology proposed by Steven Evans and presented at a number of meetings over the last few years, i.e. the PPR. While a useful way of summarising data from large databases, it does collapse all other drugs in a database into one comparator group. The reasons for this from a statistical viewpoint are reasonably clear, in that it avoids the problem of sparse cell distribution and also ensures that a "meaningful" statistical analysis can be performed. What it does not do, however, is allow for comparisons of drugs against those in the same class or family, which may be a useful analysis in its own right. Outcome: subgroup analyses may be performed (see section 4.7 subgroup analyses and stratification). Further subgroup analyses functionalities (such as a subgrouping by anatomo-therapeutic class) will be implemented in the second phase of the release of the EV DAS.

General considerations:
The guideline is an excellent initiate, which may push the area of pharmacovigilance in the right direction. In addition, to traditional pharmacovigilance methods, statistical methods for signal detection are expected to be valuable for detection of possible signals at an earlier stage.

Knowledge of data and the nature of the database are of high importance, when using statistical methods for signal detection. More in depth information on the Eudravigilance database and data entered is thus needed, when calculating and analysing SDR (Signals of Disproportionate Reporting). A detailed description of the Eudravigilance database and how to use data is needed.
It should be outlined which events and products are included and entered in the database and by whom.
A description of the overall structure and content of the database (variable and table description) should be included.

As quality of data is of high importance for reliability of calculated statistical values, the quality of data in the Eudravigilance database should be described. In case data has been cleaned to improve quality of data (e.g. identification of duplicates, names of products etc), a description of how data has been cleaned, should be included.

In case data in the Eudravigilance database are useful for analysis of combination treatment, it would be nice to include a section in the guideline, describing how to perform such analysis.

Outcome: A detailed description of the duplicate detection algorithm and of the recoding practices would make the guideline difficult to read and understand. These issues are addressed in the training which is offered to the users of the EV DAS. The analyses can be performed with or without the identified duplicates. Considering that the dataset is changing constantly (which would require a periodic update of the guideline with limited added value) and considering that it would make the guideline much more difficult to read, no detailed description of the dataset has been included in the guideline. The data set can be accessed directly with the EudraVigilance Data Analysis System.

Addition of statistical references
It would be very helpful if references were provided routinely for each of the statistical statements made in the document, including a discussion about the appropriateness of the assumptions required (e.g. a repeatable sampling mechanism and asymptotic arguments) in the context of SRS data. This would help to highlight the cases where these should be regarded as exploratory techniques.
There are obviously significant problems in utilising inferential statistical methods in an exploratory analysis and we feel that the document needs to address the issues that this raises, particularly the possibility that statistical validity may wrongly be implied for purely speculative analyses. Some specific comments are made in the relevant sections, but, given the number of available choices (none of which are convincingly validated) for designing a data mining analysis, we would particularly warn of the dangers of 'retro-fitting' an analysis such that it fulfills pre-existing expectations.

Outcome: Additional relevant statistical references have been added to the guideline.

Which linked guidelines are already in place; which are to be expected?

Outcome: The legal basis makes a cross-reference to the implementing guidelines of Council Directive 2001/83/EC on the Community code relating to medicinal products for human use and Regulation (EC) No 726/2004 laying down the Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency i.e. Volume 9A of the Rules Governing Medicinal Products in the European Union.

Independent of the methods used for disproportionality analysis, there are many issues that needed to be considered when using spontaneous adverse event data. The adverse event databases contain voluntary reports from patients and healthcare professionals where there may be suspicions that one or more drug that the patient is taking may have caused one or more unintended effects. The voluntary nature of the cases creates inherent limitations in the data interpretation. While each case may represent one or more drug-event pairs of interest, there is no true denominator to define a true reporting rate. Moreover, the reported adverse events can not be compared to either the drug utilization rate, or the background disease rate in the population. The true population itself is undefined, so basic comparisons between cohorts within the overall reporting sample can not be adequately adjusted for relevant confounding factors. Of particular note for the EMEA is that population differences across participating countries can not be accounted for. While some disproportionality analysis metrics use stratification as a means to perform some adjustment on a limited set of factors (e.g. MGPS stratifies by age, gender and year of report), many reports do not provide sufficient information about the relevant covariates to adjust even if the method allowed for correction. Another limitation with spontaneous adverse event reporting system is the quality of the data; cases can be redundantly reported and there are challenges with the coding of drugs and conditions. Many organizations, like EMEA, use MedDRA for classifying adverse events, but cases can be misclassified or preferred terms can be aggregated into term groups that are not clinically meaningful. Outcome: The limitations of the spontaneous reporting system databases are highlighted in the guideline. To avoid any misinterpretation of the comment, it is noteworthy to highlight that stratification can be perform with other quantitative methods such as PRR, RRR, OR or BCPNN.
There are several threats to internal validity within spontaneous adverse event reporting data. "History" is a threat when an observed effect might be due to an event which takes place between the pretest and posttest[9]. In the case of adverse event reports, increased media coverage poses a direct 'history' threat; as health professionals become more aware of potential concerns, they become more likely to report those events (independent of how often they actually observe the drug-event relation). Maturation bias is another related threat, when the observed effect might be due to the respondent's growing older, wiser, and more experienced. In our case, it is quite often the case that drug-event associations are not commonly established until knowledge of the potential signal in available. That is, physicians may be less likely to report a potential association between a drug and condition, if he knows no relation between the two, until he observes more instances of the association (at which time, his reporting rate will increase). This is a primary argument for considering concomitant medications, in addition to 'suspect' or 'interacting' drugs in disproportionality analysis; maturation bias can lead to misclassification of the drugs within the adverse event report. The most significant threat to internal validity is selection bias- when the effect may be due to differences between the kinds of people in one group as opposed to another. Here, there is clear reporting bias that is completely uncontrolled: patients and health professionals self-select as to whether they will report any cases, as well as self-select which cases they chose to report. It is well accepted that there is a significant amount of underreporting, but it is not understood the degree to which underreporting varies by drug or condition. Parks et al[10] demonstrated the impact that underreporting can have on different disproportionality metrics. Because of threats due to history, maturation, and selection (as well as the interactions between them), spontaneous adverse event data has low internal validity. Outcome: The limitations of spontaneous reporting databases have been discussed in the guideline.

<<name>> is developing methods to leverage observational data, such as insurance claims and electronic medical records, for its pharmacovigilance activities. Proper observational study can eliminate much of the limitations associated with disproportionality analysis on spontaneous adverse event data. In particular, observational data provide a large, defined cohort from the real-world population from which to sample subgroups and conduct analyses to base meaningful comparisons. In general, these data sources provide a high degree of detail about the patients, such as demographic factors, comorbidities and concomitant medications, and also provide significant exposure time and long follow-up to enable thorough longitudinal analyses. The data quality enables methods to be used with strong precision, while the quantity and distribution of the sources provide good external validity. More generally, observational data provides a comprehensive patient sample to act as a true denominator to calculate background disease rates and medication usage. Such information can facilitate the identification of true adverse events by assessing the temporal associations between drugs and conditions and monitoring how these background incidence rates are affected by the introduction of new medicines within specific patient populations. <<name>> has developed an independent database matched cohorts comparable group design that facilitates both the screening and evaluation of safety signals through the integrated use of disparate data sources and methods. Such an approach holds great promise for the industry as a basis of supplementing current pharmacovigilance activities as well as addressing many of the ongoing challenges associated with both disproportionality analysis and spontaneous adverse event reporting data more generally.

Outcome: The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment). Some pharmacoepidemiology studies will also be conducted with the use of another health care database and EudraVigilance. We acknowledge that this may be useful however the performances of this approach are not defined at this point in time.

ACCESS POLICIES, ROLES AND RESPONSIBILITIES

There seems to be a slight disconnect in this document between market authorisation holders and the use of this tool. It is not clear whether this data will be presented to companies, or whether companies will have access to the tool. We would like this to be further clarified.

It is not clear to whom this guidance document is directed. In the Executive Summary the "stakeholders" are mentioned as the analyzers of adverse event data but it is not clear who they are. Footnote 1 on page 3/21 does not give any further indication. To the best of our knowledge there is no access to the data contained in the EudraVigilance Data Analysis System by external parties (like MAH's or academic organizations).
In general, the draft guideline rather looks like an internal guidance for EMEA co-workers on how to use the EudraVigilance Data Analysis System properly. This notion seems to be supported by the fact that more or less the complete chapter 5 of the draft guideline reads like an extended user manual to this EudraVigilance Data Analysis System. Moreover, the EudraVigilance Data Analysis System seems to be already operational (see last sentence first para in Executive Summary, page 1 and e.g. large parts of chapter 5). If the draft guideline is thought for internal use, we wonder why it is disseminated as an EMEA guidance document to the public.
So according to the real addressee of this document, either chapter 5 may be dropped – or chapter 5 remains and the whole document may be turned into an internal (EMEA) user manual of the EudraVigilance Data Analysis System.

Use of the document

It is not clear to whom this guidance document is directed. In the Executive Summary the "stakeholders" are mentioned as the analyzers of adverse event data but it is not clear who they are. Footnote 1 on page 3/21 does not give any further indication. In section 8, 1st paragraph it is referred to the risk management plan of Applicants or MAHs, so this relates to industry as user. However, in the Executive Summary, 2nd paragraph, 1st line it reads that new SDRs are screened regularly and in section 5, 1st bullet that data summaries are generated periodically. Are these two latter activities mandatory part of any PV plan or is it here referred to use by EMEA itself? Is there a recommendation on how often this process will be performed?

Regarding stakeholders' access to EudraVigilance Analysis System, it is not clear what intention the EMEA has to make data mining statistics available to parties other than the regulatory agencies and MAHs, e.g. healthcare professionals and the general public (including law firms).

EFPIA has significant concerns if it is intended to provide unrestricted access to the EudraVigilance analysis tools and statistical disproportionality reports. Whilst we support the principle of transparent public access to important safety information, the value of allowing public access to PRR Static and Dynamic Reports is highly questionable as a communication tool, particularly as it will be open to significant misinterpretation. Furthermore, it is quite possible that data mining statistics could be considered inappropriately as being scientifically rigorous even though the underlying adverse drug reaction data and statistical methods have known limitations. Signals of disproportionate reporting should always be medically assessed, with completion of reviewing steps and documentation as outlined in Section 7.1 of the guideline. We therefore urge that any outputs from the system are tailored to the needs of stakeholders with the appropriate context (e.g. Caveats Document) to facilitate understanding.

It should also be emphasized that the data mining techniques described in this guideline are simply one part in a larger context of overall pharmacovigilance programs and would be no substitute for a well-designed study with an analytic design (e.g. clinical trials and epidemiological studies).

Please refer in the guideline to the following questions/issues:

1. Who are the expected users of the EudraVigilance Analysis System? If different user groups are covered by this guideline, what are the scopes of these user groups?
2. Please provide a clear distinction between the general description of signal detection methods and a specific description of EudraVigilance capabilities.
3. Who is generating SDRs? Is this EMEA only, other competent authorities, the MAHs, etc?
4. Is there a routine signal detection process planned? All SDRs, DMEs, TMEs? In what time frame?
5. For which products is the signal detection process intended?
6. Please describe the signal detection and management process in more detail. Show responsibilities and timelines. Who is responsible for further analyses?
7. How is the communication planned between EMEA and MAHs?
8. What are the expectations of the MAHs by the EMEA?

Will all user groups, including the MAH's have access to the data as analyzed with the EudraVigilance Data Analysis System?

Outcome: Many of these issues relate to the roles and responsibilities for the conduct of Pharmacovigilance activities in the EU. These aspects are or will be further detailed in the community legislation and guidelines. Finally, some of these issues are addressed in the training given to the EV DAS users.

According to section 5.1, the Reaction Monitoring Weekly/Monthly report will be generated at defined intervals. It is not clear which events, drugs, or drug-event combinations these reports will be based on. Will all the new cases received during the period covered be included in the reports?

Furthermore, the guideline should indicate which parties (MAHs, regulatory agencies, or other institutions) these reports will be generated for. Finally, after receiving the reports and/or if there are pre-defined signals detected in the reports, it is not clear what follow-up measures these parties should take.

Outcome: the guideline specifies that the table includes all the ICSRs received in EudraVigilance during the period covered by the Reaction Monitoring Weekly/Monthly reports.

Outcome of these comments on the access to EudraVigilance/definition of the stakeholders: Access policies to EudraVigilance are being developed. The guideline will be changed (and will make reference to these access policies) when these access policies will be implemented.

INTEGRATION WITH OTHER OF PHARMACOVIGILANCE ACTIVITIES

In general, the guideline provides a detailed description on the application of a specific statistical method to the analysis of pharmacovigilance data. It recognizes the challenges in interpreting the results originating from such statistical calculations of the data that are generally poor in quality. The emphasis on comprehensive medical assessment on a statistical signal is evident. It could be made clearer that a statistical association does not imply any kind of causal relationship between the administration of the drug and the occurrence of the adverse event. The guideline says little about the stakeholders and their authority, responsibility, and accountability in the use of statistical signal detection methods. It is unclear why the proportional reporting ratio (PRR) is specified for implementation, raising some concerns that the use of alternatives or more sophisticated statistical methods may get discouraged. Nonetheless, the guideline reflects the EMEA's efforts in providing quantitative tools to facilitate the identification, evaluation, and ongoing monitoring of potential safety issues.

Outcome: N/A

My first comments relate to the statement in Section 7.1 that "There is scientific consensus that signals of disproportionate reporting identified with statistical methods should always be medically assessed." Once a signal has been identified, there are some additional statistical steps that can be used, which we call "signal investigation" (as opposed to "signal detection"), which can reduce the number of signals that reach the medical reviewer and can facilitate the review process. For instance, there are automated de-duplication algorithms that can reduce the investigator burden for identification of potential duplicates, and for evaluation of signals that are due to duplication.

We have adopted and evaluated a signal investigation process at <<name>> that combines these additional computer-based steps with a human-expert-based adjudication process in order to optimize the utilization of valuable ICSR review time. There is substantial overlap between this process and the one described in section 7.1, but ours includes some additional statistical pre-processing steps and considerations. Our signal investigation process is described as follows:

The first step of the process is to obtain a statistical view of the medications that are most frequently co-prescribed with the medicinal product under study. This information facilitates the identification of potential bystander effects and drug interactions.

Once a signal of disproportionate reporting has been identified, the second step of our signal investigation process is to obtain a second statistical view of the co-prescribed medications for the medicinal product under study, this time including only those cases where the adverse event of interest is present. Combining this information with the information from the first step can further identify potential bystander effects and drug interactions. Often the statistical evidence for a bystander effect is extremely strong and an SDR can be confidently adjudicated to be a bystander effect without a detailed review of ICSRs.

The next step in signal investigation is to evaluate the trajectories of those signals over time. In the <<name>> system, we evaluate both changes in the reporting ratio and in its statistical significance. Most genuine signals should show increasing statistical significance over time as cases accumulate. An SDR with more than ten cases which shows decreasing statistical significance over time should be questioned. It may be a happenstance finding or it may be an adverse event for which an effective risk-management plan is already in place. In evaluating the time course of signals, the <<name>> software plots a line or a series of arrows for the reporting ratio on the horizontal axis and its statistical significance on the vertical axis with time as a parameter. When plotted in this manner, legitimate SDRs appear as a series of arrows moving upward. This plot has a more compelling appearance than a plot of either reporting ratio or statistical significance over time, and is very useful in communicating signal trends effectively to non-statisticians.

Another component of signal investigation is what we refer to as an "event-focused" analysis. Here we consider the adverse event of interest and look for signals of disproportionate reporting for the various medicinal products in the adverse-event database—a reversal of the usual paradigm. In other words, we produce a list of disproportionate drugs for an event, rather than events for a drug. This analysis can be useful in identifying class effects or, conversely, medicinal products that behave differently from other members of their class or indication. The event-focused analysis can also identify confounding factors for a given adverse event that should be noted in medical review, and should be taken into account if an observational study of the adverse event is called for.

In some cases, an event-focused analysis will trigger the study of medicinal products related to the product of interest by class or indication. Tabular and graphical representations of the SDRs for several related products can be helpful in establishing the context for the review of a particular medicinal product.

With these pieces of information in hand, our signal investigation process permits the investigator to adjudicate an SDR into one of several categories at this point: a) The adverse event term is uninterpretable or is too generic for meaningful further investigation. In the MedDRA terminology, examples include unexpected therapeutic effect and feeling abnormal; b) The SDR is due to a bystander effect from a concomitant medication; c) The SDR is due to confounding with the indication for the medicinal product, e.g. bacterial infection for an antibiotic; d) The SDR is due to confounding with clinical or demographic characteristics of the patients using the medicinal product, e.g. hiv wasting syndrome associated with an antifungal drug; e) The SDR covers an adverse event which is already adequately described in prescribing information; f) The SDR falls in none of the categories above and requires further investigation.

Detailed review of ICSRs takes place only for SDRs in the last category. There is, of course, a non-zero risk in adjudicating SDRs through these mechanisms prior to a review of ICSRs. This risk must be balanced against the benefit of allowing the medical reviewers, a scarce resource, to spend more time studying the ICSRs for signals that are more likely to represent legitimate safety issues for medicinal products.

The Working Group should provide a more detailed description of the signal investigation process, including the potential role of computer algorithms in reducing the need for detailed case review, and a description of conditions under which detailed review of ICSRs is considered unnecessary. The sense that every SDR must be subjected to a detailed medical ICSR review may hinder the adoption of automated statistical signal detection algorithms, to the net detriment of the drug safety system.

A validation study of our signal investigation process, in terms of its ability to identify SDRs that eventually resulted in changes to US Prescribing Information or the issuance of warnings or recalls will appear in *Drug Information Journal* in the fall.

Outcome: The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).

There is also an opportunity to investigate whether signal detection algorithms can be used to monitor the effectiveness of risk management plans. Theoretically, an effective risk management plan might result in the reduction or disappearance of an SDR over time. Using signal-detection algorithms in this manner should be considered.

Outcome: The checking of the effectiveness is a component of the risk minimisation activities and a key element of risk management. Some methods have been proposed to evaluate the effectiveness of regulatory action (see for example Waller P and Tilson HH Managing drug safety issues with marketed products in Stephen's detection of new adverse drug reactions 5th edition. Wiley. Chichester 2004). The spontaneous reporting is one of these methods. Therefore, the role of EudraVigilance in the context of risk management activities has been emphasised in the guideline.

The guideline should emphasize the importance of placing the statistical data mining methods in a larger context of overall pharmacovigilance programs. In particular, determining thresholds for statistical methods are often inevitably arbitrary. Complementing statistical findings with traditional pharmacovigilance approaches would be essential. Outcome: The comment is valid and is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 8 seems to be out of scope of this document as it is beyond statistical signal detection. For monitoring TMEs, other methods than quantitative signal detection are more appropriate. Outcome: the comment was not taken into account in the revised guideline since the detection of new signals is a key component of the risk management activities (see Guideline on risk management systems for medicinal products for human use EMEA/CHMP/96268/2005).
SIGNAL MANAGEMENT (PRIORITISATION, EVALUATION AND COMMUNICATION)
A signal-detection algorithm can serve to rank SDRs in a meaningful order, in addition to detecting (and rejecting) SDRs. Signal-detection software should provide such ranked lists for review. Ranking serves a psychological function that can facilitate review by focusing particular attention on the top-ranked signals. It also provides a context which contributes to the credibility of signals. For the PRR algorithm, ranking signals on the basis of chi-squared rather than by PRR itself can reduce the tendency of this algorithm to put medically-obscure signals near the top of the list. For <<name>>, statistical unexpectedness (1/p-value) is used for ranking. For MGPS, either EBGM or EB05 can serve as ranking variables. In the validation study for <<name>>, we found that highly-ranked (top 12) SDRs for medically-significant adverse events often preceded the strengthening of safety warnings for events that were previously just listed in the "Adverse Events" section of the Prescribing Information or simply appeared in an adverse-event table. Thus the extra attention to highly ranked signals appears to be justified. Outcome: It is recognised that the prioritisation of SDRs should be based on the potential impact of the signal in terms of Public Health. The rest of the comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
In the implementation of this guideline, there is an opportunity to monitor the performance of signal-detection algorithms as they are deployed over the next few years. This could improve the situation regarding validation of signal detection methods discussed in Section 6. The Working Group should consider a mechanism for tracking SDRs detected, and enumerating which of these eventually have an impact in terms of product labeling, warnings, or recalls, and which do not. This tally could be accompanied by additional information that might lead to the refinement of algorithms and the improvement of their accuracy. The fact that action was taken on a signal is not the same thing as a scientific determination of causality for the drug-event combination, but it is an objective, measurable benchmark that may be useful in the evaluation of automated signal detection methods. Outcome: For that purpose a signal tracking tool has been developed and is being piloted by the EMEA.

It is likely that this process could lead to a significant duplication of signal generation activities between the EU competent authorities, and across the EU, FDA and WHO Uppsala Monitoring Centre, with each party using overlapping (but not identical) data and different statistical methods. As a result, MAH's could face an unprecedented barrage of similar (but not identical) regulatory enquiries if these signal generating activities are not co-ordinated across the various parties. What measures do the EMEA intend in order to reduce the likelihood of redundant signal generation and evaluation activities by the competent authorities and the MAHs?

Outcome: The guideline emphasises that any communication with the different EU stakeholders in the framework of signal detection should be done strictly in accordance with the rules laid down in the existing Community legislation and guidance documents. In addition, the coordination of the detection of new signals at the level of the CHMP-PhVWP will streamline (rationalise) the communication between the EU National Competent Authorities and the Marketing Authorisation Holders. The exchange of information with other non-EU Regulatory Agency under the confidentiality agreement (when such confidentiality agreement has been established) is also a way to exchange information on signals which also reduces the duplication of efforts for the Marketing Authorisation Holders. However, when regulatory decisions must be made, the MAH can be requested to provide the necessary scientific supporting evidence to all the relevant Regultory Authorities to characterise, refute or confirm a signal in order to make evidence-based decisions.

There should be guidance about:

- The selection of cases and data fields (e.g. E2B fields) to be considered for signal detection, and
- The documentation of the signal detection process

i.e. the definitions which have been made before running the signal tool as e.g. database datalock point, definition of the drug of interest, definition of the event of interest, definition of strata, etc. and the signal output.

The rationale for this suggestion is as follows:

1. Results from a signal detection process should always be reproducible.
2. It is assumed that the competent authority will contact the MAHs to assess certain SDRs identified with the EudraVigilance Analysis System.

The documentation should not only include outputs as given in tables 2 to 7 in the guideline, but also the definitions made before generating the SDR so that the MAH is able to adequately respond and further investigate the issue.

The guideline would also benefit from having a detailed description of the data/report elements available for analysis (e.g. availability of narratives), and a separate discussion on the limitations of the EudraVigilance database as they relate to the interpretation of SDRs (e.g. proportion of duplicate reports).

Outcome: Concerning the first aspect of the question relating to the fields used for signal detection purposes and the standard terminologies, the signal detection is performed on the standards and terminologies developed under the auspices of ICH. Further practical guidance is given in the EudraVigilance Data Analysis System training given to the users. EudraVigilance Access policies are currently under development and will define the appropriate level of access given to the system to all the stakeholders taking into account the individual data protection and protection of commercially confidential information. The access to the case narratives will be addressed in these access policies. Secondly, the EudraVigilance Data Analysis System has been subject to a technical validation which ensures that the results obtained with the EV DAS are reproducible. Finally, the EU legislation and guidance documents clarifies the roles and responsibilities of all the stakeholders involved in the pharmacovigilance for the products authorised in the EU (including the communication with the Marketing Authorisation during the evaluation of the signals), the guideline emphasises that the communication with the MAH should be done in line with the EU regulatory pharmacovigilance procedures.

There are appropriate and necessary cautions in sections 1, 4(d), 6, 7 on the interpretation of the methods described. However, these omit some important additional caveats of the method proposed, for example, stronger recognition of the need to consider bias implications from the use of observational databases. If background information (e.g., severity of condition) will not be available from the database, such confounding factors will unknowingly influence the proposed signal detection procedure.

Outcome: The issue was further highlighted in the guideline.

It is recommended that this statistical guidance be directly linked to a clinical guideline that details the intent and objective of signal monitoring.

Outcome: The comment is taken into account.

The process of generating 'signals of disproportionate reporting' (SDR) is clearly described. The guideline emphasizes that such signals require further steps of analysis, in particular quality control and medical/clinical assessment. GEHC appreciate the approach towards more formalised detection of safety signals from individual case safety reports.

Outcome: N/A.

MISC. STATISTICAL CONSIDERATIONS

Caution should be exercised in calling for analyses as multiple levels of the MedDRA hierarchy (Footnote 3) and for stratified as well as unstratified analyses (Section 4.5(g) and 4.7), particularly when these analyses are ad hoc rather than pre-specified. When multiple analyses are preformed, it may be appropriate to apply a multiplicity adjustment to the detection threshold for SDRs. Appropriate adjustment methods should be considered and described.

Outcome: There is no adjustment for multiplicity in the PRR method. Distinction is made between standard analyses for signal detection and additional exploratory analyses.

We note a brief mention of drug interactions in Section 8 but the guideline contains no specific recommendations as to how drug interactions should be detected or ruled out. Further discussion of drug interactions should be included.

Outcome: Outcome: The detection of drug-drug interactions using DMAs is a complicated topic which requires some additional research. This aspect will be further developed in EudraVigilance in the future.

We would recommend including a statistician in the review of the document as there are numerous inconsistencies and methodological shortcomings in the document, e.g.

- chisquare statistic is not a measure of association
- the threshold introduced in section 4.6 used for the chisquare statistic are not based on scientific rationales
- the use of confidence intervals rather than test statistics should be pointed out and implemented in the tool. The consistency between test statistics and confidence intervals should be discussed.
- the x^2 -formula in section 5.2.2 is not correct

Furthermore, the guideline should clearly state that beyond the request of a medical assessment, a statistical assessment is required as well.

Outcome: These mistakes have been corrected.

Subgroup analysis, Stratification and sensitivity

Reference to the first two points is given briefly in paragraph 4.7. There is no mention to a sensitivity analysis for example after exclusion of individual cases or for a limited population.

Due to the exploratory characteristics of this kind of analysis, reference to sensitivity analysis and in general on how to deal with inconsistent results could be appropriate.

Outcome: Further investigation of the signal may be based on subgroups or stratification, interpretation of these results will not be based on inferential statistics but on further medical evaluation of the signal.

Other statistical comments

The method of comparing a single product to all other products seems rather arbitrary, it would be more relevant to compare like-with-like, e.g., within class or within indication of drug.

The method proposed weighs all data equally (just like constructing a mean value), which is efficient in a retrospective setting when one analysis is performed after all data have been collected. However, during surveillance, the data are analysed sequentially. Additionally, the method assumes that the length of exposure is the same in each group, rather than accounting for variation in follow-up time.

The intention is presumably that the criteria are 'and', in that both the PRR and the individual case criteria need to be satisfied; the current text implies in places that individual cases >= 3 alone is sufficient.

The drill-down option can show the data by age, sex, etc but interpreting those data will be additionally challenging for several reasons, for example:

- How different is different enough to conclude a different effect?
- What difference is acceptable considering that medical products are aimed at different patient populations with distinct characteristics of demographics and medical history?

Section 4.5 (c) states that the thresholds for detecting SDRs are a trade-off, which is true. However, the given criteria appear rather high (equivalent to $p<0.05$) which, given the volume of tests being proposed, would by chance identify many false-positives.

In general the properties of the PRR method are not well known, but some comments regarding properties for surveillance methods could be helpful, to describe the information available and also the limitations. The “type 1 error” is usually characterised by the average run length to a false alarm or the probability of a false alarm. Typically, the detection ability depends on when the change occurs and is usually evaluated using measures such as the expected delay in an alarm etc. The traditional concept of “power” cannot be used.

We would recommend to include a statistician in the review of the document as there are numerous inconsistencies and methodological shortcomings in the document, e.g. chi-square statistic is not a measure of association, the threshold introduced in section 4.6 used for the chi-square statistic are not based on scientific rationales, the use of confidence intervals rather than test statistics should be pointed out and implemented in the tool, the consistency between test statistics and confidence intervals should be discussed, the x^2 -formula in section 5.2.2 is not correct.

Furthermore, the guideline should clearly state that beyond the request of a medical assessment, a statistical assessment is required as well.

Outcome: the comment is acknowledged and the guideline has been modified accordingly.

This guideline utilises PRRs without stratification for the detection of signals in the EudraVigilance database - we would appreciate the rationale for using unstratified PRRs as selected method for signal detection by the EMEA.
Sections 5.1 and 5.2 suggest that the EMEA might use adjusted or stratified PRR (by age, gender, country etc.). Such estimates would be more efficient than the crude PRRs as SDRs can often be explained by differences in the distribution of such factors. The EMEA might be wise to consider implementing stratified PRR as routine output.

Outcome: Stratified analyses will be implemented at a later stage and will require additional validation.

PRR, as originally defined by Evans et al[2], is an easy-to-interpret and computational efficient method for calculating the relative reporting rate of a given event with a drug of interest as compared to the reporting rate of that event in all other drugs. However, the metric is unstable when counts of reports as recorded in the cells within the 2x2 contingency table are small; in particular, the two counts that cause the greatest variation are the number of the reports of the drug-event pair of interest and the number of reports with the event of interest but not the drug of interest. In fact, if either of these cells have 0 reports, the standard error for PRR cannot be calculated. As originally presented by Evans and as currently proposed in the EudraVigilance guidance, PRR is calculated as an unadjusted metric. Because it is well understood that are there significant reporting differences across subpopulations of interest, it is recommended that appropriate adjustments are made. PRR is amenable to adjustment through stratification by using the Mantel-Haenszel method.

Outcome: Stratified analyses will be implemented in a second phase and will require additional validation.

PERFORMANCES OF THE QUANTITATIVE METHODS (SENSITIVITY/SPECIFICITY AND THRESHOLDS)

The possibility of false negative results should be noted. PRR is the measure of the number of reported events for a drug over all the reported events for all other drugs. The maturity of the database will substantially affect what is identified as a signal or not, some events that may be considered clinically as a signal for a drug but may not be detected because the comparison group may already contain a sample of the events reported for other drugs. A rare event for a given drug may be masked by other compounds in the database for which the event is not so rare.

Outcome: These effects will be adjusted in future versions.

Conversely, it is our opinion that a PRR threshold defined by the EMEA is too low. This raises the possibility that a large number of false positives may be generated for a given drug. This will cause unwarranted regulatory worry and stress to a company's Pharmacovigilance function forcing resources to be spent examining false signals. We would like the PRR threshold of 2 to be reconsidered and raised to reduce the likelihood of false positives.

Outcome: More insight on the chosen thresholds will be obtained with the ongoing validation study. .

Section 7.0 states that "Traditional Pharmacovigilance methods are generally satisfactory when the number of reports in a database is small". What will be the impact of this initiative on smaller companies? Will they need to undertake traditional signal activities and also act on the SDR's provided by the EudraVigilance?

Outcome: The two methods are not mutually exclusive, therefore regardless of the type of product on which a surveillance is performed, all the methods could and should be used. However, when interpreting the results of any Data Mining Algorithm the limitations of the methods should be borne in mind.

A number of studies have shown that the PRR algorithm is a sensitive detector of SDRs, but is relatively nonspecific. Compared to other algorithms, a larger proportion of SDRs identified by PRR lack supporting evidence that they represent a clinically-significant safety issue. Other algorithms, including <<name>>, BCPNN, and MGPS, can provide greater specificity and should be considered. The use of multiple methods should also be considered. SDRs that are detected by multiple methods may have a higher likelihood of representing medically-significant safety issues, compared to those detected by a single method alone.

PRR and some other data mining methods require a reporting ratio of 2.0 or more in order to detect a signal. This restriction is imposed to avoid false positives. For medical events with a significant "background" rate of occurrence independent of drug adverse events, such as stroke, myocardial infarction, seizure, etc, this can represent an impediment to the detection of important SDRs. The Working Group should consider the use of a method which uses a different basis to discriminate SDRs, and is sensitive to at least some SDRs with reporting ratios between 1.0 and 2.0.

Outcome: A validation study is being conducted on EudraVigilance to assess the performances of the methods implemented in the EV DAS. If other methods were to be implemented, the whole validation exercise would have to be repeated. There is no restriction which is imposed on the size of the PRR. The masking effect of the PRR is also known, some methods such as subgrouping may help to minimise this masking effect. A good knowledge of the background is important to interpret the SDRs.

The key area of focus not detailed in either paper though is the appropriate use of thresholds in disproportionality analysis. Estimates for the reporting rate should be used as a means of prioritizing drug-event pairs to monitor, not as a means of eliminating drug-event pairs from consideration that fail to meet pre-specified criteria. If we intend to use the thresholds to create strict delineations, then we can easily adjust them accordingly to increase the number of signals at the expense of specificity (or vice versa) without affecting the underlying data; either PRR or MGPS could be more sensitive or specific depending of how we define the two thresholds. However, if the metrics are used to guide the exploration of adverse event reports, then the most effective metric is one which rank-orders drug-event pairs in the most meaningful manner. To that regard, it is clear that stratification should be employed to eliminate spurious associations due to demographic factors. Additionally, since PRR is susceptible to generating large signals in cases with low counts and MGPS provides a means of adjusting for small numbers, it is recommended that MGPS be used as the primary method for prioritizing drug-event pairs within the disproportionality analyses of spontaneous adverse event data.

Outcome: The guideline already discusses the trade-off between sensitivity and specificity. In addition, the validation study conducted on the EV DAS will provide some better insight on the thresholds used in EudraVigilance. Stratified analyses will be implemented in a second phase and will require additional validation. Finally, the validation study will also provide a better understanding of the use of the disproportionality analyses performed with the PRR in situations where the number of reports is low (orphan drugs, paediatrics, etc …) and how they best supplement the traditional methods of signal detection
.

OTHER QUANTITATIVE APPROACHES TO SIGNAL DETECTION (INCLUDING BAYESIAN METHODS)

Section for discussion in the subgroup

As long as the proportional reporting ratio (PRR) has been already implemented in the EudraVigilance Data Analysis System as the only measure of disproportionality (as mentioned in last sentence of the first para in the Executive Summary) there seems to be no need for a guideline on the "use of statistical methods" in the EudraVigilance Data Analysis System. If there is a software tool that only allows one specific statistic to be calculated, guidance may be required only on i) how to use the software at all ("user manual") and ii) the interpretation of results.
For a guideline under discussion it seems to be even more worthwhile to discuss the options of statistical methods to be used for signal detection and if possible provide guidance on which one to prefer in which situation based on the detailed content of database used (if there are differences in this regard between the different statistical methods available).

As it is by far not obvious which method is preferable or even "optimal" in general for the detection of disproportionalities in the (routine) reporting of adverse events, we would suggest to provide indeed guidance on the options available and which method(s) will be preferred. In addition, there is already increasing evidence from published reports available, indicating that the PRR has considerable drawbacks as directly compared to the empirical Bayes multi-item gamma Poisson shrinker (MGPS) [c.f. 1]. In addition, there might be other measures with better statistical and decision theoretic characteristics as the PRR, although there exist not in every case published direct comparisons to the PRR yet [2, 3, 4, 5, 6]. However, the other options beside the PRR shall be critically reviewed and discussed in a document of the scope as currently expected given the title of the guideline.

Alternate statistical methods / future work

The proposed method is empirical, with previous use having proved useful. Two commonly used Bayesian methods are the Gamma-Poisson Shrinker (GPS) [DuMouchel (1999)] and the Bayesian Confidence Propagation Neural Networks (BCPNN) [Bate et.al (1999)]. These have been more recently developed, and could provide a stronger approach. It would be worthwhile to discuss the options of statistical methods to be used for signal detection and if possible provide guidance on which one to prefer in which situation based on the detailed content of database used (if there are differences in this regard between the different statistical methods available). In general, it is difficult to realise the benefit of SDRs when the interpretations have such limitations. We also suggest the addition of a section on directions for further research, aimed at encouraging more statisticians to look at the difficult inferential problems inherent in this data mining exercise

It is evident that the EMEA has selected the proportional reporting ratio (PRR) as the signal generation system of choice. The guideline only briefly mentions that, in addition to PRR method, other algorithms will be considered for future implementation. We would like to see a more extended discussion on what other automated screening algorithms are being considered (e.g. Bayesian methods), including approximate implementation timelines.
It is true that none of the alternative statistical methods has been shown to be generally as advantageous over the PRR method. However, providing only one method in the tool may be seen as a technical drawback, particularly because:
- the WHO Uppsala Monitoring Centre uses another statistical method.

most of the commercially available software tools support other methods in addition to the PRR.

It is unclear why Proportional Reporting Ratio (PRR) is the only method being used in the EudraVigilance Data Analysis System. Selection of the PRR method is strange given published comparisons with the MGPS alternative. It is also evident that the false-positive rate from the PRR method is too high for a large pharmacovigilance database, and will lead to much unnecessary work with little extra sensitivity in detecting emerging signals. It would be helpful to:
- Describe other methods available and their advantages relative to PRR
- Explain the reasons(s) why those methods are not included as options in the current guideline, and
- Provide the timeline when those other methods will be implemented.

Other authorities and organizations, e.g. FDA, WHO and MHRA, apply other methods. Comparisons of the selected method PRR and the FDA-applied method MGPS have been done and are published.

1. Why do different authorities/organizations apply different quantitative signal detection methods to overlapping sets of spontaneous data?
2. How will the interactions work with the MAHs in case of different SDRs?
3. Why have you seemingly not considered the results of the comparison of the signal detection methods PRR vs. MGPS ([2] Almenoff et al.)? i.e.:
    - Confounding by demographic / temporal factors
    - Performance when the number of reports is small
    - Adaptation of conventional thresholds

The proposed method is empirical, with previous use having proved useful. Two commonly used Bayesian methods are the Gamma-Poisson Shrinker (GPS) [DuMouchel (1999)] and the Bayesian Confidence Propagation Neural Networks (BCPNN) [Bate et.al (1999)]. These have been more recently developed, and could provide a stronger approach. It would be useful to at least provide some comments concerning these. In general, it is difficult to realise the benefit of SDRs when the interpretations have such limitations. As the process allows for multiple data presentations for each product, it would be beneficial for the authors of the guideline to obtain further input from industry (and academia?) to develop this draft to enable the stakeholders to most appropriately review their products for signals.

<u>Alternate statistical methods / future work</u>

The proposed method is empirical, with previous use having proved useful. Two commonly used Bayesian methods are the Gamma-Poisson Shrinker (GPS) [DuMouchel (1999)] and the Bayesian Confidence Propagation Neural Networks (BCPNN) [Bate et.al (1999)]. These have been more recently developed, and could provide a stronger approach. It would be useful to at least provide some comments concerning these. In general, it is difficult to realise the benefit of SDRs when the interpretations have such limitations. We also suggest the addition of a section on directions for further research, aimed at encouraging more statisticians to look at the difficult inferential problems inherent in this data mining exercise.

Since the EudraVigilance Data Analysis System is work in progress (section 4.1 'Other methods will be considered for future implementation'), perhaps the MGPS approach could be added to the system with some guidance on how to interpret differing results. Other regulatory agencies have commented that these safety signal detection techniques should take into account the length of exposure to drug. This does not seem to be taken into account in section 4.2, and is likely related to the number of false positives quoted thereafter. In terms of safety, this approach adopts a policy of grading events relative to other events for the same product in section 4.2. Several comments on this: - This method does not look to match the intent of the practice, that is to detect and monitor signals of harm from drugs in the population of concern. Data from other drugs in the population of concern is not included. Nor is information on background rates in a population not treated with drug. - Events are not weighted for the importance of the event i.e. a death is not the same as a tooth-ache, but both constitute an event.
It is recommended that the EMEA consider following the best practices established by the FDA and MHRA in regards to disproportionality analysis on spontaneous adverse event data. In particular, EMEA should consider adopting additional statistical methods, like Multi-item Gamma Poisson Shrinker (MGPS), as part of its pharmacovigilance process. If the EMEA chooses to continue to move forward with PRR as its metric of choice, it is recommended they use stratification on age, gender, year of report, and potentially country of origin, to improve the metric performance. Moreover, it is recommended they agree upon one common threshold (either using 90% CI of PRR or x^2) and use that threshold only as a means of identifying and prioritizing candidate signals, not in evaluating the drug-event pairs. Throughout the pharmacovigilance process, it is recommended the EMEA highlight the limitations of the methods employed, as well as the biases and limitations of the spontaneous data itself. Finally, it should be noted that there have been many advances in use of other data sources, like observational claims and electronic medical records, for pharmacovigilance activities that address many of the limitations of spontaneous adverse event data.
Disproportionality analysis (DA) is the primary class of methods developed for use in spontaneous adverse event reporting database to assess how much the observed frequency of a given drug-combination pair deviates from the expected frequency, given statistical independence between drug and condition. With all methods, a high relative reporting rate does not necessarily indicate a high incidence of the condition or suggest a causal relationship between the drug and the condition. There have been many approaches developed, including: proportional reporting ratio (PRR as is recommended in this guidance), reporting odds ratio (ROR), the x^2 test with Yates' correction (similar to guidance, but preferred for use with contingency tables), multi-item Gamma Poisson Shrinker (MGPS), and Bayesian confidence propagation neural network (BCPNN). While there is no agreed best practice and no definitive approach to determine the optimal method, this document highlights the strengths and limitations of PRR and MGPS as a basis of comparison. A more detailed discussion of the methods was produced by Fedorov and Lin[1].

Multi-Item Gamma Poisson Shrinker[3,4] is a data mining algorithm that computes the empiric Bayes geometric mean (EBGM) and corresponding 2-sided 90% confidence interval (EB05 < EB95) for each observed drug-condition pair in a reporting database. Currently, the FDA, MHRA, <<NAME>> and other industry partners work with <<NAME>> to use MGPS as part of their routine pharmacovigilance activities, using both the FDA AERS data as well as company-specific databases, such as <<NAME>>'s OCEANS database. EBGM values represent relative reporting rates (after Bayesian smoothing or "shrinkage") for drug-condition pairs in a given database. An EBGM of 5 means that a drug-condition pair has been reported 5 times as frequently as would be expected if reports involving the drug and reports of the condition were independent (i.e., no reporting association). The FDA has used EB05 22 as a threshold for signal detection. This threshold ensures with a high degree of confidence that regardless of the number of reports, a particular drug-condition combination is being reported at least twice as often as would be expected if there were no association between the drug and the condition[5]. The "standard" internal stratification variables used in MGPS are age, gender, and year of report. Stratification minimizes the detection of apparent drug-condition associations that are actually due to independent relationships that may exist between a drug and a strata variable and a condition and the same strata variable. Analyses are stratified by year of report to reduce the chance of detecting signals that might arise due to "trendiness" in the reporting of specific drugs and/or events[6]. MGPS is computationally intensive and not as immediately interpretable as PRR. However, MGPS is theoretically appealing, because it makes use of the Poisson distribution as part of its Bayesian prior; adverse event reports occur randomly at a homogenous rate over time, even for rare events, following the classical Poisson process characteristics of orderliness and memorylessness.

It should be noted that stratified PRR and MGPS are expected to converge with sufficiently large sample size, but a primary interest is to review drug-event pairs within the domain of small case counts. The issue of small counts is at the heart of pharmacovigilance, since our primary aim is to detect and respond to potential drug-event associations quickly so as to maintain the proper benefit-risk profiles of our medicines. As such, our goal should be to address concerns with small counts of the drug-event pair of interest. Small counts can commonly occur for two primary reasons: the drug of interest is recently marketed and/or has not had many reports; and the event of interest is not commonly reported (such as infrequently used medDRA terms). In this context, there has been much work evaluating the relative performance of the two metrics. The concept of a drug-event pair being reported at a disproportionate rate is not strictly analogous to the relation of interest that a drug has a causal association with an event. As such, there is no means of evaluating the true sensitivity and specificity of the metrics since 'false positives' and 'false negatives' generally refer to the ability of the test (DPA method) to predict the truth (causal relation between drug and event). However, as a surrogate to assess sensitivity and specificity, Hauben et al[7] and Almenoff et al[8] demonstrated properties of the methods through real-world and simulated examples. Using thresholds of PRR>2 and EB05>2, Hauben showed that PRR produced more signals, thus increasing 'sensitivity'. They argue that additional 'false positives' are not of public safety concern and simply require additional triage. Almenoff showed, using the same thresholds, that MGPS created fewer signals in drug-event pairs with low counts and demonstrated the impact of stratification in eliminating associations attributable to gender, age and year of report. The paper argues that higher specificity, in combination with ongoing medical triage, is a more resource-efficient process.

Outcome: The comments on the implementation of Bayesian methods are acknowledged. The draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with all the content of these comments. The idea which may suggest that false positives and false negatives have not been considered when implementing the PRR method in EudraVigilance is not correct. In addition, the issue of false positives and false negatives must also be considered in a Public Health perspective. The importance of both false positives and false negatives has been highlighted in the past in the scientific literature (see for example Hauben Drug Safety 2007; 30(7):627-630 and Hauben *et al*. Expert Opin Drug Saf 2005 4(5): 929-948). The potential Public Health importance of missing true signals (false negatives) must be emphasised and deserves a thorough assessment. Finally, the comments must also be qualified by the fact that some studies aimed at evaluating the performances of some Bayesian methods mentioned above have shown that true signals can also be detected earlier than these quantitative methods, by traditional signal detection methods (see Szarfman Drug Safety 2002; 25(6): 381-392). The implementation of Bayesian methods will be considered in the future taking into account the comments received on the guideline, the elements mentioned above and the fact that some of these methods are computationally demanding.
DEFINITION OF TERMS
Many guidelines have a section for 'Definitions' and it would be helpful if this one did (even though most of the different terms are defined in the text). Specifically, it would be helpful to have a section with definitions for: Statistical signal; signal of disproportional reporting; potential signal; signal; statistical association; statistical noise; individual case safety report, proportional reporting ratio; drug–event pair; suspect; interacting. It may be that 'statistical signal' and 'statistical association' are the same thing, in which case one of the terms could be dropped.
Section 1: The seeming reluctance to use the term 'signal' is troublesome. The CIOMS VI definition is very clear that a signal is a finding "worthy of further investigation", and as such has a lack of certainty built into it. It just confuses the issue when a guideline starts talking of "Signal of Disproportionate Reporting", "potential signals" and "statistical signals". Additionally, the description of non-causal factors (bottom of page 3) does not include the possibility of false positives arising from the large number of comparisons that are carried out. Whilst it could be argued that it is implied in "statistical noise", it is not clear in this respect. The final paragraph of section 1 is also of concern when it talks of "stakeholder training, ...". PPR's are by no means the only methodology available for signal detection. A variety of other methods are available, such as the Berry & Berry methodology and the Empirical Bayesian Shrinkage methodology of DuMouchel. It is a concern that the EMEA are taking such a restricted approach to the analysis of pharmacovigilance databases.
On pages 3 and 8, what are 'non-interventional clinical trials'? There is no such thing! Clinical trials are interventional by their very nature. It would be better to use 'observational studies' or 'excluding interventional clinical trials' in the text.

Please add an overview of abbreviations used in the guideline (e.g. EEA, E2B(M), PIL) and a glossary of terms used (e.g. signal, potential signal, TME etc.). Many guidelines have a section for 'Definitions' and it would be helpful if this one did (even though most of the different terms are defined in the text). Specifically, it would be helpful to have a section with definitions for: - Statistical signal - Signal of disproportional reporting - Potential signal - Statistical signal - Statistical association - It may be that 'statistical signal' and 'statistical association' are the same thing, in which case one of the terms could be dropped. - Statistical noise - Individual case safety report - Proportional reporting ratio - Drug–event pair - Suspect - Interacting Please be consistent in using terms throughout the guideline e.g. 'SDRs' instead of 'potential signals'.		
Outcome: A glossary has been added to the guideline with a cross-reference to the Volume 9A of the rules governing medicinal products for human use.		
GUIDELINE SECTION TITLE: 1. EXECUTIVE SUMMARY		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Para 1 ,line 3	It is unclear where the "*stakeholders*" refers to MAH or regulatory agency. It would be helpful to specify that who would have access to this data system and run this type of analysis.	The guideline will be updated once the EudraVigilance Access Policies will be adopted.
Para 2 ,lines 3-4	Will the signal generation be based on combined data or separately for spontaneous reports and data from non-interventional clinical trial reports? What is the rationale if a combined dataset is used, given that reports from non-interventional studies are often solicited, if not just stimulated (I.e. they are not spontaneous)?	The quantitative methods of signal detection have primarily been designed to be used on spontaneous reporting systems databases. It is difficult to know without any prior scientific assessment whether a spontaneous report has been solicited or stimulated. Therefore, the quantitative methods implemented in the EudraVigilance Data Analysis system will be applied to the reports arising from the non-organised methods of collection of data (E2B(M) field A.1.4 type of report spontaneous report).
Para 2, lines 3-4	To ensure the consistency on what ICSR is referred to. Here it says "*spontaneous reports or reports from non-interventional products*". In the SCOPE (page 4), it states "*spontaneous reports or non-organized*	The guideline was modified to refer to the non-organised methods of collection of data (in opposition to the organised methods of collection of

	methods of collection of data". In section 5.1 (Standard outputs of SDRs), it states "*spontaneous reports and reports from non-interventional clinical trials*". Please use consistent language, presumably "*spontaneous reports and reports from non-interventional clinical trials*", when referring to ICSRs.	data see ICH Topic E2D).
para 2, line 5	The text states "*…and/or the number of individual cases…*" which implies that the number of cases alone might be sufficient to qualify as a 'signal'. Usually the number of cases is used together with the measures of disproportionality, and not by itself. Either change the '*and/or'* to '*together with'* or add some more words of explanation.	'*and/or'* changed to '*together with'*.
para 2, lines 6-7	Please clarify that interpretation of SDRs requires not only thorough knowledge of both the data available and the statistical methods but also the medical understanding of the drug and the disease indication. Please add medical/clinical knowledge and understanding of the drug and the indicated disease as important elements in the interpretation of SDRs.	The text now reads as follows: The interpretation of SDRs is often complex and requires thorough knowledge of the data available in the EudraVigilance Data Analysis System, the statistical methods applied as well as the medical knowledge of the medicinal product and the medical condition or underlying disease of the patients treated with this medical product.
para 3, line 2	To ensure the consistency on what 'data quality control' means. Here it says "*presence of potential duplicate reports and controls of data quality in terms of e.g., completeness or coding of data*". In section 4.5 "*the coding practices*" is mentioned. In section 7.1 it says "*identification of potential duplicates*" on individual cases and "*data quality check*" on "*completeness of the information provided, the coding practices, …*". To use either duplicate reports or duplicate cases throughout. Otherwise, indicate that the terms 'reports' and 'cases' are interchangeable. It is preferred to use 'cases' in this situation. To use 'the coding practices' rather than 'coding of data'.	The executive summary explicitly refers to the elimination of the case reports which are duplicates of original case reports. The wording "duplicate reports" has been kept. The reference to the coding practices has been implemented.
para 3 lines 3-5 [& Section 5.2	The document gives the impression that EudraVigilance can also serve as database for explorative analyses of SDRs. It should be clearly presented that EudraVigilance can be used to:	The comment was not taken into account. Once a SDR is identified, EV is used both in the signal detection and initial signal evaluation processes.

first sentence]	1. Generate SDRs 2. Investigate if signals identified elsewhere are confirmed as SDRs in EudraVigilance but that EudraVigilance is not suitable to confirm and evaluate SDRs.	
Page 3, Footnote	As one of the stakeholders, footnoted in page 3 "*in line with the 'EudraVigilance Access Policies' currently being elaborated by the EV-EWG in accordance with Community legislation*," would the MAH have access to the EudraVigilance data (perhaps limited to their own marketed products) and the statistical tool? Once the '*EudraVigilance Access Policies*" are finalized, the guideline should be updated to make it very clear to whom these guidelines are being targeted and for what purpose. If the purpose is to allow industry access to the database, the guideline should consider a situation where a different statistical method is used and comes to a different conclusion than the one using PRR method.	The guideline will be updated once the EudraVigilance Access Policies will be adopted.
Executive summary (footnote 1):	What is meant by 'stakeholders'? The guideline explains (footnote 1): "In line with the 'EudraVigilance Access Policies' currently being elaborated by the EV-EWG in accordance with Community legislation." It would be desirable if Market Authorisation Holders were able to utilise this tool for internal signal generation efforts.	The guideline will be updated once the EudraVigilance Access Policies will be adopted... The comment on the access to Marketing Authorisation Holders is acknowledged.

GUIDELINE SECTION TITLE: 1. INTRODUCTION (background)		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Para 2, line 3	Please recognize that it is widely considered as poor practice to evaluate a potential signal using the database that generated the signal in the first place. Amend to: 'It provides tools that facilitate the identification ~~, evaluation and ongoing monitoring of 'potential signals' o~~f SDRs…'.	The comment was not implemented: The signal evaluation process should include the evaluation of all scientific evidence available. Therefore, It is recognised that the evaluation of SDRs is the first step of the evaluation process of a signal arising from a spontaneous reporting system. At a second stage, the signal evaluation may take some other evidence into account (quality, non-clinical, clinical and other clinical efficacy and safety data from post-authorisation safety / observational studies as appropriate).
Para 4, lines 2-4	It is not accurate to say "…, *based on the frequency of ICSRs on the reported drug and the frequency of ICSRs of a specific adverse event*". Please add a sentence on what the null hypothesis is in addition to the alternative hypothesis. Please amend to: "…, based on the frequency of ICSRs of a specific adverse event on the reported drug and the frequency of ICSRs of all other adverse events on all other drugs".	Under the hypothesis of independence, the null hypothesis is that there is no disproportionality of reporting. Concerning the hypothesis testing for the chi-square statistics see for example DG Rees. Essential statistics 4th edition. Chapman & Hall/CRC Boca Raton 2001 or Collett D. Modelling binary data 2nd edition. Chapman & Hall/CRC Boca Raton 2003. The guideline was not modified.
Para 5, line 7	The phrase '*cannot be established a priori in the context of data analysis*' is too complex Amend to: '…cannot be established based on data analysis alone.'	The comment was implemented.
Para 6, line 7	For the '*stakeholder training*', the strengths and limitations of the ICSR data and statistical methods should be included.	The strengths and limitations of spontaneous reporting systems are discussed in many pharmacovigilance textbooks. Even if important for the interpretation of the quantitative methods (in particular for the assumptions underlying the use of disproportionality analysis), a detailed description of these limitations is out of the scope of the guideline. Reference is made to these textbooks.
Introduction	*"It provides tools that facilitate the identification, evaluation and ongoing monitoring of 'potential signals'"* Disproportionality analysis (DPA) on spontaneous adverse event databases can only be used as one mechanism for signal detection. The	The evaluation of individual case safety reports is the first step of the signal evaluation process even if the evaluation is a holistic process which may also take some other evidence into account (quality, non-clinical and other clinical efficacy and safety data including data from

	information contained within adverse event reports is not sufficient to conduct a comprehensive signal evaluation, which requires other information sources and clinical judgment to assess the causal association between the drug and the event.	post-authorisation safety / observational studies as appropriate).
Introduction, para 6	*"In general, the use of statistical methods in detecting 'potential signals' is useful but requires a profound knowledge of the available data in EudraVigilance to interpret correctly the results originating from calculations"* The are two primary issues with the expertise required for disproportionality analysis on spontaneous adverse event reporting data; the reviewer needs to understand the issues and limitations of the statistical method, as well the issues and limitations of the data itself. In particular, because the PRR method is particularly unstable with low case counts, SDRs generated using the proposed PRR threshold should be evaluated carefully since the method may not be as precise. Perhaps more importantly, any SDR needs to be considered the context of the biases associated with the data source to prevent both 'false positives' and 'false negatives'.	The issue of false positive is discussed in the guideline.
Introduction and Section 5.1, line 5	what are 'non-interventional clinical trials'? Better to say observational studies or 'excluding interventional clinical trials'	The wording used in Directive 2001/20/EC was used in the guideline. A reference to Directive 2001/20/EC will be added for clarity.

GUIDELINE SECTION TITLE: 2. SCOPE		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Para 1, lines 3-6	With regards to the statement "*It encompasses the use of … from health care professionals and associated to authorized medicinal products* …", does this mean that 'consumer reports' are not considered? If so, what is the rationale? Are there any other restrictions, e.g. only serious case reports?	The ICSRs contained in EudraVigilance result from the reporting obligations of Marketing Authorisation Holders in accordance with the community legislation (Directive 2001/83/EC, as amended and Regulation (EC) No 726/2004). Only medically confirmed ICSRs are transmitted to EudraVigilance.
Para 1, lines 5-6	The term '*non-organised methods of collection of data*' seems odd, particularly since in section 4.1 the guideline refers to '*formal epidemiological studies*'.	The terminology is taken from ICH topic E2D.

	Please amend to: '…i.e. spontaneous reports or other ad hoc methods of data collection.'	
Para 1, lines 5-6	The term 'non-organised methods of collection of data' is a bit odd (particularly since in section 4.1, the guideline refers to 'formal epidemiological studies'. Suggest: 'i.e. spontaneous reports or other *ad hoc* methods of data collection.'	The terminology is taken from ICH topic E2D.
Para 2, lines 1-2	With regards to the statement "*Regulatory steps that can arise following the confirmation of a SDR are described in Community legislation and related pharmacovigilance guidelines.*", the Guideline needs to specify how SDRs are confirmed. It should specify whether confirmed SDRs only are included in published reports. Handling of discrepancies when there are conflicting results between PRR analysis and the confirmation process should be described. A specific section on confirmation of SDRs could be included after section 4.6 (Thresholds defining SDRs in EudraVigilance) and before section 4.7 (Subgroup analyses and stratification).	The signal evaluation, confirmation and communication is beyond the scope if this guideline, reference is made to the relevant regulatory procedures are described in the community legislation and guidelines (see Volume 9A of the Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use)

3. LEGAL BASIS		
Line no. + para no.	**Comment and Rationale**	**Outcome**
	No comment received on this section.	N/A.

GUIDELINE SECTION TITLE: 4. GENERAL CONSIDERATIONS ON POTENTIAL SIGNALS GENERATED BY STATISTICAL METHODS		
GUIDELINE SECTION TITLE: 4.1. Signals of disproportionate reporting		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 4.1, lines 3-5	The different quantitative signal detection methods have been compared and published in numerous articles, e.g. Almenoff et al. Please add a brief overview of the different methods with their	Such overview is out of the scope of the guideline which addresses the methods which have been implemented in the EudraVigilance Data Analysis System. Review articles on all the current methods have been

	respective advantages and disadvantages. Please provide the rationale for using PRR as the current signal detection method. State which methods are intended for implementation in the future.	published in the literature.
Section 4.1, line 2	The term "*potential signal*" used here and elsewhere in the document has not been clearly defined. How is it different from the WHO definition of signal (Edwards et al, 1994)? Suggest using the term 'signal' throughout the text and provide reference to the signal definition, or include a table of "Definition of Terms"	Even if the term "signal" is widely used in the pharmacovigilance community, there is no definition of the term signal in the Community legislation and consequently in the implementing guidelines. The term "signal" is consistently used in the guideline to express the existence of a risk associated with the use of a medicinal product authorised in the Community without making any judgement on the causal relationship between the administration of the product and the occurrence of the reaction.
Section 4.1 lines 5-6	With regards to the statement *"Other methods will be considered for future implementation"*, other signal detection methods that might be considered include the Empirical Bayes Geometric Mean (EBGM) or the Reporting Odds Ratio (ROR) methods. Furthermore, the guideline should describe how signal detection that came to different conclusions using different methods (either finding a signal or not) are handled, no matter what database was used. We propose to add a section for the EBGM method and one for the ROR method, highlighting advantages and disadvantages compared to the PRR method.	Such overview is out of the scope of the guideline which addresses the methods which have been implemented in the EudraVigilance Data Analysis System. Review articles on all the current methods have been published in the literature.
Section 4.1, lines 1-3	Please clearly explain the relationship between SDRs and potential signals. Furthermore, what are the key differences between the potential signals originating from SDRs and those arising from individual case analysis or epidemiological studies? In which order should such potential signals should be prioritized? Please add explanatory text in this section as well as in section 7 (Integration of statistical methods with the classical methods of signal detection in pharmacovigilance).	A signal of disproportionate reporting must be viewed in a medical context. When considered necessary, before any decision is made the SDR require a systematic medical evaluation to assess the strength of the causal association between the administration of the medicine and the occurrence of the adverse reaction. The signal confirmation step may require the conduct of additional studies (including epidemiological studies). A signal is confirmed when the level of scientific evidence suggesting a causal association is judged to be appropriate to make the appropriate regulatory decisions concerning the authorisation of the medicinal product. It is acknowledged that the signals should be prioritised according to their potential impact on Public Health.

Section 4	Describe the considerations for the analysis data set (i.e. number of cases, possible subset of data, types of data, data quality and completeness, duplications). Include in a paragraph at the start of section 4.	Considering that the dataset is changing constantly (which would require a periodic update of the guideline with limited added value) and considering that it would make the guideline much more difficult to read, no detailed description of the dataset has been included in the guideline. The data set can be accessed directly with the EudraVigilance Data Analysis System.

GUIDELINE SECTION TITLE: 4.2. The proportional reporting ratio (PRR)		
Line no. + para no.	**Comment and Rationale**	**Outcome**
section 4.2, para 2, line 1	"In this table the elements counted are the individual cases …": It is unclear how the information is selected. If the individual case is the unit, I assume that only one drug (most suspect?) and only one event (most suspect?) are used for the calculation. Clarify criteria to select one drug and one event.	No selection is necessary. The 2x2 tables used for the PRR are the marginal tables from 2^n table were n is the total No of drugs and adverse events in the system.
Section 4.2, para 4	This para (beginning 'Following usual PV practice…' seems misplaced here. Suggest move this para to end of section 4.2 (after example 2).	The paragraph was moved.
Section 4.2, para 3, line 2	The word 'interacting' here is confusing. I think it here means 'concommitant', and not drugs that interact statistically. If it does mean interacting statistically, then how this is identified needs to be stated. Don't use 'interacting' except for statistical interactions.	The comment was taken into account in the revised guideline.
Section 4.2, para 2, lines 1-3	*"Thus, a given individual case may contribution to only one of the cells of the table, even if the individual case refers to multiple medicinal products or multiple adverse events."* It is recommended that the document specifically outline the implementation of PRR to ensure that outside readers properly interpret the resulting metric. Because an adverse event report is composed of a set of one or more drugs (suspect or otherwise) and a set of one or more events, one adverse event report can represent multiple potential drug-event pairs of interest. It is very important to not 'rule out' any of these drug-event pairs in the classification of the reports. It is possible that, upon a user specifying a particular target drug, the EVDAS recalculates counts of drug-event pairs based on the sets of drugs and events in each report to provide the listing of PRRs for all events co-reported with the target drug; this approach would calculate potentially acceptable cell counts (though would be underreporting B, C, D, as shown in Table 1. Another possible scenario that may be of cause for concern is that the	In EudraVigilance the approach of performing the computations of the PRR on the individual case counts instead of number of ADRs has been chosen to keep the independence between the variables used to compute the PRR so that the variance of the PRR will not be underestimated.

	set of drug-event pairs are pre-specified from the adverse event reports, and then PRRs are calculated on this set; in this case, all counts in Table 1 could be significantly underrepresented (most significantly, cell A, which is of most particular interest). It is recommended that the specific implementation of this calculation be further detailed such that participating organizations can evaluate the approach appropriately.	
Section 4.2, para 3, lines 1-3	*"Following usual pharmacovigilance practices, the table takes into account the medicinal products reported as 'suspect' or 'interacting'. Concomitant medication are not normally taken into account in the calculations"* A preliminary focus on only 'suspect' or 'interacting' drugs is warranted. However, this raises the concern of reporting selection bias on spontaneous adverse event data. Typically, the professional who reports the AE makes a determination of drug relation, but this assessment may not be correct. It is possible that a drug that was thought to be a concomitant medication could be the true cause of the adverse event. As such, it is recommended that additional analyses be performed without conditioning the drugs on AE relation. The disproportionality rate using all drugs should be considered as part of the evidence to determine if a given drug-event pair is a SDR warranting further evaluation. It also bears noting that signal detection and evaluation methods that leverage observational data are not subject to the same biases that are exposed here by the subject assessment of the reporter. That is, drug-condition pairs can be assessed in the context of all potential confounding factors, such that true event rates can be calculated given concomitant medications and comorbidities.	The EudraVigilance Data Analysis System has been modified in such a way that concomitant medications may be taken into account in the computation of the PRR.
Section 4.2, lines 1-2	PRR is not a method, but a ratio. It is one of the measures of the disproportionality of reporting method. Please amend to "*PRR is a* ~~statistical method~~ a measure of disproportionality of reporting *used to detect SDRs in pharmacovigilance databases such as EudraVigilance*."	The comment was implemented in the guideline.
Section 4.2,	"*In this table the elements counted are the individual cases* …": it is	No selection is necessary. The 2x2 tables used for the PRR are the

para 2, line 1	unclear how the information is selected. If the individual case is the unit, can we assume that only one drug (most suspect?) and only one event (most suspect?) are used for the calculation? Please clarify the criteria to select one drug and one event.	marginal tables from 2^n table were n is the total No of drugs and adverse events in the system.
Section 4.2, para 3, line 1	This paragraph, beginning '*Following usual PV practice…*' seems misplaced here. Suggest moving this paragraph to the end of section 4.2 (after example 2).	The comment was implemented.
Section 4.2, para 3, line 2	The word '*interacting*' here is confusing. I think it here means 'concomitant', and not drugs that interact statistically. If it does mean interacting statistically, then how this is identified needs to be stated. Don't use 'interacting' except for statistical interactions.	The guideline will clarify and specify the meaning of the word "interacting" as appropriate whether it refers to a drug-drug interaction or a statistical interaction.
Section 4.2, Table	In this table, "*all other medicinal products*" is used as the comparator. There are situations in which a specific medicinal product or a group of medicinal products (for example, those with similar indication) may be more appropriate to be used in the analysis. It would be helpful to provide a general guideline on the selection of comparison groups or comparator drugs in the analysis.	Subgroup analyses may be performed (see section 4.7 subgroup analyses and stratification). Further subgroup analyses functionalities (such as a subgrouping by anatomo-therapeutic class) will be implemented in the second phase of the release of the EV DAS.
Section 4.2, para 1	Awkward wording is used in para 1: the second sentence is presented the wrong way round: the principle is surely that the fact of more frequent AE reporting leads to the identification of an SDR.	The comment was taken into account.
Section 4.2, Table	With regards to the paragraph after Table 1 and the footnote - each event is counted once - does that mean individual PPRs are calculated, say, for Serevent and Flixotide, or budesonide and formoterol, in the database? What if the generated PPR differs widely for the two components? Is there no mechanism for assessing potential synergy between the components of combination products, which may alleviate a potential signal? By doing the components separately, are we not starting to investigate potential causal relationships, which is a contradiction to paragraph 5 of section 1? In addition, multiple reports could be an indicator of an important aspect of the potential signal, but by only counting once for each subject this is not picked up on. One can understand the reasons for taking this approach, but it reinforces the fact that this is effectively a screening tool for further evaluation, as	The detection of drug-drug interactions using DMAs is a complicated topic which requires some additional research. This aspect will be further developed in EudraVigilance in the future.

	opposed to a definitive tool.	
Section 4.2, Example 2	Example 2 is not clear as “nausea” & “diarrhoea” appear used interchangeably! Please use nausea or diarrhoea in this example.	The comment was implemented.
Sections 4.2 – 4.4	Please provide appropriate references of the formulas used for ‘PRR’, ‘s’, and ‘chi-squared’.	The comment was implemented.
Section 4.2, Example 2	There is a typographical error in the document. On page 6/21, line 2; ’15 reports of diarrhoea’. This should be ’15 reports of nausea’.	The comment was implemented.
Section 4.2	The calculation of the confidence interval for PRR is approximate whereas an exact method exist (Gart and Nam). In situations that the value **A** in table 1 of section 4.2 is rather low, the difference between exact and approximate methods may not be trivial. With the exact method of Gart and Nam it is possible to calculate the interval when **C**=0 (see footnote 3 in section 4.2) and so the arbitrary set of PRR=99.9 can be skipped.	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 4.2, Table 1	In usual "contingency table" analyses, we might assume that A,B,C,D have a multinomial distribution. If we make this assumption and we consider all possible events R and pick out the one with the highest A (or largest A/(A+B) or PRR etc) the joint distribution is no longer multinomial and our inference method should reflect this. The guidance should address this issue. If this table is constructed for all (drug, event) pairs, the multiplicity issue could be huge. It would be good if the guideline could also address this issue.	The PRR is not computed under the assumption that the joint probability distribution function is multinomial. In addition, contrary to what is suggested in the comment, the disproportionality analysis is not an inferential exercise (i.e. the exercise is not aimed at drawing conclusions about a parent population on the basis of evidence obtained from a random sample from this population). We want to emphasise that such distributional assumptions cannot be reliably adopted and checked in the context of SRS data. The PRR is not computed under the assumption that the joint probability distribution function is multinomial. The only reliable way of validating methods relying on disproportionality analysis is through empirical approach. The comment more specifically relates to the use of generalised linear models used sometimes for signal detection purposes (see in particular the work from E. van Puijenbroek). The guideline does not address the use of generalised linear models for the analysis of binary data. Under the assumption that each cell of the contingency table follows a Poisson distribution, the joint distribution of the Poisson random variables will follow a multinomial distribution. Under this assumption, nominal

		logistic regression models can be used. In some instances, as it is suggested in the comment, a natural order among the response categories could be used in the model. Under these circumstances, ordinal logistic regression models should be used. Finally, there are circumstances (clinical trials for example) where some constraints imposed on the data included in the contingency table imply that the joint distribution is not multinomial but the product multinomial distribution. In these circumstances log-linear models should be used. More information on the use of generalised linear models and analysis of binary data are given for example in the following textbooks: Collett D. Modelling binary data. Texts in statistical science. Chapman & Hall/CRCBoca Raton 2003 and Dobson AJ. An introduction to generalised linear models 2nd edition Texts in statistical science. Chapman & Hall/CRCBoca Raton 2002. The issue of multiplicity is known for the PRR and no adjustment for multiplicity is performed on this method.
Section 4.2, para 4	The PRR should not be described as being "run", but as "calculated".	The comment was implemented.
Section 4.2, Footnotes	Capitalise C (C=0)	The comment was implemented.

GUIDELINE SECTION TITLE: 4.3. The 95% CI of the PRR		
Line no. + para no.	**Comment and Rationale**	**Outcome**
4.3, line 7	Reference is missing for the formula of how to estimate standard deviation of the natural logarithm of the PRR.	A reference has been added to the guideline.
Section 4.3.	Reference is missing for the formula (sec. 6, 4.3) of how to estimate standard deviation of the natural logarithm of the PRR.	A reference has been added to the guideline.
Section 4.3, line 2, para 1	It is probably more usual to refer to the standard error of the log of the PRR, not the standard deviation (although different texts do differ on this point). The formula should refer to the standard error, not standard deviation, of ln(PRR). Should the formula for the sd read: s=sqrt (1/A - 1/C - 1/(A+B) + 1/(C+D))? Refer to 'standard error' not 'standard deviation' and use '*se*' in the equation, not '*s*'.	The comment concerning the standard error was implemented. The comment concerning the formula was not implemented. A correct formula is given for example by S Greenland and KJ Rothman. Introduction to categorical statistics in KJ Rothman and S Greenland. Modern epidemiology 2nd edition. Lippincott Williams & Wilkins. Philadelphia 1998 (quoted in reference No 27 of the article by Puijenbroek *et al.*) and by Woodward M. Epidemiology: Study Design and Data Analysis, Second Edition. Chapman & Hall/CRC Texts in Statistical Science Series . Volume 64. Boca Raton 2005.
Section 4.3	A confidence interval is an accepted way of evaluating the variability around a given point estimate. However, it is recommended that both the formula and the width of the confidence interval be refined. As defined by van Puijenbroek et al[11], the proper formula for standard error of a PRR is: $SE(\ln PRR) = .V[1/A - 1/(A+B) + 1/C + 1/(C+D)]$ Where $(1\text{-}a)CI = e^{\ln(PRR) +/- Z(a)*SE(\ln PRR)}$ The formula provided uses the common assumption that the error is normally distributed. Counts of rare events tend to follow the Poisson distribution and the normal approximation tends to bias the estimate	The comment was not implemented. A correct formula is given for example by S Greenland and KJ Rothman. Introduction to categorical statistics in KJ Rothman and S Greenland. Modern epidemiology 2nd edition. Lippincott Williams & Wilkins. Philadelphia 1998 (quoted in reference No 27 of the article by Puijenbroek *et al.*) and by Woodward M. Epidemiology: Study Design and Data Analysis, Second Edition. Chapman & Hall/CRC Texts in Statistical Science Series. Volume 64. Boca Raton 2005.

	away from no effect. <<NAME>> actively uses the lower bound of its DPA score as a basis for setting thresholds, but have used 90% confidence intervals, not 95%; the intuitive interpretation of <<NAME>>'s use is that we are 95% confident that the drug-event pair has a true disproportionality rate above the lower bound of the 90% CI. Please refine the formula as indicated.	

GUIDELINE SECTION TITLE: 4.4. The chi-square (X^2) statistics		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 4.4, line 1	The description 'which is traditionally used in disproportionality analyses' doesn't provide much information. The word 'which' should be 'that' in this sentence. Also, I find the words 'goodness of fit' very useful in describing the use of chi-square statistics, and I think they would help here.	The Chi-square statistics is now described in the guideline as an approximate statistics used to test the independence of categorical variables displayed in contingency tables. The sentence containing the word "which" instead of "that" has been deleted from the text.
Section 4.4, line 1	Why do 'certain queries' report confidence intervals and others chi-squared?	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used.
Section 4.4	Should some comment be made about small 'expected frequencies' and the problems of chi-squared in some cases?	An explanation on the normal approximation of the Poisson distribution has been added to the guideline with relevant references. More information can be found in these textbooks and in other statistical textbooks.
Section 4	Please add a additional sub-section to describe the selection of cases and data fields to be considered for signal detection. It should also be clear which data cleaning activities are done before signal detection (e.g. identification of duplicate reports). Also coding activities, e.g. MedDRA code refresh, should be described, because this may influence the signal detection results.	Concerning the first aspect of the question relating to the fields used for signal detection purposes and the standard terminologies, the signal detection is performed on the standards and terminologies developed under the auspices of ICH. Further practical guidance is given in the EudraVigilance Data Analysis System training given to the users. Concerning the use of MedDRA, The coding adverse reactions should in accordance with the latest version of the MedDRA points to consider (at the present time MedDRA Term Selection: Points to Consider Release 3.8 [Based on MedDRA Version 10.0]).

Section 4.4	Chi-square tests have been historically used as a means of testing the independence between the cells in Table 1. However, largely the outcome of such a test is similar to the confidence interval of PRR- it provides a measure of variability. Hence, it is not recommended to construct signal thresholds on both the CI of PRR and x^2. The committee should decide on one common threshold (CI or x^2) and provide the appropriate statistics on all views. If x^2 is to be used, the formula should be refined. It is well known that the x^2 statistic is biased away from the null hypothesis when the any cell count in a 2x2 table is less than 5. Because Table 1 represents a contingency table, it is recommended that the calculation use Yates' correction for continuity, which prevents the overestimation with small data. $x^2_{Yates} = \sum_i (|O_i-E_i| - .5)^2/(E_i)$.	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used. The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment). The x^2 formula is very well known and was deleted from the document to avoid any confusion and to simplify the guideline. The x^2 formula used in EudraVigilance is correct.
Section 4.4	The fact that the term "chi-square" is being used suggests that this statistic has a chi-square distribution (at least approximately). Is there a reference for this? And what are the underlying assumptions? We again raise the question about whether this is truly an inferential exercise and recommend that wording be added that some of the description above may not fulfil the assumptions underlying statistical inference. Please add reference and or a description of the underlying assumptions	The disproportionality analysis is not an inferential exercise (i.e. the exercise is not aimed at drawing conclusions about a parent population on the basis of evidence obtained from a random sample from this population). The Chi-square statistics is an inexact test which results from the normal approximation of the Poisson distribution (see for example Altman DG. Practical Statistics for Medical Research. Chapman & Hall/CRC Texts in Statistical Science Series Volume 12. CRC Press Boca Raton 1990 or G Grimmett G, Stirzaker D. Probability and Random Processes third edition, Oxford University press. Oxford 2001).

GUIDELINE SECTION TITLE: 4.5. Interpretation of signals of SDRs		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 4.5, Title	The words '*signals of*' are redundant in the heading Amend to: '4.5. Interpretation of SDRs'	The comment was implemented.
Section 4.5, line 5-6 (a)	To ensure the consistency in the statement that PRR does not imply any kind of causal relationship as suggested for Introduction paragraph 3 line 5-6. Please delete "necessarily" and replace with "any kind of".	The comment was implemented.
Section 4.5 line 8 (a)	It is not correct to suggest that all SDRs should <u>always</u> be medically assessed. Many of the SDRs identified by this statistical method will be well known drug effects and already in the relevant SPC, and hence will not require further medical assessment. Please delete "always" in this section as well as in section 7.1 (Systematic evaluation of SDRs).	A drug-event pair provides only very limited information on a case report. In many instances, even if the drug-reaction term association is expected, other elements from the case may lead to the fact that the case is in fact unexpected (examples include a change in seriousness or a change in the outcome of the reaction, a signal in a different population, an off-label use, drug-drug interaction, etc …). Therefore, the fact that the nature of the reaction itself is expected may be insufficient to disregard the case as a signal. For that reason, all SDRs must be considered and assessed in the context of the case reports and therefore all SDRs should be subject to the medical evaluation.
Section 4.5 (a)	A number of biases that may affect PRR should be mentioned, including confounding by indication and the Weber effect.	These biases have been highlighted in the guideline.
Section 4.5, lines 3-4 (b)	*"b…There is no 'gold standard' on the thresholds that should be adopted for SDRs."* As the disproportionality analysis of spontaneous adverse event data is fundamentally an exercise in data mining, it is important to note that strict thresholds and absolute criteria are not necessary. Instead, reviewers should use the analyses provided as additional information to consider in their evaluation. <<NAME>> endorses the use of DPA scores as being part of the 'preponderance of the evidence' in the evaluation of drug-event pairs. We emphasize that just as drug-event pairs with high scores may not be 'true signals', low scores do not	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).

	reflect the absence of a signal. It bears noting that that the 'preponderance of the evidence' is at the heart of <<NAME>>'s signal evaluation process and at the center of the exploratory model <<NAME>> has developed in its use of observational data for signal detection and evaluation. The model enables safety scientists to explore the data and the relations between drugs and conditions, and integrates multiple disparate data sources and analytical methodologies to provide a full context around the effects of our medicine on real-world populations. Such an exploratory approach is intentionally not prescriptive, but is conducted on the basis of best practices for scientific investigation.	
Section 4.5, line 1 (b)	To ensure the consistency in use of the term "*a drug-event pair*". Here it says "*a drug-event combination*". Amend to "a drug-event pair".	The comment was implemented.
Section 4.5 line 3 (c)	If the threshold is too high, there is a risk of missing true signals (not 'potential signals'). Amend to: '…or missing true signals if this threshold is too high.'	The comment was implemented.
Section 4.5, lines 1-3 (c)	*"c. The thresholds commonly used to detect SDRs are a trade-off between two options"* The selection of the threshold does determine the sensitivity and specificity of the analysis. That is, a lower threshold may increase the total number of signals, which is likely to increase the 'false positive' rate but also decrease the 'false negative' rate, whereas a higher threshold that yields fewer signals may have less 'false positives' but more 'false negatives'. The key challenge to note, however, is that the true sensitivity and specificity can never be determined using spontaneous adverse event data alone. That is, we do not know 'truth' for which to evaluate the false positives and false negatives; a 'true' signal would be a drug that has a causal relationship to an event, but such causal relations can not be inferred from adverse event reports (let alone that the disproportionality of those reports). If effect, all drug-event combinations should be considered as candidate signals: disproportionality analysis can simply be a means of prioritizing those drug-event pairs, not eliminating the candidacy of pairs that don't meet a pre-defined threshold.	The validation study will provide more insight on the choice of the thresholds.

Section 4.5 (d)	The list presents cautions in the interpretation of results with the PRR method. This list should be expanded as a separate section on limitations of the PRR method. This is a method with limited capabilities, where many assumptions are necessary. Key reference publications on limitations to the PRR method are included at the end of this table of comments.	The guideline already discusses extensively the limitations of the PRR method. The PRR like the other quantitative methods based on disproportionality analysis work on very similar underlying assumptions (mostly concerning the mechanisms of reporting of adverse drug reactions). Consequently, the other methods mentioned in the references included at the end of the table of comments suffer from similar limitations due to the underlying assumptions. The studies comparing the performances of different quantitative methods need to be interpreted with caution. Further work aimed at better understanding the performances of the methods which have been (or will be) implemented in EudraVigilance will be performed.
Section 4.5, line 9 (e)	The issue is more about reliability of results if analysis is carried out in an undisciplined manner, not 'generalizability', *per se*. Amend to: '… results may not be reliable.'	The sentence was deleted.
Section 4.5 (e)	The comment that undisciplined analysis entails subject decision is true but subject judgement is what is being advocated (and rightly so) at the beginning of this paragraph – i.e. 'what is the medical context?' Suggest deleting the last 3 lines of text in this point. Even though they are true, they seem at odds with what else is being said.	The sentence was deleted.
Section 4.5 (e)	PRR calculation is based on the assumption of independence. When there is a hint that this assumption is violated because of public awareness or different reporting behavior, the PRR cannot be interpreted. For prioritization of SDRs, the potential pact on public health (event fatality, magnitude of reporting rate, etc.) should also be considered. List some of the biases that occur in spontaneous data analyses, e.g. public awareness. List some of the biases that might influence the PRR value: e.g. non-representative background, innocent-bystander phenomenon, etc.	The guideline already mentions the reporting artefacts as a factor which may dramatically influence the interpretation of the PRR (under- and over-reporting and uneven reporting between medicinal products). The prioritization of SDRs should be made on the impact of the signal in terms of Public Health. The innocent-bystander will not influence the PRR but induce the identification of a SDR which does not reflect the presence of a true signal but a drug association.
Section 4.5	We would like to see some text in this area indicating that the PRR is not as robust when case counts are small.	In certain situations particularly when the cell counts are low, an exact test should be performed (Fisher exact test). No adjustment for

	With the number of ratios being computed, there is a high chance of spurious results. How will they account for multiplicity?	multiplicity is performed with the PRR.
Section 4.5 (d)	Consider adding "- The frequency/timing and magnitude of database updates"	By definition the PRR changes as new reports are transmitted to EudraVigilance. The comment was not taken into account.
Section 4.5, (b)	To ensure the consistency in use of the term "**a drug-event pair**". Here it says "a drug-event combination". To change to "a drug-event pair"	The comment was implemented.
Section 4.5, line 2, (b)	To specify that "**pre-defined** threshold applied". The threshold should not be changed after To add "pre-defined threshold" as the same term is used in 4.5 (e).	The comment was not taken into account. A signal may be reviewed even if the value of the DMA does not exceed a pre-defined threshold (e.g. DME, TME). A medical judgement should also be used to decide when to review (or not) a SDR.

GUIDELINE SECTION TITLE: 4.6. Thresholds defining SDRs in EudraVigilance		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 4.6, line 4	This implies some threshold should be applied routinely. '… to assess if there are any thresholds that should be applied routinely…'	By definition the methods can only be used if some thresholds are defined.
Section 4.6, bullet points (a) and (b)	As noted above, it is not clear why some situations give chi-squared and others give confidence intervals. It would help if this was commented on (somewhere in the document).	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used.
Section 4.6, bullet points (a) and (b)	Bullets (a) and (b) will give raise different drug–event pairs depending on the (somewhat arbitrary) style of output. Even though the thresholds are arbitrary, this discrepancy is likely to lead to confusion and ambiguity. As the system 'colour-flags' the SDRs, why not use $x^2 > 3.84$? It doesn't need the human eye to search for 3.84 (or 4 or something else). More importantly, the requirement for PRR>2 could give very different sets of SDRs to the confidence interval method. Should not a PRR>2 also be included as a requirement in bullet (a)?	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used. Bullets (a) and (b) reflect these different screening criteria. In addition, it is not known at the moment what are the exact performances of the methods implemented in EudraVigilance. More insight and a better understanding of the performances of these methods will be gained with the conduct of the validation study. There is little practical difference between 3.84 and 4. The use of the lower bound of the 95CI is likely to improve the specificity of the PRR method (particularly when combined with a threshold for the number of cases). As mentioned in the guideline, no threshold for the PRR is used for the main weekly/monthly signal monitoring report.
Section 4.6, line 3-5	*"Further practical experience as well as formal validation studies are necessary to assess which thresholds should be applied routinely in the EudraVigilance Data Analysis System."* Because adverse event reports are submitted at a non-random rate from an unknown distribution of the general population, it is impossible to conduct a formal validation study to identify the 'correct' threshold. Simulation studies can approximate the reporting distributions, but Parks et al[10] have shown empirically that underreporting bias alone can cause significant signal misclassification. The primary value of setting	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).

	thresholds is simply to enable prioritization across drug-event pairs to ensure that limited resources focus on the signals of most interest. However, it is imperative, given that 'false negatives' will fail to reach the threshold, that all drug-event pairs to considered as part of the exploratory process. To this regard, <<NAME>> has actively endorsed an exploratory model for pharmacovigilance, whereby drug-event pairs are prioritized on the basis of disproportionality, but all pairs are considered through multiple views as part of the routine signal detection process.	
Section 4.6, bullet point (a)	*"a. When the PRR is displayed with its 95% confidence interval…The number of individual cases greater or equal to 3"* It is well understood that the PRR metric is unstable with small counts, particularly when cells A or C are small in Table 1. To a certain extent, the requirement that n>3 addresses some of this instability, but also eliminates many particular events of interest. Specifically, very rare but serious events, like Stevens Johnson Syndrome, may warrant consideration after only 1 or 2 reports. As such, it is not recommended to use any thresholds which require a specific number of cases before consideration.	In certain situations particularly when the cell counts are low, an exact test should be performed (Fisher exact test). Stevens Johnson Syndrome is a typical DME, the concept of DME and TME is described in section 7 of the guideline.
Section 4.6, bullet point (b)	*" b. When the PRR is displayed with the x^2 statistic"* The first threshold, using the confidence interval of the PRR, accounts for some of the variability inherent to the calculation of the reporting rate. It is unclear what, if anything, is gained by the addition of the x^2 statistic. Even when x^2 scores are displayed, it seems feasible to maintain the same (a) threshold, rather than constructing a different threshold which has a different interpretation and meaning.	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used.
Section 4.6, lines 1-2	This uses the term '*statistical signal'*. Paragraph 1, page 3, says that 'statistical signals' are referred to as SDRs. Amend to: '… that establishes universal thresholds for signals of disproportionate reporting.'	The comment was implemented.
Section 4.6, line 3	With regards to "… *formal validation studies are necessary*…", any validation work needs a gold standard to compare with. It should be	The detailed protocol of the validation study was not included in the guideline.

	suggested that what kind of 'gold standard' could be used for the validation studies.	
Section 4.6, lines1-2	With regards to the statement '… *on the entire EudraVigilance database*', does this also mean consumer reports? Are both serious and non-serious adverse events included?	The ICSRs contained in EudraVigilance result from the reporting obligations of Marketing Authorisation Holders in accordance with the community legislation (Directive 2001/83/EC, as amended and Regulation (EC) No 726/2004). Only both serious and medically confirmed ICSRs are transmitted to EudraVigilance.
Section 4.6	Please clarify the basis for these criteria. Was this to still determine a signal when counts are low.	The guideline was clarified that either a) or b) could satisfy the SDR criteria (the word "or" was added between a) and b)).
Section 4.6	Please make it clear that all the bulleted items need to be satisfied simultaneously	The guideline was clarified that either a) or b) could satisfy the SDR criteria (the word "or" was added between a) and b)).
Section 4.6,	Please explain why interventional clinical trials are excluded. What about reports from other types of studies?	Disproportionality analyses (DA) were primarily designed to be used on spontaneous reporting systems and not on organised methods of collection of reports such as clinical trials (taken in a broad sense either interventional or observational). Since the assumptions underlying the use of disproportionality analyses do not apply to any organised method of collection of ICSRs, these methods are not designed to be applied to data arising from clinical trials. Some methodological aspects relating to the conduct of clinical trials (e.g. blinding and randomisation schemes) make it difficult to perform any meaningful unplanned analysis without interfering with the integrity of the trial.
Section 4.6, bullet points (a) and (b)	It should be made clear that either a) or b) could satisfy the SDR criteria. Add "or" between a) and b).	The comment was implemented.
Section 4.6, bullet points a) and b)	It should be stated that rules a) and b) are equivalent. It should also be clear that the "*number of individual cases*" refers to the total number of cases reported for P involving an adverse event R, i.e. A.	The rules a) and b) refer to different disproportionality analyses methods and therefore cannot be considered to be equivalent.
Section 4.6, bullet points a) and b)	Section 4.6 discusses two thresholds for defining SDRs, one relying on the lower limit of a 95% Confidence Interval and the other on a chi-square and absolute PRR value. Concerning point a) on the 95% CI, it should be made clear what the	The definition of the SDRs using the lower bound of the 95CI of the PRR is not ambiguous in the guideline. In addition, some medical judgement (i.e. knowledge of the medical product and of the underlying disease) should also be exercised when using disproportionality analyses. In that respect, pancytopenia could also be considered to be a DME/TME. The

	interpretation of a CI is in the context where no samples are identified. Usually, a CI refers back to a parameter in a sample. The problem arises in an example from the Guideline in Table 2, page 10. We can see that the PRR for pancytopenia is 1.81 with a lower end CI of 1.31.It can be argued that with 30 reported cases and a PRR of less than 2.0, this is not a signal. A strong definition for the CI interpretation should enable identifying meaning signals.	medical evaluation of such reports might not rely exclusively on the use of data mining algorithms.
Section 4.6 bullet point (a)	"*The lower bound of the 95% confidence interval greater or equal to one*" seems to be inconsistent with concerns mentioned elsewhere about the issue of too many false positives. Please consider re-setting the lower bound of the confidence interval.	The thresholds implemented in EudraVigilance will be reviewed in the light of the experience gained with the system (i.e. validation study). The comment was not taken into account.
Section 4.6	The PRR and the chi-squared test are not exactly the same thing, so it is conceivable the two will give different results. Has any work been done on the consistency of read-outs from the two sets of thresholds proposed? In particular, it has been noted that if C=0 the PRR is set to 99.9, which automatically biases the result if we have, say, the first report of an event. The reasons for this are valid in this setting (new events are automatically flagged as a cause for concern), but the wide confidence interval means the second method here is the most likely way such signals would be detected. We would be happier about both being presented if some work was published on the similarity of read-outs.	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used.
Section 4.6	Reference to problem of not having a gold standard for the threshold determination is given but a retrospective analysis on the effect of using the selected threshold on different data sets could have been used providing more evidence to this point.	This comment is one crucial methodological aspect of the validation study conducted on EudraVigilance.
Section 4.6	In the described context, the upper limit of the confidence interval for PRR has no value. It does also not play a role in the considerations on thresholds in section 4.6. It is suggested that one-sided intervals are used throughout the guideline. Question is to change then to 95% one-sided or to stick to 97.5% one-sided approach.	Two-sided intervals are used.

Section 4.6	Unclear definitions of cases a) and b) of the SDR determination. Add 'and' to all the bullet points	The guideline was clarified that either a) or b) could satisfy the SDR criteria (the word "or" was added between a) and b)).
Section 4.6, (b)	In case of a 95% one-sided method, the chi-square threshold on 2nd bullet could be changed from 4 to 3.	Two-sided intervals are used.
Section 4.6	(a) and (b) both have the flavour of decision rules. What is known about their operating characteristics? Are there any supporting references? If it is unclear, or there are very different assessments of the characteristics, the document should clearly explain that the characteristics are unknown or may be situation-dependent. Please add supporting references for these rules.	Several studies aimed at assessing the operating characteristics (sensitivity and specificity for the chosen thresholds, ROC curves) of different quantitative methods of signal detection have been published (see for example the works published by Roux, Thiessard, Szarfman, etc …). Some limitations of these validation studies have been briefly discussed in the guideline. The rules implemented in EudraVigilance were mostly chosen based on the results of some of these studies but require further empirical validation.
Section 4.6	*"Further practical experience as well as formal validation studies are necessary to assess which thresholds should be applied routinely in the EudraVigilance Data Analysis System."* Because adverse event reports are submitted at a non-random rate from an unknown distribution of the general population, it is impossible to conduct a formal validation study to identify the 'correct' threshold. Simulation studies can approximate the reporting distributions, but Parks et al have shown empirically that underreporting bias alone can cause significant signal misclassification. The primary value of setting thresholds is simply to enable prioritization across drug-event pairs to ensure that limited resources focus on the signals of most interest. However, it is imperative, given that 'false negatives' will fail to reach the threshold, that all drug-event pairs to considered as part of the exploratory process.	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 4.6	*"When the PRR is displayed with the x^2 statistic"* The first threshold, using the confidence interval of the PRR, accounts for some of the variability inherent to the calculation of the reporting rate. It is unclear what, if anything, is gained by the addition of the x^2 statistic. Even when x^2 scores are displayed, it seems feasible to maintain the same (a) threshold, rather than constructing a different threshold which has a different interpretation and meaning.	Different groups have defined screening criteria in different ways. Some used the Chi-square and other the confidence interval of the PRR both methods are presented in EudraVigilance in order to allow either criteria to be used.
Section 4.6, Para 2, Line	To clarify that interpretation of SDRs requires not only thorough knowledge of both the data available and the statistical methods but also	The comment was implemented.

6-7	the medical understanding of the drug and the disease indication. To add medical/clinical knowledge and understanding of the drug and the indicated disease for interpretation of SDR.	
GUIDELINE SECTION TITLE: 4.7. Subgroup analyses and stratification		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 4.7, all lines of that section.	Subgroup analyses are referred to, but there is not much discussion, especially about the limitations. Since the idea behind the PRR analysis is that the background set of drugs forms a sort of 'average drug' pattern, care must be taken that reports from a sufficient number of drugs are present in the subgroup so that you don't lose this 'average' experience.	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 4.7	With regard to this section, the ability to perform meaningful analyses of targeted sub-groups is clearly of value to safety evaluators. However, in practice, the creation of sub-groups will inevitably limit the number of ICSRs included within the analysis to some extent. Later in the document it is quite correctly and clearly stated that "*the power of the statistical screening are likely to be limited with low numbers of ICSRs*". Whilst the minimum size of the data set required to perform a 'meaningful analysis' has not been fully validated, the addition of the sentence shown may be useful. Please add: "*When sub-group analyses are performed, it is important to consider whether the overall numbers of ICSRs available for analysis within the specific sub-group is adequate for meaningful statistical screening, as the power of the statistical screening is likely to be limited for sub-groups with low overall numbers of ICSRs.*"	The comment is acknowledged. It is difficult to provide specific guidance on the best way to perform sub-group analyses. The number of ICSRs is one aspect of the sub-grouping; other important aspects include the choice of medical products used to perform the comparison in the disporportionality analysis. Such analyses should be performed with some common medical sense and should be interpreted with the appropriate caution. The sub-grouping may also be very helpful to remove the ICSRs which induce a masking effect of the PRR.
Section 4.7	The role of stratification as an aid to reducing bias due to confounding should be emphasized. This would apply to all analyses, not just subgroup analyses.	Stratification will be part of the future releases of the EudraVigilance Data Analysis System.

Section 4.7	When performing subgroup analyses with small sample sizes, the likelihood for false positives will increase using PRR.	The comment is acknowledged. It is difficult to provide specific guidance on the best way to perform sub-group analyses. The number of ICSRs is one aspect of the sub-grouping; other important aspects include the choice of medical products used to perform the comparison in the disporportionality analysis. Such analyses should be performed with some common medical sense and should be interpreted with the appropriate caution. The sub-grouping may also be very helpful to remove the ICSRs which induce a masking effect of the PRR.
Section 4.7	It seems that much of the criterion for defining a signal comes from the limitations of the test. Why has the group not attempted to incorporate the EBGM method that seems to perform better when numbers of counts are small? Is this because the EBGM method is more difficult to incorporate into the EudraVigilance Data Analysis system?	The comment is acknowledged. The implementation of Bayesian methods will be considered in the future.
Section 4.7	We suggest to add the stratification by calendar year. Please amend to "…the computation of the static PRR will be adjusted for age, gender and calendar year by stratification…"	The comment was implemented.
Section 4.7	The role of stratification as an aid to reducing bias due to confounding should be emphasized. This would apply to all analyses, not just subgroup analyses. Performing subgroup analyses, however with small sample sizes the likelihood for false positives will likely increase using PRR. We support the use of stratification as opposed to subgroup analyses. The risk of accusations of "data dredging", however valid they may or may not be in this context, cannot be ignored.	Stratification will be part of the future releases of the EudraVigilance Data Analysis System.

GUIDELINE SECTION TITLE: 5. DESCRIPTION OF THE EV DATA ANALYSIS QUERIES		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5	This section reads as training and user manual because it gives general functionalities of the EudraVigilance Analysis System. What is the reason for including this into a guideline? Is the EudraVigilance Analysis System used only by EMEA employees or also others?	The section provides an overview of the main technical functionalities of the EudraVigilance Data Analysis System (incl. choice of the filters). However, this section puts the use of the quantitative methods in the practical context of EudraVigilance which was felt to be very useful to the users. More detailed practical guidance is given during the EudraVigilance Data Analysis System training course.
Section 5	As the system will get enhanced functions with its maintenance, it does not seem to be appropriate to update the guideline each time. How will this be handled?	The guideline is not intended to cover the full functionalities of the system. However, the guideline like any other scientific / regulatory guideline will have to be updated according to the scientific knowledge gathered on the use of quantitative methods of signal detection on EudraVigilance.
Section 5, para 1	With regards to "*To periodically generate data summaries and SDRs for reported medicinal products*", information on which medicinal products will be included in the data is not clear. This should be clarified	The scope of the guideline refers to all the medicinal products authorised in the EU (according to the provisions of Directive 2001/83/EC, as amended and Regulation EC No 726/2004).
Section 5, para 1, line 2	The term "*Data summaries*" could be more specific. What type of summary does it allude to?	These data summaries refer to the static reports implemented in the EudraVigilance Data Analysis system. Further practical information will be given in the EudraVigilance Data Analysis System training.
Section 5, para 1, line 7	The phrase "*Exploratory and descriptive manner*" is vague. It merits further explanation.	The phrase has been deleted and replaced by "This allows ad-hoc scientific evaluation of 'potential signals'". Further practical information will be given in the EudraVigilance Data Analysis System training.
Section 5, para 2	The bullet "*To conduct dedicated analyses on sub-sets of individual cases using different analyses filters. This allows ad-hoc investigation of 'potential signals' in an exploratory and descriptive manner*." gives the impression that analyses can be made within the EudraVigilance database rather than simply using it for signal generation.	This is true. Further practical information will be given in the EudraVigilance Data Analysis System training.
Section 5,	What are '*SDR functionalities'*? The phrase '*signal of disproportionate*	The comment was implemented.

para 2, line 1	*reporting functionality'* doesn't make any sense.	
Section 5, para 2, line 3	With regards to the phrase "…*the training and user material needs to be consulted*.", where are they? Will there be access to these documents by the MAHs? It will be helpful to put a footnote or a reference about where to get the training and user material.	Access policies to EudraVigilance are being developed and will be implemented in the future. The guideline will be changed (and will make reference to these access policies) when these access policies will be implemented. A EudraVigilance Data Analysis System training for all stakeholders is currently under-development, additional information on the training will be published on the EudraVigilance website. The training and user material will be handed-out during this training. http://eudravigilance.emea.europa.eu/human/evComDataAnalysisSystem.asp
GUIDELINE SECTION TITLE: 5.1. Standard outputs of SDRs		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.1, para 1	Which cases will be included in the output? Will any cases that are considered signals be included in the output? What levels (PT, HLT, etc.) will be used in the reports? To whom this output will be sent? An explanation (or a reference) of '*suspect'* and '*interacting*' would be helpful.	The ICSRs transmitted to EudraVigilance according to the provisions included in Directive 2001/83/EC, as amended and Regulation EC No 726/2004 are included in the queries (post-authorisation clinical studies are excluded). The SUSARs transmitted to EudraVigilance in accordance to Directive 2001/20/EC, as amended are not included in these reports. Therefore, post-authorisation ICSR potentially involved in a signal will be included in these reports. The levels of MedDRA used in the report from Table 2 cannot be changed. The analysis is performed at the PT level and grouped at the SOC level. The appropriate access given to EudraVigilance will be detailed in the EudraVigilance access policies. Definitions and additional guidance of the terms used in the guideline are available in the Volume 9A of the Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use.
Section 5.1, para 1	The guideline indicates that 'Reaction Reports' (shown in Table 2) are generated on a weekly or monthly basis. It is possible that this will be unnecessarily frequent for the majority of products. A short time since the previous report will result only in insignificant changes in PRR values for most drug-event combinations, and in small numbers for 'New EEA' and 'New non-EEA' events all of which will be difficult to	The comment has not been implemented

	interpret. Suggest changing the frequency of new reports generation by default to once every three months for most drugs, with the exception of products for which weekly or monthly numbers of new reports involve hundreds of cases.	
Section 5.1, para 1	SDRs are duly reported in the 'Reaction Monitoring Weekly' report. However section 4.6 highlights that there is no gold standard for the threshold defining SDRs. Then, especially since readers of Reaction Monitoring Weekly will likely not have access to all the data to understand all the issues outlined in section 4.5, the list of SDRs published may only cause panic in certain patients/providers. This is one of the main criticisms of PRR, that the method tends to inflate the disproportionality, especially compared to EBGM and other methods. Access to "Reaction Monitoring Weekly" should be restricted to people having full access to all the data and having relevant competencies to interpret the results.	The appropriate level of access to EudraVigilance will be defined in the EudraVigilance access policies. The guideline will be updated after the implementation of these policies. The necessary training will also be given to make sure that the SDRs are appropriately interpreted, assessed and put in the context of other information available (including the information contained in the terms of the Marketing Authorisation). The guideline details and emphasises the elements which must be taken into account when interpreting the SDRs.
Section 5.1, para 1, line 6,	The phrase '*are taken into account*' implies some sort of analysis. Simpler text might help. Amend to '… are included.'	The comment was implemented.
Section 5.1, para 2, line 8	The label for lower and upper bounds is not obviously understandable. Change to 'PRR LB' or 'PRR 95LB' for the lower bound.	The guideline has been clarified.
Section 5.1, para 2, line 17	What does '*all relevant cases reported to EudraVigilance*' mean? Please clarify.	The ICSRs transmitted to EudraVigilance according to the provisions included in Directive 2001/83/EC, as amended and Regulation EC No 726/2004 are included in the queries (post-authorisation clinical studies are excluded). The SUSARs transmitted to EudraVigilance in accordance to Directive 2001/20/EC, as amended are not included in these reports. Therefore, post-authorisation ICSR potentially involved in a signal will be included in these reports.
Section 5.1, para 2, line 18,	What does '*fatal'* cases mean? What is the difference from '*new fatal'* cases?	Fatal cases refer to all the ICSRs included in the report for which the seriousness criterion as been flagged at death. The "new fatal" cases, refer to the cases for which the seriousness criterion was death during the

		period covered by the report.
Section 5.1, Table 2	Table 2 and the explanation on previous pages are hard to follow. Please consider putting the table first followed by the explanation.	The comment was implemented.
Section 5.1, Table 2	It should be mentioned that the PRR as calculated in the standard output (table 2) is based on Total reports.	The comment was implemented.
Section 5.1, Table 2	It seems that the standard output (section 5.1, table 2) comprises more information than those outputs referred to under section 5.2 "Exploratory functionalities". What is the rationale?	A report similar to the standard output is available in the EudraVigilance Data Analysis System and can be customised by the user.
Section 5.1, line 1	Weekly/Monthly Quarterly may be sufficient	The comment was not taken into account.
section 5.1, line 5	The term interacting is again used here, and somewhere there needs to be more explanation about what that means. Don't use 'interacting' except for statistical interactions.	The guideline will clarify and specify the meaning of the word "interacting" as appropriate whether it refers to a drug-drug interaction or a statistical interaction.
section 5.1, lines7-9	I assume that the PRR is calculated from the cumulative numbers. Would make sense to clarify this	The comment was implemented.
section 5.1, lines9-11	Why report EEA and non-EEA reports when the numbers are not accurate and do not add up to the total number of reports? Report Total data.	It is important to assess the potential Public Health impact of the signal in the EU. Therefore, the country of origin of the reports is an important element to take into consideration in the prioritisation of the evaluation of the signals.
Section 5.1, line 6	The phrase '… are taken into account' implies some sort of analysis. Simpler text might help. '… are included.'	The comment was implemented.
Section 5.1, Table 2	Some knowledge about the joint distribution of the PRRs would be extremely valuable.	As mentioned above in the document, we want to emphasise that such distributional assumptions cannot be reliably adopted and checked in the context of SRS data.
Section 5.1, Table 2	Table 2 and explanation on previous pages are hard to follow. To put the Table first followed by explanation.	The comment was implemented.
Sections 5.1 and 5.2	Sections 5.1 and 5.2 suggest the potential for estimating adjusted or stratified PRR (by age, gender, country etc.) Such estimates would be	Stratified analyses will be implemented in a second phase and will require additional validation.

	more efficient than the crude PRRs as often SDR can be explained by differences in the distribution of such factors. It might be considered to implement stratified PRR as routine output.	
GUIDELINE SECTION TITLE: 5.2.Exploratory Functionalities		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2	The limitations regarding the quality of spontaneous data should be taken into account, especially when looking at concomitant medication. Is would be highly desirable if MAHs (e.g., Qualified Person) were in the position to perform analyses on their own for their products.	The limitations of spontaneous reporting system and the importance of data quality are emphasised in the guideline. The EudraVigilance access policies will define the conditions of access to EudraVigilance by the Marketing Authorisation Holders.
Section 5.2, line 4	In the 2nd sentence, why should the variables mentioned be *"…indicative of the medical condition of the patients"*?	The comment was implemented.
Section 5.2, line 1	'exploratory functionalities' of the analysis system. Several filtering options are of interest for MAHs, e.g. customisation of MedDRA term selection	The user can also choose the criteria to filter the output, which will be produced by the query. A filter can be applied e.g. at any level of the MedDRA hierarchy from System Organ Class (SOC) to PT level as well as Standard MedDRA Queries (SMQs). The query can also be customised to select specific MedDRA terms and medicinal products.

GUIDELINE SECTION TITLE: 5.2.1. Standard query options		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2.1, para 1, line 9	Change "Preferred Term" to "PT" for consistency.	The comment was implemented.
Section 5.2.1, para 1, lines3-6	*"The user can select as query parameter the appropriate medicinal product hierarchy level (EudraVigilance Medicinal Product Dictionary) and the MedDRA hierarchy level. Results are calculated on the basis of the specified level e.g. if the Preferred Term (PT) is selected, then all statistics are calculated at PT level."* It is recommended that the document specifically outline the implementation of hierarchy level aggregation to ensure that users properly interpret the resulting metrics. In particular, it is recommended that algorithm be presented to outline when terms are re-classified and at what stage calculations are made, so as to address potential concerns about double-counting and biases that can be introduced through aggregation. The query approach outlined in this guidance is an important first step to enable data exploration, but there are significant consequences in its limitations. While the EVMRD provides one classification for drugs and MedDRA provides one classification of adverse events, it is quite common that users may require reports to be classified in more general terms that do not fit these schemes. <<NAME>> commonly reviews SMQs, 'grouped' MedDRA terms that do not follow the MedDRA hierarchical structure but are clinically relevant. While it can be desirable to aggregate all data to a particular level of the hierarchy, it can be more valuable to aggregate only parts of the hierarchy for purposes of analysis. As an example, it may be desirable to look for events for all Cox-2 inhibitors, as compared to all other drugs; this analysis would require aggregating EVMPD terms for Vioxx, Celebrex and Bextra to the drug class 'Cox2', but leaving other	The comment will be taken into account (when the query functionalities using grouping of MedDRA terms and medicinal products will be implemented).

	reports disaggregated. <<NAME>> has tools that enable the construction of these custom terms to use in exploratory analyses of spontaneous data. As part of its efforts in leveraging observational data, <<NAME>> uses large biomedical ontologies to its exploratory analyses. These ontologies enable analysis across disparate data sources using a common language. They also facilitate aggregation to meaningful concepts at varying levels of generalization to meet the needs of the analysis.	
Section 5.2.1, para 2, line 10	It is unclear why terms such as "*Sponsor study number*", "*study reports*" are mentioned here as the whole scope of the quantitative method is for spontaneous reports. AE reports from studies are not part of the scope. Consider deleting these terms.	The signal detection methods described in the guideline should not be used to perform some signal detection on the SUSARs received from clinical trials in accordance with the Directive 2001/20/EC, some queries implemented in the EudraVigilance Data Analysis System have been implemented to support the analysis of the information transmitted to EudraVigilance in the context of interventional clinical trials.
Section 5.2.1, para 2	Information may not always be available for these filters. Should mention that since filters may not always have the designated information (missing data), the more filters used the more data likely to be excluded.	The filters will select the ICSRs for which the field was populated by the reported information. Adding more filters always implies that more data will be excluded by the analysis. It is also possible to select "Not specified" for the value of the field. The comment also emphasises the importance of data quality in signal detection (see paragraph 9).
Section 5.2.1, para 2	With regards to "The type of ICSRs (spontaneous reports, study reports)", isn't "non-interventional study reports" more correct?	No. "Study reports" covers reports from clinical trials whether the study is interventional or not.
Section 5.2.1, para 2	What is "*Sponsor study number*"?	The Sponsor study number is the number of the study as it was identified by the sponsor of the clinical trial. The EudraCT number in the format YYYY-NNNNNN-CC# should be inserted at the start of the ICH E2B M2 message field, A.2.3.1 Study Name. It is possible to query on the basis of the EudraCT number.
Section 5.2.1, para 2	An additional filter "*calendar year*" should be listed.	The comment is acknowledged (this functionality will be considered for future releases). However, the calendar year is too imprecise, to which E2B(M) field this calendar year would refer to?
5.2.1, para 2	Additional filters should be available to user:	The queries can be performed using different identifiers for the medicinal

	• Tradename (manufacturer) • Product alias (a synonym name linking all identical products regardless of tradename and manufacture). • Indication	products (e.g. INN, invented name, presentation level, etc …).
Section 5.2.1, para 2	*"Other filters available to the user are as follows:"* <<NAME>> actively considers the time of event as part of its routine pharmacovigilance, and creates analyses that partition reports during specific time periods and plot DPA scores over time. Because of the history bias associated with media coverage that is inherent to spontaneous adverse event reporting databases, it is important for the specific time period of the analysis to be considered.	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 5.2.1, para 3, line 16	"drilling functionality" Please clarify if this is to patient level, showing co-med/co-morb. (as in AERS)?	The drilling functionalities will be addressed in detail during the EudraVigilance Data Analysis training.
Section 5.2.1, para 3-4	Who will use the 'drilling functionalities'? What is the content of the drilling functionalities (CIOMS level?)? Please clarify if this is to patient level showing co-medication/co-morbidity, as in AERS? Shouldn't this be part of a user manual? 'Drilling' suggests going down into finer detail, yet the example given is about moving to a higher level.	The drilling functionalities will be addressed in detail during the EudraVigilance Data Analysis training.
Section 5.2.1	In the penultimate paragraph, 'drilling' suggests going down into finer detail, yet the example given is about moving to a higher level Please re-word	The comment was implemented.

GUIDELINE SECTION TITLE: 5.2.2. Static PRR Table		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2.2	This section seems redundant. The document is supposed to be describing the statistical methods, not training on how to use the system and what the different output means. A formula for chi-squared has previously been given (in section 4.4). This section repeats that formula (which isn't necessary) and then gives an alternative, more complex version of it (again, unnecessary). This more complex version then needs a substantive footnote to explain some of is terminology. The explanation of how to get A, B, C and D from the table (with A, A+B, A+C and A+B+C+D) is computer/system, training, and not part of the 'methods'. Delete this section. Alternatively, it might be moved to an Appendix.	The x^2 formula is very well known and was deleted from the document to avoid any confusion and to simplify the document. The x^2 formula used in EudraVigilance is correct.
Section 5.2.2	The 2nd chi-sq formula is wrong. It should be the sum of $(Obs - Exp)^2 / Exp$ components. In any case, why does the formula have to be expanded in this tedious way? It is conventional to present in terms of (Obs - Exp) rather than (Exp - Obs).	The x^2 formula is very well known and was deleted from the document to avoid any confusion and to simplify the document. The x^2 formula used in EudraVigilance is correct.
Section 5.2.2, table 3	In table 3, the lower and upper 95% confidence limits, PRR(-) and PRR(+), are slightly wrong. They should be 0.25 and 2.47, respectively.	The way the CI is computated will be checked.
Section 5.2.2, equation 2, and footnote 4	The x^2 formula is not correct. The denominators 'N' must be replaced by the expected values per term. $x^2 = (ExpA-ObsA)^2/ExpA + (ExpB-ObsB)^2/ExpB + (ExpC-ObsC)^2/ExpC + (ExpD-ObsD)^2/ExpD$	The x^2 formula is very well known and was deleted from the document to avoid any confusion and to simplify the document. The x^2 formula used in EudraVigilance is correct.
Section 5.2.2,	Delete the 3rd sentence: "*The expected value of the cell A….by the total number of reports*" because it does not provide any further information than the following sentence, "*Expected value for A is ….*"	The comment was implemented (section 5.2.2 was simplified and clarified). There is some degree of uncertainty surrounding the computation of the PRR (different formulae have been used by different

Footnote	Please delete as suggested.	persons). Therefore, this section has been added for two reasons: it avoids any misunderstanding on the actual computation of the PRR in EudraVigilance. Secondly, this section of the guideline had been added to support the training of the users. Relevant references to statistical textbooks have been added.
Section 5.2.2, Table 3	There is a "*" in the first line for "*Metrics*", but there is no footnote for this. Please clarify.	The typo was corrected.
Section 5.2.2, Table 3	The high precision of Expected A, B, C, D is not necessary.	This high precision may be useful for very low expected values (see Table 3). We are not able to modify this feature in the system.
Section 5.2.2, Table 3	Table 3 has a lot of numbers in it. Do any of them have meaning? How are they to be interpreted? To consider modifying the table accordingly.	The comment was implemented (section 5.2.2 was simplified and clarified). There is some degree of uncertainty surrounding the computation of the PRR (different formulae have been used by different persons). Therefore, this section has been added for two reasons: it avoids any misunderstanding on the actual computation of the PRR in EudraVigilance. Secondly, this section of the guideline had been added to support the training of the users. Relevant references to statistical textbooks have been added.
Section 5.2.2, Table 3	Wwhat is the significance of producing the individual components of the chi-squared test? For me, they provide a source of confusion to non-statisticians and only complicate the issue.	The comment was implemented (section 5.2.2 was simplified and clarified). There is some degree of uncertainty surrounding the computation of the PRR (different formulae have been used by different persons). Therefore, this section has been added for two reasons: it avoids any misunderstanding on the actual computation of the PRR in EudraVigilance. Secondly, this section of the guideline had been added to support the training of the users. Relevant references to statistical textbooks have been added.
Table 5.2.2, Table 3	Reference to the data set A clear reference to exclusions, subset, and subgroup could be useful in all tables.	The comment is acknowledged, this information can be obtained in the "report filter" which is accessible via the EudraVigilance Data Analysis System.
Section 5.2.2, Footnote	The term "expected value" isn't being used in the usual, commonly accepted sense. This seems to suggest that the expected value of a statistic is another statistic (and not a function of the parameters in a statistical model)? Perhaps "estimated expected value" is what is really	The wording is similar to the wording used in statistical textbooks.

	meant.	

GUIDELINE SECTION TITLE: 5.2.3. Static PRR Report		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2.3	This section is much more helpful than section 5.2.2 and serves as a good example of the results one might get back from running the system. No change	N/A.
Section 5.2.3	We suggest combining this section into section 5.2.2.	The two sections cannot be combined because they reflect different statistical reports available in the EudraVigilance data analysis system.
Section 5.2.3	We assume that the use of PRR >2, x^2>4 or A>3 being used as a criterion is an effort to take into account PRR's poor performance characteristics when counts are low. Is this correct?	These thresholds are arbitrary but have been used in previous studies using the PRR (see Validation studies).
Section 5.2.3, Table 4	Almost all of the numbers of cases given in Table 4 are small Please provide a better example table.	The Table 4 provides a good "real life" example where the actual number of SDRs is low. The majority of the drug-event pair in the pharmacovigilance databases is populated with a very low number of reports.
Sections 5.2.3-5.2.5 Tables 4-6	The red color for highlighting PRR values for drug-event combinations that hit the defined threshold is used in all tables but the meaning of the red color is only described for table 6. In table 4, different types of red color are used without explanation.	The comment has been implemented and the guideline has been clarified.
GUIDELINE SECTION TITLE: 5.2.4. Graphic PRR Monitor (CHI^2)		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2.4	Again, this section is much more helpful than section 5.2.2, giving a good example of the results one might get back. The different size of the 'bubbles' could be commented on. Add text: 'The size of the bubbles indicates….'	The comment was implemented.
Section 5.2.4	We are not convinced that there will be additional value added by generating this type of report/graph. It doesn't provide additional	The comment was not implemented.

	information to what is presented in Table 4 (Static PRR Report) and only a small number of SDRs can be displayed without affecting the usability of the screening results. Suggest removing this report from the list of available standard outputs.	
Section 5.2.4, line 5	Replace "*numbers of individual cases*" by "numbers of individual cases involving the adverse event of interest" Amend to: "numbers of individual cases involving the adverse event of interest".	The comment was implemented.
Section 5.2.4, Table 5	The size of the bubbles in the figure seems to be proportional to the number of ADR cases. If so, it would be better to state it clearly so as not to get it confused with the width of the confidence interval.	The comment was implemented.
Section 5.2.4, Table 5	Bubble plots should be constructed with area proportional to number, not diameter, because it is the area that will be visually assessed by the reader.	The feature is inherent to the system.
Section 5.2.4, Table 5	Displaying the bubbles dependent on number of cases might be impractical, if the cases numbers are big. Presentation of confidence intervals of PRRs would give more information (in a e.g. forest plot).	The forest plot is not available; the confidence interval is graphically displayed on 5.2.6 which has been added to the guidance.
Section 5.2.4, Table 5	It is recommended to mark/highlight the section defining SDRs.	The comment is acknowledged. Unfortunately this is not possible with the current system.
Section 5.2.4, Table 5	Compared to Table 4, the graph does not provide any additional information. It is actually confusing. Also, the fact that it is a log scale should be mentioned. We suggest that it could be deleted.	Table 4 was not deleted from the guidance but the scale will be explained in the guideline.
Section 5.2.4, Table 5	Graphic representation The figure seems uninformative- the table is sufficient.	The comment was not implemented.
GUIDELINE SECTION TITLE: 5.2.5. Static PRR Monitor		
Line no. + para no.	**Comment and Rationale**	**Outcome**

Section 5.2.5	Again, much more helpful just to present this example, than all the notational details in section 5.2.2. No change	N/A.
Section 5.2.5	The point estimate of PRR is as much a measure of association between a medicinal product and the occurrence of a specific event as is its confidence interval. Please amend the 2nd sentence to: "The report highlights (in red) any PRR for which the PRR $\geq$ 2 and the lower bound of the 95% confidence interval of the PRR $\geq$ 1 and the number of ICSRs $\geq$ 3."	The comment was not taken into account (see guideline section 4.6 bullet point b).
Section 5.2.5, Table 6	One column showing the PRR would be sufficient. Can also primary and secondary MedDRA paths be shown?	The comment was not taken into account since the report uses all the levels of MedDRA hierarchy (from SOC to PT).
5.2.5, Table 6	The table is hard to read. To make a better example table	The comment is acknowledged unfortunately we have used a "real life" example.

GUIDELINE SECTION TITLE: 5.2.6. Dynamic PRR Report

Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 5.2.6, final para	With regards to the statement *"Initial results on the use of the dynamic PRR are described in the scientific literature [14]."*, the limitation of the similar data does not appear in the guideline. It is known that in the FDA AERS database with over 43 million possible drug-AE pairs, less than 7% of these potential combinations have been reported, and half of these only appear once. The sparseness of this database, and presumably that of EudraVigilance as well, will lead to unstable estimates no matter what method is employed. The limitation of similar data and unstable estimates should be identified and highlighted.	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).
Section 5.2.6, last	The last sentence provides only one reference to the scientific literature on the use of "*the dynamic PRR*". However, the referred publication is a	The comment was implemented.

sentence	conference abstract and not a peer-reviewed article. Please provide additional references on the potential uses of dynamic PRR reports and guidance on PRR trend interpretation.	
Section 5.2.6, Table 7	It should be noted that the PRRs per month and year are cumulative PRRs.	The comment is acknowledged.
Section 5.2.6, Table 7	It is recommended to mark/highlight the section defining SDRs.	The comment is acknowledged. Unfortunately this is not possible with the current system.
Section 5.2.6, Table 7	The display of confidence intervals is unclear. The central line should be more prominent, and the outer lines less so, preferably without the dots marking the points, which are unnecessary in the outer lines.	The comment is acknowledged and will be taken into account for next releases of the system.
Section 5.2.6 Table 7	*The axes should be labeled.* "PRR and 95% CI" for Y and "Time" for X	The comment is acknowledged and will be taken into account for next releases of the system.
Section 5.2.6 Table 7	Identification of the PRR line is not clear: "*The PRR is indicated by the dotted line in the middle of the graph. The two other …*". It is hard to see the difference in the pattern of the line. More distinctive patterns (and colours) should be used for PRR and the 95% confidence interval.	The comment is acknowledged and will be taken into account for next releases of the system.
Section 5.2.6, Table 7	- Unclear display of confidence intervals; The central line should be more prominent, and the outer lines less so, preferably without the dots marking the points, which are unnecessary in the outer lines.	The comment is acknowledged and will be taken into account for next releases of the system.
GUIDELINE SECTION TITLE: 6. VALIDATION STUDIES ON STATISTICAL METHODS FOR SIGNAL DETECTION		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 6, line 5	*"The following conclusions can be drawn from these validation studies:"* As stated previously, spontaneous adverse event reporting databases have no external validity. The various methods proposed can impact the precision of the reporting rate estimate, but can not make the results generalizable to a different population. Moreover, given the limitations	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment).

	of the data, it is not possible to conduct a validation study to determine the appropriate thresholds. The threshold will only determine the number of signals to be flagged, which will impact the sensitivity and specificity indirectly, but will not impact the degree of evidence to evaluate the causal relationship.	
Section 6	Also in paragraph 6, it is unclear the consequence of lack of validation studies for statistical methods for signal detection. A rational for not having any requirement regarding minimum sample size of cases for the analysis could be provided.	Some validation studies of the current methods have been conducted but these studies have some limitations which are not discussed in the guideline (for example they were conducted on the basis of an unclear definition of what constitutes a signal, other studies did not use any case counts No, etc …). These limitations make it difficult to understand how these methods perform in practical in real situation. A performance study will be performed on EudraVigilance. This study will provide a better understanding of the performances of the quantitative methods compared to the "traditional" methods of pharmacovigilance when the two methods are applied, in parallel, independently.
Section 6	Interpretation Reference the following: Use for comparing drugs Inference about causality Interpretation when the number of cases is small Consistency across methods	The use of disproportionality analyses to compare the safety profile of different medicines should be used with extreme caution since the reporting rate of adverse reactions may be completely different across products. The disproportionality analysis is not an inferential exercise (i.e. the exercise is not aimed at drawing conclusions about a parent population on the basis of evidence obtained from a random sample from this population), therefore the comment on inference about causality was not taken into account. When considered necessary, the SDRs should be medically confirmed to assess the "strength" of the signal (i.e. the level of scientific evidence in favour or against a causal relationship between the administration of the medicinal product and the occurrence of the reaction). The PRR may artificially be high when the number of cases is very low (a case count threshold of 3 reports is used in the guideline). In circumstances when the No of report is likely to be very low (orphan drugs, special populations of patients) traditional methods of pharmacovigilance may be used. The consistency across methods relates to the performances of the

		different methods of signal detection. Please refer to the validation studies conducted on the different methods (out of the scope of the guideline).
Section 6	The issues outlined here are critical to this and any other "signal detection" system - to have only two paragraphs devoted to this, and not to provide any guidance on how to deal with and handle these issues, seems insufficient. This should be expanded to suggest strategies for dealing with the issues described.	The use of other quantitative methods than the methods implemented in EudraVigilance is out of the scope of the guideline. The performance of these methods but also the performance of the traditional methods need to be better understood and further work should be (and will be) performed in that direction in the future.
Section 6, a) & b)	It would be useful if the guideline could discuss the trade-off between these two considerations (sensitivity and specificity). For example, could one increase the threshold to 3 or 4 (instead of 2) to reduce the false positive rate? The thresholds should not be set in a biological vacuum, and there should be some recognition that thresholds will necessarily vary according to the importance of the events.	The comment is acknowledged. The guideline discusses this trade-off under section 4.5 interpretation of SDRs. The validation study will provide more insight in the choice of thresholds.

GUIDELINE SECTION TITLE: 7. INTEGRATION OF STATS METHODS WITH …		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 7	We would like to see some comments around the problems with using spontaneous report data for signal detection. The lack of denominator, duplicates, biases depending on market conditions, descriptions of the actual AEs, under and over reporting issues make using these data very unreliable. Only by using additional data sources and medical evaluation can signals be fully evaluated. The problem with the data tends to be much more of an issue than the analysis method being used.	The limitations of the spontaneous reporting have been discussed and highlighted in different parts of the guideline (introduction,
Section 7, para 4, line 6	What are the "*traditional methods*"?	The term "traditional methods" of pharmacovigilance is used in opposition to the quantitative methods to identify the methods which were (and are still) conventionally used before the recent introduction of the quantitative methods.
Section 7, para 4, last line.	Poor grammar. Please amend to "…such an approach…"	The comment was implemented.
Section 7, para 5	If an event is reported for only the drug but no other drugs, then chances are there are few reports of the event at all. Setting the PRR arbitrarily to a high value in this circumstance could be misleading, since the uncertainty about the true relative risk is so great that no reliable value of PRR can be asserted. For important risks, DMEs or TMEs can be identified and would be followed up regardless of the number. (The author of the comment requested ")To illustrate this scenario".	It is difficult to illustrate the scenario without giving a real example The concept of DME/TME is discussed in the guideline
Section 7, para 5, line 20	Poor grammar. Please amend to "…the power of the statistical screening is likely…"	The comment was implemented.
Section 7	Guidance on the extent of SDR documentation for various types of events is needed. Similarly, situations when the result of SDR evaluations by MAH should be communicated to other stakeholders	It is considered difficult to provide some additional guidance since many different scenarios may occur. The communication of the signal is out of the scope of this guideline and is addressed in other Community

	need further clarification.	pharmacovigilance guidelines (see Volume 9A of the Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use).
Section 7, para 5, line 18	*"When a medicinal product is new to the market and only a small number of ICSRs has been received, it is feasible and probably more appropriate to review these ICSRs individually than to rely on statistical methods."* Statistical methods on spontaneous adverse event databases may be inappropriate for use with newly marketed medicines. <<NAME>> is developing methods for leveraging observational data to evaluate drugs that are newly marketed by evaluating on the temporal relations between the drug use and incident events, using large comparator cohorts to estimate the relative risk increase of the new medicine. These methods should provide valuable supplement to the pharmacovigilance process in these settings where spontaneous adverse event data alone may not be sufficient.	The comment is acknowledged.
Section 7, last sentence	Consider changing the last sentence to read "...opportunities to *potentially* improve the statistical data analysis."	The comment was implemented.

GUIDELINE SECTION TITLE: 7.1. Systematic evaluation of SDRs		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 7.1, para 4, para 1	Since the text in italics is such a strong statement (beginning 'There is scientific consensus…') it might be worth supporting with published reference(s). Add (or refer to existing) published paper(s)	The statement was slighty qualified since there are different interpretations of what constitutes a medical evaluation. The SDRs should be considered with a medical judgement reflects the current scientific consensus on the management of SDRs. An appropriate reference was added.
Section 7.1, Diagram 1	'systematic evaluation of SDRs'. Diagram 1 visualises the process where communication with relevant stakeholders is the *final* step. Other steps in the above mentioned diagram 1 include a review of the marketing authorisation (SPC, PIL) with regard to an SDR and different types of documents submitted to the Agency, e.g. PSURs, SUSARs, Risk Management Plans. GEHC would suggest that the guideline identified a potential role of a concerned MAH in relevant steps of the evaluation process.	The comment is acknowledged, the communication and the involvement of the Marketing Authorisation Holder is addressed in other Community pharmacovigilance regulatory guidelines (see for example Volume 9A of the Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use).
Section 7.1, Diagram 1	"*Identification of potential duplicates*" and "*data quality checking*" should be the first step. Although the footnote (1) mentions "*routine duplicate and data quality checking*", this step is very important and should be mentioned in the process as the first step.	The comment is acknowledged, this step may not be necessary if the SDR is considered to be expected.
Section 7.1, Diagram 1	It appears that the term '*MAH'* is mentioned for the first time here – use of the full term (Marketing Authorisation Holder) would be better.	The comment was implemented.
Section 7.1, Diagram 1	The guideline suggests that all steps of SDR review should be fully documented and communicated to relevant stakeholders. From the reviewer experience, the number of SDRs, especially for mature products, could be quite large with many of them representing known/expected patterns of reporting (e.g. events related to underlying illness for which the drug was used). It is not clear to what extent should those types of events be documented and whether the results of their review need to be communicated.	A reference to internationally agreed quality standards is made.

Section 7.1, Diagram 1	Prior to the flowchart diagram (section 7.1, para 1) it is stated that **"***signals of disproportionate reporting identified with statistical methods should always be medically assessed***".** It is also stated that the steps outlined in the diagram provide **"***guidance on the reviewing process of SDRs, which relies on medical judgment***".** However, it isn't clear within the process when initial medical/scientific judgment is to be applied. When the analysis is being conducted by the EMEA, would this initial judgment take place before communicating the SDR to the MAH for further evaluation or action? Following the flow diagram through its various steps, it might appear that <u>any</u> SDR, which is either: i) Not based on duplicate or poor quality ICSR data, or ii) Is not addressed by the SmPC other documentation supplied by or agreed with the MAH (e.g. PSURs, RMPs, post-marketing commitments, etc) could be 'communicated' to stakeholders (i.e. the MAH) for further evaluation/action (and presumably a formal response would then be required). Generally documentation supplied by or agreed with the MAH (e.g. SmPC, PSURs, RMPs, post-marketing commitments) tends to focus upon already identified risks (either potential or known). Risks that have, quite reasonably, been excluded from targeted review (e.g. false positive signals, spurious signals, or those with clear alternative causal associations) are not routinely described in such documents. Hence an assessment of the information available within these documents would only be likely to identify those SDRs that are already 'identified' as potential/known risks.	The guideline emphasises the importance of a systematic medical evaluation of the SDRs for several reasons: - There is no relation a priori between the presence of a SDRs and a causal relationship between the administration of a medicinal product and the occurrence of the reaction. - The PRR like other methods may generate some "false positive" (i.e. SDRs which are considered not to be true signals during the signal evaluation / confirmation phase. - Depending on the chosen thresholds, the number of false positive signals may be significant. Therefore, before any communication takes place between the stakeholders involves in the process, it is important that the person who detected the SDR takes the appropriate steps to evaluate whether the SDR may or may not be a "true signal". This evaluation step and the relevant methods are described in the Community legislation and guidance documents. Once this evaluation has been performed and once the SDR is confirmed to be a possible signal, the communication between stakeholders should follow the rules and procedures described in the Community legislation and in the guidance documents (see for example Volume 9A of the Rules Governing Medicinal Products in the European Union: Pharmacovigilance for medicinal products for human use).

	The chosen method of analysis (PRR) is known to be associated with the generation of significant numbers of false positive 'signals', particularly in situations where the total number of ICSRs is low. This is acknowledged throughout the document. Therefore, there is the potential for a considerable proportion of SDRs generated via this system to actually be false positive signals. Is there an intention for SDRs to undergo an additional (initial) step of medical/scientific evaluation by the body conducting the analysis (e.g. the EMEA), prior to communication to the stakeholder (i.e. the MAH), to assess whether the SDR is likely to represent a safety topic requiring further detailed evaluation (i.e. there are reasonable grounds for assuming possible causality). Or is this evaluation to be performed entirely by the MAH following the communication of essentially a 'raw' list of SDRs that are not described within existing documents? If the stakeholder (i.e. the MAH) is to be requested to perform the medical/scientific evaluation of <u>all</u> communicated SDRs, without any initial medical/scientific pre-screening by the body conducting the analysis, are there any acceptable criteria for prioritizing the order in which SDRs are subsequently addressed? Assuming there may not always be adequate resource to evaluate large numbers of generated SDRs within a short timeframe, such criteria could be based on seriousness, listedness, DME status, level of disproportionality, etc.	
Section 7.1, Diagram 1	Please add an additional step (and additional box) on the flow diagram stating the following: *Medical/scientific assessment of the SDR to determine whether SDR represents a potential signal that requires further detailed evaluation.* This box could appear either before or after the final box currently shown on the flow diagram, depending on whether this step is expected	The purpose of the diagram is to provide a framework to perform a thorough (detailed) scientific evaluation of all available evidence which may substantiate or refute the existence of a true signal. The procedures and actions used in the presence of a signal (signal evaluation, signal confirmation, signal communication and appropriate regulatory actions) are detailed in the Community legislation and guidelines. The table was renamed; the word "review" was replaced by the word "evaluation".

	of the organisation generating the SDR (e.g. the EMEA), or whether it is routinely expected of the stakeholder who has the information about the identified SDR communicated to them. If this step is to be conducted by the stakeholder following communication of the SDR, it would be useful if the text could include guidance on how topics may be prioritised for review in situations where multiple SDRs are highlighted for a given medicinal product.	
GUIDELINE SECTION TITLE: 7.2. Specific aspects of the use of quantitative methods for vaccines and medicines used in children		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 7.2	Other areas of special interest should be considered. Each area with special populations, e.g. cancer therapy, drugs used only by male or female, etc. needs specific requirements.	The comment is acknowledged. Some areas have already been identified (clinical trials, vaccines, medicines used in children, etc …). The list is not limitative.
GUIDELINE SECTION TITLE: 8. TARGETED MONITORING AND RISK MANAGEMENT PLANS		
Line no. + para no.	**Comment and Rationale**	**Outcome**
section 8, para 2, para 1	'monitoring of these risks' (identified and potential risks from RMP): note that if risks are labelled, PRR is likely to be raised ('me-too reporting' effect).	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment). The opposite effect on the PRR may also happen.
Section 8	It is implied in Section 8 that the EMEA would monitor the events that the MAH specifies in the Risk Management Plan (RMP) as targeted medical events when they apply the statistical methods described in this guideline. It is not clear if the guideline would advise the EMEA and member states to get pre-specified agreements with the MAH regarding the targeted medical events and parameter specifications (e.g., search criteria based on MedDRA terms as implied in Table 8; thresholds for SDRs). In a typical application of data mining techniques, any SDRs that meet	Such agreement should be reach with the adoption of the risk management plan; some additional guidance is given in the EMEA template for EU risk-management plans including the Annex I: Interface between EU-RMP and EudraVigilance published on the EudraVigilance website. http://eudravigilance.emea.europa.eu/human/EURiskManagementPlans.asp

	pre-specified criteria could be considered the basis for further investigations (e.g., additional pharmacovigilance activities or requirements by the EMEA or member states to the MAH). Would the monitoring for pre-specified events be handled differently from the monitoring of other events?	
Section 8, para 2	It is not clear how will the screening for additional Targeted Medical Events (such as drug misuse, off-label use and drug interactions) be performed. These types of events are difficult to detect consistently in automated fashion. Moreover, the TME search strategy will need to be tailored to individual products/groups of products, otherwise low specificity searches will produce large number of reports requiring intensive manual review.	Reference is made to the risk management activities and EU risk management guidelines. EudraVigilance is at the core of the technical implementation of the EU risk management activities (see also documents published by the European Risk Management Strategy).
Section 8, para 2, line 4	With regards to "*monitoring of these risks*" (identified and potential risks from RMP), please note that if the risks are already labeled, the PRR is likely to be raised (i.e. a 'me-too reporting' effect).	The comment is acknowledged (the draft guideline was not modified, it is important to highlight that the EV-EWG does not necessarily agree [or disagree] with the content of the comment). The opposite may also happen.
Section 8, Table 8	The table structure seems more complex than the information presented would require (the same thing is repeated for the two categories of risks; why need four rows in the table?).	The comment was implemented and the paragraph was updated to make reference to the following documents: Committee for Medicinal Products for Human use (CHMP). Guideline on risk management systems for medicinal products for human use. Annex C: Template for EU Risk Management Plan (EU – RMP). Adopted on 27 September 2006. Doc.Ref. EMEA/192632/2006. Interface between European Risk Management Plan (EU-RMP) and EudraVigilance as referred to in Annex 1 of "Annex C: Template for EU Risk Management Plan (EU-RMP)" ver. 1.0.0.
Section 8, Table 8	"*List of terms*" only refers to at least one of the considered terms per report. It does not consider the concurrent reporting of more than one of those terms as e.g. to describe hypersensitivity (e.g. asthma and edema)	The comment was implemented and the paragraph was updated to make reference to the following documents: Committee for Medicinal Products for Human use (CHMP). Guideline on risk management systems for medicinal products for human use. Annex C: Template for EU Risk Management Plan (EU – RMP). Adopted on 27 September 2006. Doc.Ref. EMEA/192632/2006. Interface between European Risk Management Plan (EU-RMP) and EudraVigilance as referred to in Annex 1 of "Annex C: Template for EU Risk Management Plan (EU-RMP)" ver. 1.0.0.

		It is also possible to use SMQs when appropriate.
Section 8, Table 8	Does the TME concept correspond to the SMQ concept as described by CIOMS VI? The terminology should be clear.	The Standard MedDRA Queries and the concept of Targeted Medical Events are different. TMEs can be HLGT, HLT, PT, LLT or SMQs.
Section 8, Table 8	It should also be possible to run the signal detection standard reports for TMEs. Which of the different reports or do all standard reports offer the option to show the signal measures on TME level?	It is of course possible to use the quantitative methods on reports which include TMEs.
Section 8, Table 8	"*Additional TMEs may be set up …*". Isn't this section out of scope of the guideline? There is no guidance at all how to run signal detection for those TMEs.	The detection of new signals is entirely part of the risk management activities. Additional guidance on the TMEs / important identified and potential risks that require further characterisation or evaluation is given in the Committee for medicinal products for human use (CHMP) guideline on risk management systems for medicinal products for human use. Adopted on 14 November 2005. EMEA/CHMP/96268/2005.
Section 8, para 2	It is not clear how TME's identified in the RMP will filter through to EudraVigilance. Who will notify EudraVigilance of any TME's if the product is not licensed centrally? Presumably this should be the company.	See Interface between European Risk Management Plan (EU-RMP) and EudraVigilance as referred to in Annex 1 of "Annex C: Template for EU Risk Management Plan (EU-RMP)" ver. 1.0.0. The guideline is available at the following URL: http://eudravigilance.emea.europa.eu/human/EURiskManagementPlans.asp
Section 8	The bullet point list of TME's should also include potential risks identified via PSUR's or other safety reviews. Again, this raises the question as to how this filters through to EudraVigilance.	In such case, the risk management plan should be updated accordingly (see Committee for medicinal products for human use (CHMP) guideline on risk management systems for medicinal products for human use. Adopted on 14 November 2005. EMEA/CHMP/96268/2005). The comment was implemented.
Section 8, Table 8	The use of LLT is not allowed in the EU-RMP template (Doc. Ref. EMEA/192632/2006). Specifically the Annex 1 template for Interface between EU-RMP and EudraVigilance dose not accept LLT. To remove LLT in the Table.	The comment was implemented.
Section 8	Set the document in context It is proposed that the first paragraph of section 8 is moved to section 1 to place the use of the described methods in context	The comment was implemented (an additional but different paragraph was included in section 1).

GUIDELINE SECTION TITLE: 9. IMPORTANCE OF DATA QUALITY IN SIGNAL DETECTION		
Line no. + para no.	**Comment and Rationale**	**Outcome**
Section 9	This section seems to appear as a 'tagged on afterthought' but should be given more prominence. Suggest move this to be between section 4.7 and 5 (i.e. make it section 5), or possibly move it to after the Introduction (i.e. make it section 2).	The comment is acknowledged but the original order seems more logical.

GUIDELINE SECTION TITLE: 10. REFERENCES		
Line no. + para no.	**Comment and Rationale**	**Outcome**
References	The references do not appear to be in any order.	Te references are presented in order of appearance.
References	Discussion and references in this area are incomplete and do not present a balanced picture of work currently being done. There is also insufficient discussion around the large number of false positives 'signals' that are generated using PRRs. A Discussion Section to include acknowledgement of the work being done to improve on the PRR such as the EBGM and BCPNN would be useful. This would also aid discussions around false positives as EBGM is designed to reduce noise.	The list of references has been extended and should be considered in light of the scope of the guideline.

SPECIFIC COMMENTS ON TEXT		
GUIDELINE SECTION TITLE: other general comments		
Line no.[1] + paragraph no.	**Comment and Rationale**	**Outcome**
Paragraph 1	To clarify some events with few number of reports may be chosen to follow up simply because of their medical importance. To add a sentence on the clarification.	This is the underlying idea of DMEs/TMEs.
	Eudravigilance will not be useful to people in industry until they can see all AEs on all drugs; just being able to view one's own company's drugs and AE reports is not useful; should follow the AERS model.	The comment is acknowledged.

[1] Where applicable