



Validation of a Novel Method to Assess the Clinical Quality of Information in Pregnancy-Related Pharmacovigilance Case Reports: A ConcePTION Project

Yrea R. J. van Rijt-Weetink^{1,2} · Toine C. G. Egberts^{3,4} · Florence P. A. M. van Hunsel^{1,2} · David J. Lewis^{5,6} · Laura M. Yates^{7,8} · Ursula Winterfeld⁹ · Eugène P. van Puijenbroek^{1,2}

Accepted: 3 December 2023 / Published online: 6 January 2024
© The Author(s) 2024

Abstract

Background To assess the causal relationship between a medicinal product and a reported event, relevant information needs to be present. Information elements for assessing cases of exposure to medicinal products during pregnancy were predefined and used in a new tool to assess the quality of information. However, the extent in which the presence or absence of these predefined information elements is associated with the overall clinical quality of these cases, as evaluated by pharmacovigilance experts, remains uncertain.

Objective We aimed to validate a novel method to assess the clinical quality of information in real-world pregnancy pharmacovigilance case reports.

Methods The clinical quality of case reports regarding medicinal product exposure and pregnancy-related outcomes was appraised from spontaneous reports, literature, Teratology Information Services (UK and Switzerland), The Dutch Pregnancy Drug Register, the Gilenya pregnancy registry and the Enhanced PV programme of Novartis. Assessment was done by means of the novel standardised tool based on the presence and relevance of information, and by expert judgement. The novel tool was validated compared to the expert assessment as the gold standard expressed as the area under the receiver operating characteristic curves, after which the sensitivity and specificity were calculated using cross-tabulations. Inter-rater variability was determined by means of weighted Cohen's kappa.

Results One hundred and eighty-six case reports were included. The clinical quality score as assessed by the novel method was divided into three categories with cut-off values of 45% (poor to intermediate) and 65% (intermediate to excellent). Sensitivity was 0.93 and 0.96 for poor to intermediate and intermediate to excellent, respectively. Specificity was respectively 0.52 and 0.73. Inter-rater variability was 0.65 (95% confidence interval 0.53–0.78) for the newly developed approach, and 0.40 (95% confidence interval 0.28–0.52) for the gold standard assessment.

Conclusions The tool described in this study using the presence and relevance of elements of information is the first designed, validated and standardised method for the assessment of the quality of information of case reports in pregnancy pharmacovigilance data. This method confers less inter-rater variability compared with a quality assessment by experts of pregnancy-related pharmacovigilance data.

Key Points

We validated a novel method for the assessment of the quality of pregnancy pharmacovigilance data.

This approach shows less inter-rater variability compared with a quality assessment by experts.

1 Introduction

As pregnant women are almost never included in clinical studies, the information regarding the safety of medicinal product exposure during pregnancy is mainly dependent on the reporting of experiences with exposure to these products in the real world. Several different data collection methods specifically aimed at the collection of safety information are currently in use to capture and analyse reports on risks associated with medicinal products used during pregnancy [1]. For example, individual case safety reports (ICSRs) are collected by national pharmacovigilance centres and marketing authorisation holders (MAHs), and then forwarded to the European Medicines Agency and the World Health Organization collaborating centre, the Uppsala Monitoring Centre [2, 3]. In the post-marketing phase, ICSRs can be reported spontaneously, solicited or based on cases published in the literature. Furthermore, Teratology Information Services (TIS), pregnancy registries or MAH-initiated enhanced pharmacovigilance programmes (in which spontaneously reported pregnancies are followed, using amongst others a structured follow-up data collection at set intervals) gather and analyse pregnancy data with the purpose of increasing the knowledge about the safety of medicinal product exposure during pregnancy [4–6].

A significant challenge in pharmacovigilance is to assess the likelihood that the medicinal product caused or contributed to the occurrence of the event experienced by the patient, i.e. causality assessment. In order to assess the possibility of a causal relationship between exposure to a medicinal product during pregnancy and an adverse pregnancy outcome, it is important that the underlying data (raw variables) and information (processing and interpretation of raw data), in individual case reports and case series, are both present and of high quality [7]. Therefore, in order to establish reliably whether there is a causal relationship between a medicinal product and a reported event, *relevant* information for the causality assessment needs to be *present*. A case report with many relevant elements of information containing the appropriate meaningful information is considered to have a high clinical quality of information. Whereas a report with only little information present that is relevant to assessing the causal relationship is considered to have a low clinical quality of information, even if other details *not* relevant to the particular event under assessment are well reported. Which information is relevant is dependent on the reported association of a medicinal product and a pregnancy-related outcome.

In the past, various approaches were developed to assess the completeness of ICSRs, including Vigigrade, developed by the Uppsala Monitoring Centre [2]. Unfortunately,

these methods do not take the relevance of information into account and base their assessment only on completeness of certain variables. An exception is the instrument designed for the assessment of the clinical quality of the information, developed by Pharmacovigilance Centre Lareb [8]. All aforementioned approaches were however not specifically designed for pregnancy data, which requires information that may differ substantially from the information needed for a non-pregnancy-related causality assessment. In the context of general pharmacovigilance, information concerning the suspected medicinal product, the event, the timing and alternative explanations is generally important. Additionally, in the context of pregnancy-related pharmacovigilance, information concerning both the mother and child needs to be taken into account, and the nature of the medicinal product–event association may differ substantially for various clinical scenarios.

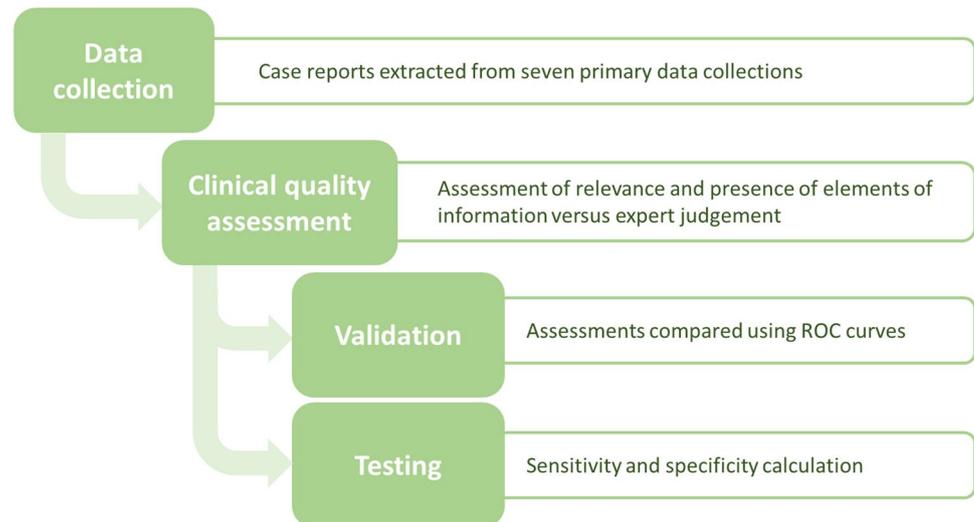
In a previous study, we identified important elements needed to assess the clinical quality of information for a risk assessment of medicinal products used during pregnancy, by means of two surveys and two focus group discussions amongst pregnancy pharmacovigilance experts [9]. However, it is still unknown how the presence or absence of these selected information elements in case reports relate to the overall actual clinical quality of the reports for assessment of the causal relationship of a specific reported association as assessed by pregnancy pharmacovigilance experts. Therefore, the aim of the current study was to validate the previously described method to assess the clinical quality of information in real-life pregnancy pharmacovigilance case reports, based on the presence of elements of information relevant to the particular outcome reported.

2 Methods

2.1 Setting and Design

This study is part of Work Package 2 of the IMI-funded ConcePTION project, in which national pharmacovigilance centres, MAHs and TIS centres collaborated in optimising the collection, analysis and interpretation of reported pregnancy pharmacovigilance data [10]. The clinical quality of case reports regarding medicinal product exposure during pregnancy was assessed by means of the newly developed tool and by expert judgement [9]. In the current study, this tool was trained and subsequently validated in comparison to expert assessment (Fig. 1). All steps are discussed in more detail below.

Fig. 1 Flow chart of the study design in which the novel method was trained and subsequently validated for the quality assessment of information of case reports in pregnancy pharmacovigilance data. *ROC* receiver operating characteristic



2.2 Data Collection

The following data sources each provided 30 randomly selected, full case reports: spontaneous reports and case reports from the literature from the Dutch national spontaneous reporting database (the Netherlands Pharmacovigilance Centre Lareb) [11]; TIS reports from both the Swiss TIS centre and the UK TIS centre [4]; pregnancy registry reports from The Dutch Pregnancy Drug Register (Lareb) [6] and the Gilenya (fingolimod) Pregnancy Registry (Novartis) [12]; and enhanced pharmacovigilance programme reports from the global Argus Safety database (Novartis) [5]. All information available in each report was extracted, except personal details or other details that could not be shared because of privacy reasons. These details were blacked out (redacted) in order to distinguish between missing (null) and hidden (masked) information.

Thirty case reports were randomly selected from each data source, with a minimum of four case reports per clinical scenario. The clinical scenarios were:

1. Pregnancy loss;
2. Congenital anomalies or chromosomal defects;
3. Foetal or neo-natal complications;
4. Infant or child complications;
5. Maternal pregnancy-related complications; and
6. Absence of adverse outcomes or complications.

The same selection of clinical scenarios was used for the development of the elements of information in a previous study [9]. If a data source did not contain at least four case reports from a clinical scenario, substitution with an additional case report from another randomly selected category was possible.

Case reports had to be completed or closed within the timeframe of 1 December, 2019 to 31 December, 2020, and had to be associated with medicinal product exposure during pregnancy. Both prospective (reported prior to the outcome being known) and retrospective case reports (reported after the pregnancy outcome became available) were included, although it was required that a possible pregnancy complication or outcome with a temporal association to a medicinal product was present. Adverse and healthy pregnancy and/or child outcomes were eligible for inclusion. Exclusion criteria included reports related to drug exposure during breast feeding only, paternal drug exposure, non-pregnancy related complications of the mother, and delivery or post-partum complications of the mother only.

All case reports were blinded for the data source, by reformatting the data to a standardised layout, and by replacing any details that may have revealed the origin of the case reports with equivalent anonymised information (e.g. geographical references, non-English terminology). Finally, the order of all case reports was randomised.

2.3 Clinical Quality Assessment

The clinical quality of the case reports was assessed in two ways: by means of the newly developed method based on the presence and relevance of information, and by expert judgement [9].

2.3.1 Novel Method Based on the Presence and Relevance of Information

In a previous study, elements deemed necessary for assessing the quality of information in pregnancy-related

pharmacovigilance case reports were determined [9]. A total of 21 elements of information were selected:

1. Association (one element): information on the combination of the medicinal product–event combination. This information is always required to assess the causal relationship.
2. Event details (three elements): information regarding the event under investigation is important to confirm the diagnosis and to provide insight into the development of the event over time.
3. Medicinal product exposure details [4]: variables that provide more details about the timing, route of administration and indication of the suspect product. Additionally, co-medication and supplements are described in this category.
4. Maternal factors (three elements): information on the characteristics of the mother that could affect the event is described in this category.
5. Pregnancy (three elements): information on previous pregnancies and the current pregnancy that could provide more detail or affect the event are included in this category.
6. Labour (three elements): information related to labour is included in this category.
7. Child (four elements): information about the child and characteristics that could affect or (partially) explain the event is described.

The previously selected elements of information [9] were assessed for both relevance and the presence of meaningful information. The clinical quality was expressed as the number of present and relevant elements divided by the number of relevant elements, $\times 100$ (Table 1). For example, for a specific association, only 16 out of 21 information elements might be relevant, whereafter the presence of the information in those elements is assessed. If 12 out of those 16 relevant information elements is present, the clinical quality is 75%.

For the clinical quality assessment of the case reports, six pregnancy pharmacovigilance experts were recruited within Pharmacovigilance Centre Lareb, over whom the reports were divided. For each case report, the assessors were asked to mark which information elements they considered to be relevant if they were to perform a causality assessment of that specific combination of exposure and outcome, and for which of those elements the required information was available. The assessors did not perform a causality assessment, they marked the elements of information that they considered relevant if they were to perform a causality assessment. The clinical quality was calculated as a percentage of the number of available information elements in relation to the number of considered relevant information elements. Reports were assessed by two different assessors. A final assessment was agreed upon via consultation of the two assessors.

Table 1 Standardised method for the assessment of the clinical quality of case reports in pregnancy-related pharmacovigilance data. [9] Presence of the association is a requirement for the method to be used

Category	Elements of information	Relevance	Presence
Association	Medicinal product–event combination		
Event details	Information to validate the diagnosis of the event (e.g. test results) Timing of occurrence or detection of the event Chronologic evolution of the event, possibly in relation to the exposure		
Medicinal product exposure details	Administration information of medicinal product (e.g. dose, route) Timing of exposure of medicinal product in relation to timing of gestation Indication for use of medicinal product Other exposures (including additional information on e.g. dose, route, timing and indication)		
Maternal factors	Medical history and concurrent disorders of the mother Maternal demographics (e.g. age, weight) Lifestyle and risk factors (e.g. smoking, alcohol)		
Pregnancy	Previous pregnancies Pregnancy-related complications of current pregnancy Prenatal testing		
Labour	Labour onset Mode of delivery Delivery complications		
Child	Gestational age at birth Apgar score (1, 5, 10 min) Breastfeeding Medical information of neonate (e.g. weight, diagnoses that are not the reported event)		

Clinical quality = number of present and relevant elements of information/number of relevant elements of information $\times 100$

2.3.2 Expert Assessment (Gold Standard)

Eighty-one pregnancy pharmacovigilance experts of Work Package 2 of the IMI-funded ConcePTION project [10] and the ISoP special interest group of women's medicines [13] were asked to participate in the comparative expert assessment of the clinical quality of a set of cases, in a questionnaire performed in the preceding study [9]. Twenty assessors were recruited. Assessors were asked to grade the quality of the case reports for the purpose of assessing the causal relationship between the suspected medicinal product and outcome (clinical quality) on a four-point scale: poor, moderate, good, excellent. They were asked to grade the quality on personal judgement only and thus did not have the quality criteria of the novel method available, but were guided on the definitions in four reference categories of poor, moderate, good, and excellent clinical quality (< 50%, < 75%, < 90%, \geq 90% of necessary information present, respectively). All reports were assessed by two different assessors from different organisations. In case of divergence, the two assessments were sent to a third assessor who decided on the final assessment.

2.4 Data Analysis

2.4.1 Training Phase

A random subset of half of the included case reports was selected to use for the training of the method. In the training phase, the parameters of a method are tuned, meaning that the cut-off percentages of the categorisation of the clinical quality were determined. For this, receiver operating characteristic (ROC) curves [14] were constructed of our novel standardised tool versus an expert assessment, for the transitions of poor to moderate, moderate to good and good to excellent. The areas under the curve (AUCs) were calculated. Cut-off percentages of the categories were determined subjectively based on the optimal balance between sensitivity and specificity determined by the ROC curves.

2.4.2 Validation Phase

In the validation phase, the performance of the fully tuned method was determined using the other half of the included case reports. Therefore, the clinical quality was categorised by the cut-off values as determined. Definitive sensitivity and specificity were calculated using cross-tabulations of the novel method versus the expert assessment. Inter-rater variability was calculated for both methods by Cohen's kappa (weighted) on the total dataset [15]. An overview of cases

used in both the training and validation phase can be found in the Electronic Supplementary Material (ESM).

3 Results

In total, 186 case reports were included, of which 93 were included in the training phase and 93 in the validation phase (Table 2).

3.1 Training

Figure 2 shows the ROC curves of the assessment of clinical quality using the novel method versus the expert assessment, in the four reference categories that were provided to the assessors. The blue line represents the transition from poor to moderate, with an AUC of 0.95 (95% CI 0.89–1.00). Red represents the transition from moderate to good, with an AUC of 0.92 (95% CI 0.86–0.95) and black represents the transition from good to excellent, with an AUC of 0.88 (95% CI 0.80–0.95). Based on the sensitivity and specificity as reflected by the curves and the size of the categories, it was determined that three, rather than four categories would suffice. Moderate and good were therefore combined into one category 'intermediate' with cut-off values of 45% (poor to intermediate) and 65% (intermediate to excellent). The tables of the coordinates of the ROC curves can be found in the ESM.

3.2 Validation

During the validation phase, the three established categories of poor, intermediate and excellent were used. Sensitivity and specificity were calculated on the second half of the assessed case reports (Table 3). Sensitivity (true positive rate) was 0.93 and 0.96 for poor to intermediate and intermediate to excellent, respectively. Specificity (true negative rate) was respectively 0.52 and 0.73. Inter-rater variability was calculated by means of the weighted Cohen's kappa, with 0 indicating no agreement and 1 indicating perfect agreement, and was 0.65 (95% CI 0.53–0.78) for the novel method (after division into final categories), and 0.40 (95% CI 0.28–0.52) for the expert assessment (after combining 'moderate' and 'good' into one category 'intermediate').

4 Discussion

4.1 Performance of the Novel Method

This study provides a novel validated and standardised method for the assessment of information in case reports in pregnancy-related pharmacovigilance data, with a sensitivity

Table 2 Description of the cohort of case reports

	Total 186 (100%)	Training 93 (100%)	Validation 93 (100%)
<i>Expert assessment</i>			
Poor	30 (16.1%)	9 (9.7%)	21 (22.6%)
Moderate	35 (18.8%)	14 (15.1%)	21 (22.6%)
Good	55 (29.6%)	30 (32.3%)	25 (26.9%)
Excellent	66 (35.5%)	40 (43.0%)	26 (28.0%)
<i>Source</i>			
Spontaneous reports	25 (13.4%)	17 (18.3%)	8 (8.6%)
Literature reports	23 (12.4%)	9 (9.7%)	14 (15.1%)
TIS (UK)	28 (15.1%)	10 (10.8%)	18 (19.4%)
TIS (Switzerland)	27 (14.5%)	17 (18.3%)	10 (10.8%)
The Dutch Pregnancy Drug Register	30 (16.1%)	12 (12.9%)	18 (19.4%)
Gilena pregnancy registry	28 (15.1%)	15 (16.1%)	13 (14.0%)
Enhanced PV programme	25 (13.4%)	13 (14.0%)	12 (12.9%)
<i>Clinical scenario</i>			
Pregnancy loss	33 (17.7%)	19 (20.4%)	14 (15.1%)
Congenital anomalies or chromosomal defects	37 (19.9%)	18 (19.4%)	19 (20.4%)
Foetal or neo-natal complications	35 (18.8%)	18 (19.4%)	17 (18.3%)
Infant or child complications	3 (1.6%)	0	3 (3.2%)
Maternal pregnancy-related complications	33 (17.7%)	18 (19.4%)	15 (16.1%)
No adverse outcomes or complications	43 (23.1%)	19 (20.4%)	24 (25.8%)
Other	2 (1.1%)	1 (1.1%)	1 (1.1%)

PV pharmacovigilance, TIS Teratology Information Services

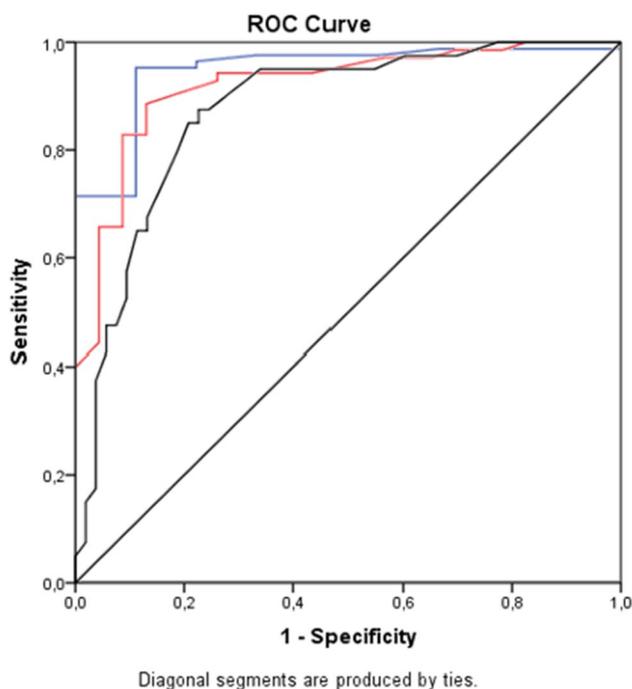


Fig. 2 Receiver operating characteristic (ROC) curves of clinical quality assessments versus expert assessment, in four categories. *Blue* = poor to moderate, *red* = moderate to good, *black* = good to excellent.

of 0.93 (poor to intermediate) and 0.96 (intermediate to excellent) and a specificity of 0.52 (poor to intermediate) and 0.73 (intermediate to excellent). A requirement for the use of this method is that the medicinal product–event association is present in the case report. In the case of absence of the association, for example, in the case where no medicinal product or pregnancy outcome is reported, the case report is unusable for a causality assessment, independent of the presence and relevance of the other elements of information. Individual case safety reports should not be submitted to the competent authorities if the association is missing [16], but TIS centres, MAHs and registries collect reports without a known adverse pregnancy outcome or complication. In the training and validation phase of this study, the medicinal product name and the outcome or complication of interest were present in all included reports. In reports used in practice, the information described may be unclear or ambiguous.

Of all items mentioned in Table 1, 73.7% were considered relevant and included in the calculation of the scores. In the training set, 23.9 % of the items were considered relevant but were not present; 46.1% of the items were considered relevant and were also present, and 30.0 % of the items were considered irrelevant. This is comparable to the validation set, in which 29.5 % of the items were considered relevant but were not present; 44.1% of the items were considered

Table 3 Cross-tabulations of the novel method versus the expert assessment for calculation of sensitivity and specificity. A: Cross-tabulation for the transition from poor to intermediate. B: Cross-tabulation for the transition from intermediate to excellent

A		Expert assessment	
		Poor	Moderate/ good/excel- lent
Novel method	< 45% (poor)	11 (52.4%)	5 (6.9%)
	≥ 45% (intermediate/excellent)	10 (47.6%)	67 (93.1%)
B		Expert assessment	
		Poor/moderate/good	Excellent
Novel method	< 65% (poor/intermediate)	49 (73.1%)	1 (3.8%)
	≥ 65% (excellent)	18 (26.9%)	25 (96.2%)

relevant and were also present, and 26.3 % of the items were considered irrelevant.

Compared with a pregnancy pharmacovigilance expert assessment, the inter-rater variability is increased from 0.40 (fair) to 0.65 (substantial) by using the novel method based on the presence and relevance of information [15]. In general, at least substantial agreement ($\kappa > 0.6$) is considered to be necessary for acceptable reliability [17]. This desired level of agreement is achieved with the novel method, wherein the two assessors scored 66.7% of case reports in the same category, with only 2.0% of case reports rated as poor by one assessor and excellent by the other. For the expert assessment, these numbers were 50.5% for the same category, and 2.7% for poor by one, and excellent by the other.

4.2 Practical Use

High-quality data collection leads to a more reliable risk assessment, and thereby to improved signal detection and characterisation [8, 18]. A standardised method of assessing the quality contributes to the transparency and reproducibility and ensures a uniform method of assessing a problem, considering all relevant aspects.

In order to improve the clinical quality of information that is collected, it is important to understand the characteristics of the current data, and the most important data elements required for a clinical assessment (i.e. the clinical context). Knowledge of the clinical quality of reports included in any signal detected would provide insight into the reliability of the assessment of the causal relationship of that signal. A signal consisting of fewer case reports with excellent clinical quality could be of more value and weight when compared with a signal consisting of more case reports with poor clinical quality. Furthermore, this method could be implemented to identify points of improvement for clinical quality, both in research settings and as good pharmacovigilance practice.

When reports from a specific data source consistently lack a certain relevant information element, while reports from another source consistently collect that element, knowledge gained from the one source can be used to improve the collection of the other source. In other words, if one source collects an information element very well, their method of collection might be implemented in a source that has more difficulty collecting that information element. This of course also applies to comparisons for reporters and reports regarding different clinical scenarios. In the process of general pharmacovigilance, the clinical quality of reports could be optimised by identifying which additional information should be collected in the process of collecting follow-up data from the reporters.

The number of ICSRs analysed in this study is too small to draw firm conclusion on differences between the various sources in the score for clinical scenarios, although it is likely that some differences might exist. In the next study, we aim to study differences among various sources and will highlight these differences in more detail.

In the majority of cases, the goal of reporting ICSRs is to assess the causal relationship between an adverse drug reaction and the exposed drug. In pregnancy pharmacovigilance, however, information on cases without an adverse outcome also contributes to the knowledge of a drug's safety. A sufficient number of cases without an adverse outcome will make the absence of a detrimental effect more likely. This category substantially differs from the other categories of ICSRs related to exposure during pregnancy. It should be noted that, similar to cases with adverse outcomes, factors such as (occupational) exposure to other potential teratogens, pregnancy, labour and development of the child contribute to a better understanding of the circumstance under which the exposure took place. The advantage of this novel method is its versatile character, which allows a tailor-made assessment for various clinical scenarios, including cases without adverse outcomes.

4.3 Setting

Previous methods designed to assess the completeness or quality of case reports did not fulfil the requirements to quantify the clinical quality of pregnancy data specifically [2, 8]. First, because these methods were only validated on ICSRs, while ICSRs are only one of the many pregnancy pharmacovigilance data collection sources that are used in daily practice [1]. Second, and more importantly, because these methods were designed for adverse drug reactions in general, while reports on pregnancy data differ greatly from general reports in the type of information needed for a risk assessment.

This novel method based on the presence and relevance of information was designed to account for important information for a causality assessment, namely relevance, completeness and precision [8, 19]. The first step is to assess which information is relevant for the evaluation of a possible causal relationship between the medicinal product and the reported event. In this way, redundant information in the assessment will not result in an overestimation of the clinical quality. In the second step, all relevant information is assessed on its presence, consisting of both completeness and precision. This prevents the overestimation of the clinical quality due to irrelevant, ambiguous or unclear information. A characteristic of this tool is that the assessment of the clinical quality using this method requires expertise in the field of pregnancy pharmacovigilance. For example, an assessor should have the knowledge to determine relevant risk factors for a specific association, in order to assess whether all necessary information is present.

4.4 Strengths and Limitations

The designed method was validated in a multitude of relevant data sources and clinical scenarios. Seven pregnancy pharmacovigilance data collection sources were selected, which include most of the available primary data sources for this purpose. For data sources that were believed to possibly differ between centres or study designs, such as TIS centres and pregnancy or drug registries, multiple centres or studies were included. The included clinical scenarios (pregnancy loss, congenital anomalies or chromosomal defects, foetal or neo-natal complications, infant or child complications, maternal pregnancy-related complications, and no adverse outcomes or complications) were based on the similarity between the profiles of information elements that are necessary for a risk assessment. This method was not validated for the assessment of the clinical quality of reports regarding, for example delivery complications, clinical scenarios related to paternal medicinal product exposure during the periconceptional period, clinical scenarios related to the use of medicinal products during breastfeeding, and maternal

non-pregnancy-related complications. The latter scenario could be assessed using the instrument for the assessment of the clinical quality of information designed by Pharmacovigilance Centre Lareb, which was not designed for pregnancy data specifically, while for the other three scenarios an adapted version of the method described in this study would have to be designed [8]. For the clinical scenarios related to paternal medicinal product exposure during the periconceptional period, the proportion of cases related to paternal exposure is rather low. Currently, the selection and assessment of cases of paternal exposure are complicated, as both maternal and paternal factors play a role. Developing an extensive list of additional elements related to paternal exposure would be necessary. For this reason, this category was not taken into account. Additionally, the scenario 'infant or child complications' was vastly under-represented in the current dataset (Table 2). This was owing to the fact that these events are often more difficult to capture because of a limited follow-up period.

All cases were randomised and blinded for assessment, so that it was as difficult as possible to determine from which data source the data were extracted. However, the medicinal product could (non-definitively) be associated with specific data sources (e.g. all cases from the Gilenya Pregnancy Registry [Novartis] and the enhanced pharmacovigilance programme [Novartis] were associated with fingolimod). Even though it is possible that other sources provided fingolimod cases, assessors most likely associated these cases with the Gilenya Pregnancy Registry and the enhanced pharmacovigilance programme.

The gold standard in this study was the expert assessment. Although this method is subjective and lacks standardisation, it is the most commonly used method for a causality assessment [20]. The limitations of an expert assessment, as demonstrated by the high inter-rater variability, can be minimised by using a Delphi method [21]. In this study, a modified Delphi method was used, in which multiple experts assessed each case report, and a third expert decided on the final assessment in the case of disagreement.

5 Conclusions

The method described in this study using the presence and relevance of elements of information is the first designed, validated and standardised method for the assessment of the quality of information of case reports in pregnancy pharmacovigilance data. It provides a method with less inter-rater variability compared with a quality assessment by experts of pregnancy-related pharmacovigilance data. In future studies, this method can be used for the comparison of the clinical quality of different pregnancy data collection sources, which

may facilitate the improvement of pregnancy pharmacovigilance data collection in general.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40264-023-01389-y>.

Acknowledgements The authors thank Khoezik Chamani for her help in the data collection and analysis as well as all assessors of the case reports providing their expert opinion. The research leading to these results was conducted as part of the ConcePTION consortium. This paper only reflects the personal views of the stated authors.

Declarations

Funding The ConcePTION project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 821520. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Conflicts of interest/competing interests David J. Lewis is a full-time employee of Novartis Pharma AG and holds shares in GlaxoSmith-Kline and Novartis. Laura M. Yates provided consultancy for Sanofi Genzyme South Africa on two occasions in 2021 relating to genetic testing in Gaucher's disease (April 2021, advisory board presentation) and a conference lecture on genetic testing in cardiac clinics (September 2021). Yrea R.J. van Rijt-Weetink, Toine C.G. Egberts, Florence P.A.M. van Hunsel, Ursula Winterfeld and Eugène P. van Puijenbroek have no conflicts of interest that are directly relevant to the content of this article.

Ethics approval Ethics approval was not needed for this study.

Consent to participate No approval or consent was needed for this study.

Consent for publication No approval or consent was needed for this study.

Availability of data and material The datasets for this article are not publicly available because of the data protection policy of Pharmacovigilance Centre Lareb. Requests to access the datasets should be directed to the corresponding author and will be granted on reasonable request.

Code availability The codes for this article are not publicly available because of the data protection policy of Pharmacovigilance Centre Lareb. Requests to access the codes should be directed to the corresponding author and will be granted on reasonable request.

Authors' contributions The original study protocol was designed by YvRW, DL, LY, UW and EvP. Data analysis was performed by YvRW, KC and EvP. The design of the manuscript was determined by YvRW, TE, FvH and EvP. All authors contributed to manuscript drafting and revision. All authors approved the final version to be published and agree to be accountable for all aspects of the work.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative

Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Benevent J, Montastruc F, Damase-Michel C. The importance of pharmacoepidemiology in pregnancy-implications for safety. *Expert Opin Drug Saf.* 2017;16(10):1181–90.
2. Bergvall T, Norén GN, Lindquist M. VigiGrade: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug Saf.* 2014;37(1):65–77.
3. European Medicines Agency. EudraVigilance. <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance>. Accessed 15 Nov 2023.
4. European Network of Teratology Information Services. ENTIS. <https://www.entis-org.eu/>. Accessed 15 Nov 2023.
5. Geissbühler Y, Rezaallah B, Moore A. An alternative to product-specific pregnancy registries? PRIM PREGnancy outcomes Intensive Monitoring. *Reprod Toxicol.* 2020;94:13–21.
6. Vorstenbosch S, Te Winkel B, van Gelder M, Kant A, Roelvelde N, van Puijenbroek E. Aim and design of pREGnant, the Dutch Pregnancy Drug Register. *Drug Saf.* 2019;42(1):1–12.
7. Alla F, Rosilio M, Funck-Brentano C, Barthélémy P, Brisset S, Cellier D, et al. How can the quality of medical data in pharmacovigilance, pharmacoepidemiology and clinical studies be guaranteed? *Thérapie.* 2013;68(4):209–23.
8. Oosterhuis I, Rolfes L, Ekhart C, Muller-Hansma A, Härmark L. First experiences with a tool to measure the level of clinical information present in adverse drug reaction reports. *Expert Opin Drug Saf.* 2018;17(2):111–5.
9. van Rijt-Weetink YRJ, Chamani K, Egberts ACG, van Hunsel FPAM, Lewis DJ, Yates LM, et al. Elements to assess the quality of information of case reports in pregnancy pharmacovigilance data: a ConcePTION project. *Front Drug Saf Regul.* 2023;3:1187888.
10. ConcePTION. <https://www.imi-conception.eu/>. Accessed 15 Nov 2023.
11. Pharmacovigilance centre Lareb. <https://lareb.nl/en/pages/about-lareb/>. Accessed 2 Jan 2024.
12. Geissbühler Y, Vile J, Koren G, Guennec M, Butzkueven H, Tilson H, et al. Evaluation of pregnancy outcomes in patients with multiple sclerosis after fingolimod exposure. *Ther Adv Neurol Disord.* 2018;11:1756286418804760.
13. The ISoP Special Interest Group on Women's Medicines. <https://isoponline.org/special-interest-groups/womens-medicines-group/>. Accessed 15 Nov 2023.
14. Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J.* 2017;34(6):357–9.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.
16. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VI: collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). 2017. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf. Accessed 15 Nov 2023.
17. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22(3):276–82.

18. Meyboom RH, Hekster YA, Egberts AC, Gribnau FW, Edwards IR. Causal or casual? The role of causality assessment in pharmacovigilance. *Drug Saf.* 1997;17(6):374–89.
19. Lindquist M. Data quality management in pharmacovigilance. *Drug Saf.* 2004;27(12):857–70.
20. Théophile H, Arimone Y, Miremont-Salamé G, Moore N, Fourrier-Réglat A, Haramburu F, et al. Comparison of three methods (consensual expert judgement, algorithmic and probabilistic approaches) of causality assessment of adverse drug reactions: an assessment using reports made to a French pharmacovigilance centre. *Drug Saf.* 2010;33(11):1045–54.
21. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast.* 1999;15(4):353–75.

Authors and Affiliations

Yrea R. J. van Rijt-Weetink^{1,2}  · Toine C. G. Egberts^{3,4}  · Florence P. A. M. van Hunsel^{1,2}  · David J. Lewis^{5,6}  · Laura M. Yates^{7,8}  · Ursula Winterfeld⁹  · Eugène P. van Puijenbroek^{1,2} 

✉ Eugène P. van Puijenbroek
e.vanpujenbroek@lareb.nl

Yrea R. J. van Rijt-Weetink
y.weetink@lareb.nl

Toine C. G. Egberts
a.c.g.egberts@uu.nl

Florence P. A. M. van Hunsel
f.vanhunsel@lareb.nl

David J. Lewis
david-1.lewis@novartis.com

Laura M. Yates
l.yates5@nhs.net

Ursula Winterfeld
ursula.winterfeld@chuv.ch

¹ Pharmacovigilance Centre Lareb, Goudsbloemvallei 7, 5237 MH s-Hertogenbosch, The Netherlands

² Groningen Research Institute of Pharmacy, Pharmacotherapy, -Epidemiology and -Economics, University of Groningen, Groningen, The Netherlands

³ Division of Pharmacoepidemiology and Clinical Pharmacology, Department of Pharmaceutical Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands

⁴ Department of Clinical Pharmacy, University Medical Centre Utrecht, Utrecht, The Netherlands

⁵ Global Drug Development, Novartis Pharma GmbH, Wehr, Germany

⁶ School of Life and Medical Sciences, University of Hertfordshire, Hatfield, UK

⁷ University of Kwazulu-Natal, KRISP, Durban, South Africa

⁸ Northern Genetics Service, Newcastle-upon-Tyne, UK

⁹ Swiss Teratogen Information Service and Clinical Pharmacology Service, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Lausanne, Switzerland