Drug Saf (2018) 41:389–401
https://doi.org/10.1007/s40264-017-0619-x

ORIGINAL RESEARCH ARTICLE

The RIMES Statement: A Checklist to Assess the Quality of Studies Evaluating Risk Minimization Programs for Medicinal Products

Meredith Y. Smith[1] · Andrea Russell[1,2] · Priya Bahri[3] · Peter G. M. Mol[4,5] · Sarah Frise[6,7] · Emily Freeman[8] · Elaine H. Morrato[9]

Published online: 7 December 2017

Abstract

Introduction Pharmaceutical risk minimization programs involve interventions designed to support safe and appropriate use of medicines. Currently, information regarding the evaluation of these programs is not publicly reported in a standardized and transparent manner. To address this gap, we developed and piloted a quality reporting checklist entitled the Reporting recommendations Intended for pharmaceutical risk Minimization Evaluation Studies (RIMES).

Methods Checklist development was guided by three sources: (1) a theoretical framework derived from program theory and process evaluation; (2) public health intervention design and evaluation principles; and (3) a review of existing quality reporting checklists. Two raters independently reviewed 10 recently published (2012–2016) risk minimization program evaluation studies using the proposed checklist. Inter-rater reliability of the checklist was assessed using Cohen's Kappa and Gwet's AC1.

Results A 43-item checklist was generated. Results indicated substantial inter-rater reliability overall ($\kappa = 0.65$, AC1 = 0.65) and for three (key information, design and evaluation) of the four subscales ($\kappa \geq 0.64$, AC1 ≥ 0.64). The fourth subscale (implementation) showed low reliability based on Cohen's Kappa, but substantial reliability based on the AC1 ($\kappa = 0.17$, AC1 = 0.61).

Conclusions The RIMES statement augments relevant elements from existing quality reporting guidelines with items that address aspects of intervention design, implementation and evaluation specific to pharmaceutical risk minimization programs. Our results show that the RIMES statement reliably measures key dimensions of reporting quality. This tailored checklist is an important first step in improving the reporting quality of risk minimization evaluation studies and may ultimately help to improve the quality of these interventions themselves.

✉ Meredith Y. Smith
mersmith@amgen.com

1 Global Patient Safety and Labeling, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA

2 Health Literacy and Learning Program, Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

3 European Medicines Agency, 30 Churchill Place, London E14 5EU, UK

4 Dutch Medicines Evaluation Board, Utrecht, The Netherlands

5 Department of Clinical Pharmacy and Pharmacology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

6 Astra Zeneca, Mississauga, ON, Canada

7 Dalla Lana School of Public Health (Epidemiology), University of Toronto, Toronto, ON, Canada

8 Pharmacovigilance and Patient Safety, AbbVie, North Chicago, IL, USA

9 Department of Health Systems, Management and Policy, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Key Points

We developed a 43-item checklist, entitled the RIMES statement, to assess the reporting quality of risk minimization evaluation studies in order to support more standardized, transparent reporting study results.

Our findings showed that the checklist had good inter-rater reliability, both overall and for the four subscales (Key information; Design; Implementation; and Evaluation).

We conclude with a proposal for further validating and refining the checklist to increase its practical appeal and usefulness.

1 Introduction

Ensuring the safe and appropriate use of medicines is an important public health priority, particularly in light of the rapid growth worldwide in prescription drug use [1]. Although product labeling serves as the basis for safe medication use, additional measures to minimize risks can be mandated by regulatory authorities in certain circumstances for products with serious safety concerns [2, 3]. These risk minimization programs can be imposed as either a condition of marketing authorization approval (most commonly), or as a condition to permit continued marketing authorization.

Marketing authorization holders of medicinal products are responsible for designing, implementing and evaluating these programs. Typically, however, sponsors must rely on healthcare professionals (alone or in conjunction with other third parties such as continuing medical education providers) to implement the actual intervention components [4, 5]. Other defining hallmarks of risk minimization interventions include the fact that they target multiple audiences (e.g. healthcare professionals, patients, caregivers, lay audiences), feature multiple measures or 'components' (e.g. risk communication, training of healthcare professionals, prescriber certification), span a range of socioecological levels (e.g. individual patient, healthcare system), involve multiple different types of implementers (e.g. physician prescriber, pharmacist, informal caregivers), and require implementation across multiple settings (e.g. inpatient, outpatient, home) and geographic areas (e.g. regions, countries, urban, rural).

Collectively, these characteristics define what is known as a 'complex' intervention [6]. Evaluating complex interventions requires ascertaining not only whether the actual intervention achieved the desired impact, but under what conditions it did so, for whom, and whether the impact was sustained over time [7]. Undesired or unanticipated impacts may also need to be captured, such as discontinuation of treatment or channeling towards inappropriate or suboptimal treatments.

Evidence of risk minimization program effectiveness is critical for demonstrating to regulatory authorities that a product's benefit-risk balance remains positive. Sponsors are encouraged to publish the results of their risk minimization program evaluations in order to build the risk minimization evidence base [3]. Additionally, the European Medicines Agency is legally required to make public both the protocols and abstracts of results of the post-authorization safety studies initiated, managed or financed by a marketing authorization holder, including those on risk minimization effectiveness [8].

Improving the effectiveness of risk minimization programs is a priority within the pharmacovigilance community [7–9]. However, to date, the number of risk minimization evaluation studies reported in the peer-reviewed literature has lagged far behind the number that have been implemented thus far [10]. Of those evaluations that have been published, methods and results have been inconsistently reported, making it difficult to evaluate their methodological quality and to interpret the results [11]. Common shortcomings in reporting include a failure to specify the intervention's purported causal mechanism(s) of risk minimization intervention and its relation to short-, intermediate- and long-term intended outcomes, inadequate information regarding the process of implementation and the healthcare context in which the intervention was delivered, limited correspondence between the stated intervention aim and the selected effectiveness measures, and an absence of predefined thresholds for effectiveness determination [12].

Over the past decade, numerous reporting checklists have been developed to standardize reporting of results of different types of studies, thereby building the evidence base for clinical and public health practices. Such checklists include, for example, CONSORT (clinical trials), STROBE and GRACE (observational and epidemiological studies), TREND (public health intervention evaluation studies), SQUIRE (healthcare systems), WIDER (knowledge transfer), and GREET (educational interventions and teaching) [13–20]. Recently, there has been a call to improve the evidence base also underlying risk minimization interventions [3, 11, 21]. In order to address this call, a standard is needed for gauging the reporting quality of risk minimization evaluation studies. Notably, however,

existing reporting checklists have limited applicability for the purposes of assessing the reporting quality of risk minimization evaluation studies. First, such checklists focus either on randomized designs or on one particular type of non-randomized design. To date, experimental study designs (e.g. randomized controlled trials), have not been used for the purposes of evaluating risk minimization programs because regulators have required that these programs be implemented across the entire targeted population. As a result, a variety of other non-randomized types of designs have been used (e.g. observational, interrupted times series), including mixed methods approaches that combine both qualitative (how, why) and quantitative (how much) research in order to gain a fuller understanding of the risk minimization program impact and the factors that contributed to its success or failure [11].

Extant checklists also fail to address why and how specific risk minimization program measures were selected, how they were designed, the process and context of program implementation, who was reached by the intervention, what 'dosage' amount was received (i.e. degree of exposure to program activities, such as, for example, completing all educational requirements), whether and to what extent different healthcare delivery settings adopted the program, and the degree to which intervention delivery was sustained over time. In particular, both the process and context of implementation are important to assess because risk minimization interventions, unlike clinical trials, are conducted under 'real-world' conditions in which both participants and participating settings are heterogeneous, implementers vary in terms of degree of commitment and relevant skills or expertise, and time and other resources are constrained. Information on the process and context of implementation can shed light on the mechanism(s) of change, help identify the circumstances under which the intervention works best, and aid in interpreting evaluation results, including negative, inconclusive, or positive intervention effects. Not least, it can also provide insight on unintended effects, whether negative or positive in nature [7, 22].

To address this gap, the Benefit-Risk Assessment, Communication, and Evaluation (BRACE) Special Interest Group (SIG) of the International Society for Pharmacoepidemiology (ISPE) sought to develop a common set of criteria to assess the quality of information reported in risk minimization evaluation studies [22, 23]. These criteria, designated as the Reporting recommendations Intended for pharmaceutical risk Minimization Evaluation Studies (RIMES) statement, were intended for use by regulatory bodies, industry, academic and journal editors and reviewers. The goals of the checklist were to (1) assess the quality of risk minimization evaluation studies; (2) improve the interpretation and usefulness of risk minimization evaluation study results; (3) increase awareness among key stakeholders regarding evidence-based standards in the field of risk minimization; (4) establish a reporting platform that bridges across the relevant sciences, including public health, health communication science, behavioral medicine, health services research and pharmacoepidemiology; and (5) promote, via reporting standardization and quality improvement, the inclusion of published risk minimization evaluation studies into systematic reviews. The latter goal is especially important given the paucity of published literature on risk minimization evaluation studies by drug or drug class. In this regard, the RIMES statement could help facilitate systematic reviews of evaluations of specific categories of risk minimization interventions (e.g. those pertaining to controlled distribution systems or healthcare provider communication plans), such as have been conducted for different types of behavioral health interventions [24].

The RIMES statement was developed explicitly as a tool for assessing the quality of the information reported in risk minimization evaluation studies, *not* as a checklist for evaluating the quality of these studies themselves. Ultimately, however, widespread adoption of the RIMES statement could lead to improvements in the quality of evaluation study design and, in so doing, generate better evidence on the effectiveness of risk minimization interventions for regulatory decision making. In turn, better evidence regarding the effectiveness of risk minimization interventions—which programs, for example, work best for whom and under what circumstances—should enhance the quality of risk minimization programs themselves [16]. The purpose of this study was to develop an initial version of the RIMES statement and test its reliability in a sample of recently published risk minimization evaluation studies.

2 Methods

We convened a multidisciplinary team of professionals (authoring team) with expertise in therapeutic risk management, regulatory science, public health, pharmacoepidemiology and behavioral medicine to develop the checklist. The development process involved a series of four consecutive steps consisting of (1) initial development of a checklist; (2) piloting; (3) individual checklist item revisions; and (4) inter-rater reliability testing. These steps are described in greater detail below.

2.1 Development of the Initial Version of the Checklist

The initial development of the checklist was guided by a theoretical framework, a review of existing reporting

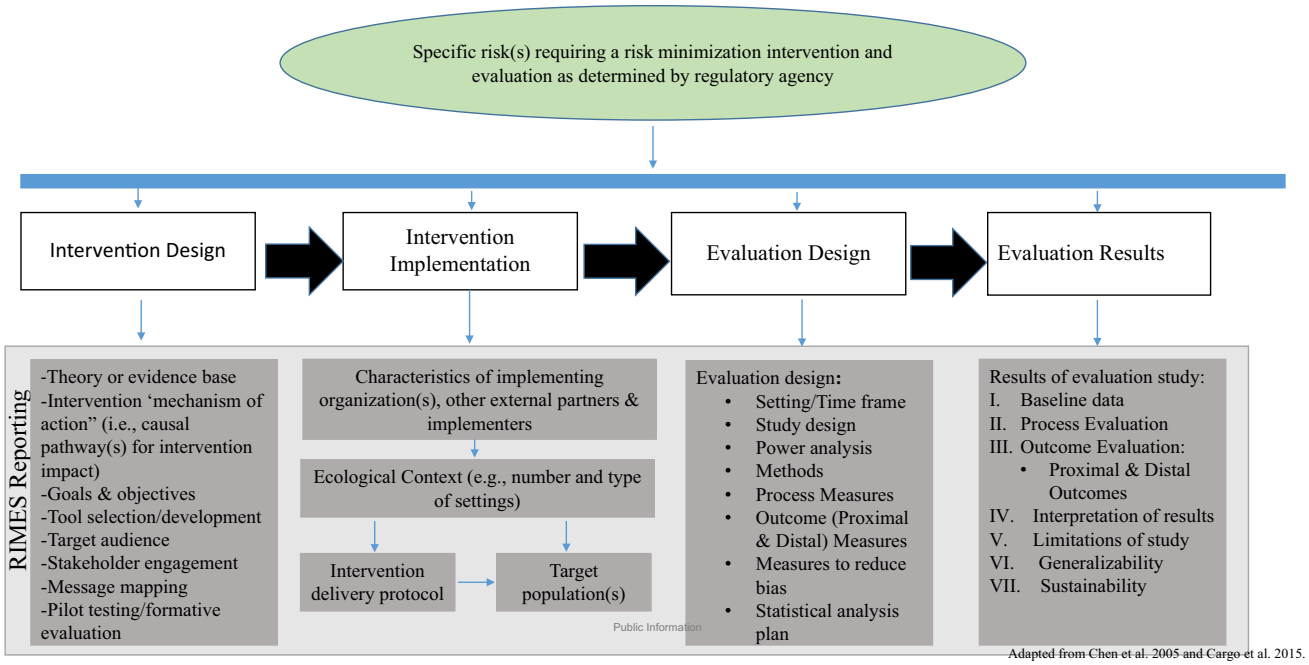

Fig. 1 Framework for risk minimization intervention evaluation study reporting criteria

checklists, and leading texts on public health and risk minimization intervention design, implementation and evaluation [14, 16, 25–34].

2.1.1 *Theoretical Framework*

To develop the RIMES statement, we adapted and combined relevant elements from existing program theory and process evaluation frameworks [7, 21, 35]. Our resulting framework (Fig. 1) emphasizes the stepwise contribution of design, implementation, and evaluation to the effectiveness of a complex intervention. Furthermore, it highlights the role of ecological context as an important contributor to intervention outcomes [32, 35]. Each of the items in our RIMES checklist falls within the elements of this framework. For example, a risk minimization program may be implemented in a range of outpatient care settings, each of which could differ in terms of leadership commitment to implementation, quality of staff training, and operating resources. An understanding of the role of, and interactions among, different contextual factors can also provide insight regarding how to optimize the fit of a risk minimization intervention to different delivery settings and how to improve its sustainability (i.e. long-term delivery) [36].

2.1.2 *Existing Checklists*

Many of the items included in the RIMES checklist refer to general research standards and are common to existing reporting checklists [14, 16, 17] but have been tailored to apply specifically to risk minimization. Items common to such reporting checklists relate to key information (author names, affiliations, conflict of interests, funding), descriptions related to study methods (participant recruitment, sample size, details of interventions, description of measures, and statistical analyses) and reporting of results (main results, limitations, generalizability and conclusions). We also consulted a checklist for implementation (Ch-IMP) and incorporated similar concepts (process metrics, implementer training, fidelity, adoption) into the RIMES checklist [7].

2.1.3 *Leading Texts in Public Health and Risk Minimization*

There are a number of known challenges to risk minimization programs and an emerging consensus regarding ways to advance the science in this field [11, 33]. For example, experts suggest that the goals of the intervention should be clearly defined, specific, measureable and time-bound. Thresholds of success should be determined *a priori*. When developing tools for communication, content should be tested among stakeholders (including intended audience) to ensure the message of risks is clearly conveyed. Furthermore, evaluations should address process outcomes (reach, adoption, implementation), as well as examine the results in the short term (effectiveness) and the success of the message in the long term (maintenance and sustainability). Each of these concepts helped inform the

contents of the draft RIMES checklist [9, 11, 14, 16, 25–34, 37].

2.2 Piloting

To explore this concept, we conducted an initial literature search of peer-reviewed published articles pertaining to formal risk minimization programs and evaluations. Specifically, we searched PubMed for English-language articles published between January 2000 and July 2016 using the following text words: ('risk minimization plan' OR 'risk evaluation and mitigation strateg*' OR 'risk management plan' OR 'risk minimization' OR 'risk minimisation' OR 'direct healthcare professional communication*' OR 'dear doctor' OR 'risk communication'.

Based on this literature search, we identified a convenience sample of 12 articles that met the following inclusion criteria: article relates to (1) a pharmaceutical product, (2) a risk communication or risk minimization intervention (including written, verbal or electronic), and (3) an assessment of the impact of the intervention [4, 5, 38–47]. Two raters (co-authors MYS and AR) separately reviewed and applied the draft checklist to each article. Of these two raters, MYS was an experienced researcher with extensive subject matter expertise. Conversely, AR had formal training and experience conducting health information communication research, but comparatively limited experience (less than 1 year) in designing and evaluating pharmaceutical risk minimization programs specifically.

2.3 Individual RIMES Checklist Item Revisions and Development of the Revised Checklist

Following the independent review of the 12 articles and application of the draft version of the RIMES checklist, the two raters met to discuss and compare item ratings. Based on that discussion, the checklist was further refined and the wording of several items was clarified to reduce ambiguity and to reflect single concepts only. Several examples were also added to guide future checklist application. The updated version of the checklist contained 45 items, with answer options scored as 0 (not reported or not applicable) or 1 (reported). These items were grouped into four domains:

1. *Key information*—includes established reporting criteria items such as adequate title, appropriate summary of the study in the abstract, valid, evidence-based study conclusions, and reporting of limitations as well as disclosures of funding and conflicts of interest.
2. *Description of the risk minimization program*—includes items that adequately describe the risk minimization program, such as the objective, design, target population and other key program elements.
3. *Implementation of the risk minimization program*—includes items that describe program implementation planning considerations, and how the program was implemented.
4. *Evaluation of the risk minimization program*—includes items that describe the study rationale, methods, implementation process measures, in particular the extent to which the program was implemented according to plan, and any factors that might have served to facilitate or impede implementation efforts, outcome measures and study results.

2.4 Inter-Rater Reliability Testing of the Revised Checklist

A second literature search of the published risk minimization evaluation literature was conducted approximately 6 months after the original search. The same search terms were employed. The inclusion criteria were the same as those used in the initial round of testing with two exceptions: (1) emphasis was placed on identifying only those articles evaluating risk minimization interventions formally required by a regulatory authority; and (2) the search timeframe was narrowed to include only those articles that had been published between January 2013 and January 2017. The purpose of restricting the timeframe was to focus the search on more recent studies that were expected to have higher reporting quality, given that they had been requested by a regulatory authority and had been published in the wake of European Union (EU) pharmacovigilance legislation that provided guidance on how sponsors were to evaluate formal risk minimization commitments. For the inter-rater reliability testing, we selected a convenience sample of the first 10 articles that met all the inclusion criteria (Table 1) [4, 5, 45–52].

The two raters independently reviewed the 10 articles, applying the revised checklist. Inter-rater reliability was reported using Cohen's kappa and Gwet's AC1 statistics. Statistical analysis was conducted in R version 3.3.2 using the 'irr' and 'lpSolve' packages. Interpretation of both statistics was based on Cohen's definition of agreement: poor (0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.0) [53]. Reporting inter-rater reliability with the kappa statistic is appropriate when the ratings have variation. The kappa statistic is sensitive to a high frequency of one score over another and may yield low reliability even when the percentage of agreement is high. This issue is known as the 'kappa paradox' and is described by Feinstein and Cicchetti [54]. In 2008, Gwet proposed and validated

Table 1 Description of articles reviewed with the RIMES statement

First author	Year	Study design	*N*	Risk communication intervention	Purpose of evaluation study
Bester et al. [48]	2016	Cross-sectional survey to assess awareness of brochure and understanding of brochure information	121 healthcare professionals	Brochure highlighting new and important adverse events	Conducted to determine the effectiveness of the educational brochure
Blanchette et al. [52]	2015	Analysis of pharmacy claims and medical claims for laboratory services at the time of drug initiation and within specified time intervals	Data from 742 patients prescribed medication	REMS elements including a dear doctor letter	Conducted to fulfill REMS commitment
Brody et al. [49]	2015	Cross-sectional survey of physicians to assess receipt of brochure and self-reported behaviors; analysis of EMR data	800 healthcare providers surveyed; Data from 7040 patients via EMR	Educational materials to inform practitioners of label changes and risks of the medication	Conducted to fulfill the Dutch Medicines Evaluation Board commitment
Cepeda et al. [4]	2016	Annual cross-sectional surveys of healthcare professionals and patients; drug utilization study of prescribing patterns; surveillance of abuse, misuse overdose, addiction and death associated with extended-release/long-acting opioids	Surveys: 612 healthcare professionals and 423 patients	Educational training course on safe and appropriate prescribing and use of extended-release/long-acting opioids	Conducted to fulfill REMS commitment
DiSantostefano et al. [51]	2017	Retrospective observational study of drug utilization via pharmacy data and employer-based claims data	Not applicable—all dispensing of medications across 7 years	REMS, including communication plan with letters to prescribers, printed and web-based information for HCPs, and letters to professional societies	Conducted to fulfill REMS commitment
Enger et al. [45]	2013	Retrospective analysis of Optum Research Database (US administrative claims database)	3568 patients	Medication guide	Conducted to fulfill REMS commitment
Hollingsworth et al. [46]	2016	Retrospective analysis of Medicare 5% sample dataset	Pre-REMS cohort: 1252 patients; post-REMS cohort: 949 patients	Black-box warning; other REMS materials not specified	Conducted to fulfill REMS commitment
Kraus et al. [5]	2013	Pre-test/post-test learning assessment	176,988 healthcare professionals	CME activity educational materials versus safe use alert	Conducted to fulfill REMS commitment
Smith et al. [47]	2012	Cross-sectional surveys	915 healthcare professionals	Distribution of product safety monograph, TB screening guidelines and TB screening checklist	Conducted to fulfill EMA 'additional risk minimization' PASS commitment
Tong et al. [50]		Pre-post intervention assessment of number of reported adverse drug reactions	36 cases of adverse events	REMS included prescribing program and medication guide	To examine incidence of AEs, trends in occurrence of AEs over a 9-year period, and clinical characteristics associated with AEs. To assess effectiveness of the REMS program

REMS risk evaluation and mitigation strategies, *EMR* electronic medical record, *HCPs* healthcare providers, *CME* continuing medical education, *TB* tuberculosis, *EMA* European Medicines Agency, *PASS* post-authorization safety study, *AEs* adverse events

the AC1 statistic as a way to address the limitations of kappa [55]. This statistic is less influenced by skewed ratings and is based on an alternative adjustment of chance that is defined as the "conditional probability that two, randomly selected rates will agree, given that no agreement will occur by chance" [56]. The Gwet's AC1 method has been used in other evaluations of inter-rater reliability of checklists and is often presented in conjunction with the kappa coefficient [7, 57, 58].

3 Results

The RIMES checklist was developed and then underwent two rounds of pilot testing. As a result of the inter-rater reliability analysis in the second round of testing, two items were deleted due to ambiguity in their phrasing—ambiguity that contributed to differing interpretations and poor inter-rater reliability. These items were (1) 'Explicit statement of causal assumptions linking intervention to a benefit for the recipient is provided', and (2) 'Upfront efforts to address potential sources of bias and confounding'. Based on further discussion, it was ultimately concluded that the first item substantially overlapped with an earlier item in the checklist ('Theoretical basis of the risk minimization program'). In addition, the second item was deemed as more accurately reflecting a quality study design item and, as such, covered by an earlier item ('Internal validity. Evaluation limitations, degree to which sources of potential bias were addressed'). After these eliminations, our final checklist consisted of 43 items (Table 2).

Rater scoring, percentage agreement, and reliability statistics for individual items can be found in Table 3. The frequency of each rater's scores are listed, where Y or N indicate the number of articles in which the rater determined the item was adequately covered or absent, respectively. For example, for Item 2b, Rater 1 determined nine articles fulfilled the criteria and one article did not fulfill criteria. For individual items, inter-rater agreement ranged from 40 to 100%, kappa coefficients ranged from -0.15 to 1.00, and AC1 coefficients ranged from -0.20 to 1.00. Slightly more than half ($n = 22$) of the kappa coefficients ranged from moderate to almost perfect, and slightly less than half were either fair ($n = 10$), slight ($n = 2$) or poor ($n = 7$). Two items (9b and 17f) had negative kappa coefficients, indicating that the reliability of raters was lower than what would be expected due to chance.

The reliability statistics for a number of items showed large discrepancies despite high or moderate percentage agreement (items 2b, 3b, 4, 5a, 8, 10a, 13a, 17f). For instance, for all of the kappa coefficients rated as poor or negative in value, the percentage agreement ranged between 70 and 90% and the AC1 statistic was 0.59 or higher. For these items, the raters' scoring patterns showed a high degree of skew such that either the item was scored consistently as being present or scored consistently as being absent.

One item resulted in 40% agreement, a slight kappa (0.12) and a negative AC1 statistic. This item passed the initial piloting of the checklist but emerged as a source of discordant ratings between the raters during reliability testing. During the final round of testing, the raters disagreed on the specificity required to give full credit on this item, with one rater being consistently more stringent than the other.

Summary statistics for the checklist overall and for subscales can be found in Table 4. We found the inter-rater reliability of the checklist overall to be substantial ($\kappa = 0.65$, AC1 = 0.65). Similarly, three of the four domains also showed substantial reliability based on the kappa: key information ($\kappa = 0.73$, AC1 = 0.80), design ($\kappa = 0.64$, AC1 = 0.64), and evaluation ($\kappa = 0.66$, AC1 = 0.69). The implementation domain ($\kappa = 0.17$, AC1 = 0.61) resulted in slight reliability based on the kappa coefficient, but higher reliability based on the AC1.

3.1 Respondent Burden

Initially, the average time raters spent reviewing and rating each article using the checklist was approximately 25 min; however, as familiarity with the checklist items increased, the average review time dropped to approximately 20 min per article.

4 Discussion

This article reports on the development of a set of criteria to describe the reporting quality of risk minimization intervention evaluation studies. Our results show that it is feasible to develop such a checklist despite the fact that these studies, by definition, must utilize non-randomized design types, may feature two or more substudies, and may employ a combination of both qualitative and quantitative research methods ('mixed methods') [4]. The checklist addresses important aspects of reporting that are vital to assessing the quality of a risk minimization evaluation study and that are under-represented in existing reporting checklists developed for other types of research studies and program evaluations. Examples of such key aspects include a description of the goals of the risk minimization program and the actual risk minimization measures used, how the program was implemented and whether implementation efforts were successful, and the inclusion of information regarding the external validity of evaluation results.

Table 2 RIMES statement: checklist of items that should be included in reports of risk minimization evaluation studies for medicinal products

Domain	Topic	Item	Descriptor and relevant examples
Key information	Declarations	1	Name(s) and affiliation(s) of the study sponsor(s) in the Conflicts of Interest statement and/or Acknowledgments statement
	Title and abstract	2a	Title mentioning type of evaluation study design, name of medicinal product(s), and target population/healthcare setting (all three required)
			Example: A drug utilization study to assess dispensing patterns at pharmacies for [*drug name*]
		2b	Structured abstract describing the purpose of the intervention and target recipient(s), evaluation methods, results and conclusions
	Discussion	3a	Summary of key results with reference to study objectives
		3b	Internal validity. Evaluation limitations, degree to which sources of potential bias were addressed, including both the direction and magnitude of any potential bias
		3c	External validity and generalizability (e.g. Will the intervention work across diverse populations and settings?)
		3d	Likelihood of sustainability. Discussion of the degree to which the intervention was integrated into the delivery setting (e.g. policies or incentives put in place to support ongoing intervention maintenance)
	Funding	4	Sources of evaluation study funding and other support, role of funders
Intervention Design	Design	5a	Goals and objectives of the risk minimization intervention
		5b	Implementation date of the risk minimization intervention
		5c	Theory or theories used to design intervention and/or risk minimization tools, including the expected causal pathway for intervention impact
			Example: The intervention was based on the theory of reasoned action
	Target population	6	Description of the key characteristics of geography and population targeted for intervention (i.e. age, sex, race/ethnicity, disease condition, socioeconomic status), enabling the reviewer/reader to determine if the evaluation study sample adequately reflected the targeted population
			Example: The risk minimization program targeted at US adults (*aged 18 years +*) *who have been prescribed* [*drug name*] *for the treatment of cardiovascular disease*
	Risk minimization tool selection and development	7a	Risk minimization tool(s) [e.g. managed distribution program; Medication Guide]
			Example: The risk minimization tools included a Dear HealthCare Professional Letter, and a Benefit-Risk Counseling tool for physicians
		7b	Pilot testing and formative evaluation of tools
		7c	Cultural sensitivity (i.e. reporting regarding whether local language, sociocultural values and traditions were considered when designing tools)
		7d	Stakeholder engagement, (i.e. patient and other stakeholder input considered/obtained in design of tools
		7e	Risk minimization tool message content (could be included in an online supplement or appendix)
			Example: Information on [drug name] risks, symptoms to watch for, and actions to take if symptoms presented themselves
		7f	Intervention distribution modality, including rationale for why a specific modality/ies were selected (the latter is recommended but not essential)
			Example: The tool was intended for distribution via Medscape email to physicians, journal advertisements, and a website posting
	Success metrics	8	A priori specification of measures and threshold for determination of intervention success

Table 2 continued

Domain	Topic	Item	Descriptor and relevant examples
Implement-ation	Setting	9a	Organizations responsible for implementing the intervention
		9b	Implementers of risk minimization intervention, including, for example, how they were selected and their qualifications
		9c	Training (i.e. did implementers receive training in the intervention and how to implement it?)
		9d	Ecological context (i.e. healthcare settings where the intervention was implemented (number, type and location[s])
	Fidelity	10a	Use of a formal protocol for implementing the intervention
		10b	Important intentional modifications made to risk minimization intervention after commencement (including at local level)
Evaluation	Hypotheses	12	Specific goals/objectives of the risk minimization evaluation study, including any hypotheses
			Example: We hypothesized that, as a result of distributing a brochure, 80% of physicians who prescribed [*drug name*] *would correctly identify the three key steps involved in screening patients for low blood pressure prior to initiating* [*drug name*] *therapy*
	Participants	13a	Eligibility requirements (i.e. inclusion and exclusion criteria) for participating in the evaluation study
			Example: Physicians were eligible to participate in the evaluation if they had prescribed [*drug name*] *to 10 + patients within the last 6 months*
		13b	Method of participant recruitment into evaluation study, including whether financial reimbursement was provided (code as zero for exceptions, e.g. secondary data analysis)
	Measures	14a	Process evaluation measures prespecified as a goal of the evaluation (e.g. reach, adoption, dose delivered, fidelity of implementation)
		14b	Primary and secondary outcome measures
		14c	Explicit link between evaluation study goals and methods in particular, and selection of processes and outcome measures
		14d	Sources of data and methods of measurement for each variable of interest
	Statistical analysis	15a	Study size calculation and power analysis (as applicable, depending on whether the study is qualitative or quantitative)
		15b	Statistical methods for analysis of primary and secondary outcomes
		15c	Explanation of missing data handling
	Results: process measures	16a	Results for each process evaluation measure
		16b	Description of factors that served to impede or facilitate intervention adoption and implementation
	Results: main outcomes	17a	A table showing baseline characteristics of the evaluation participants and evaluation settings (e.g. demographic, clinical, social, setting type, number and locations)
		17b	Results of participant recruitment (for human subjects research only), including dates and reasons for non-response or attrition rates (a participant flow diagram is strongly recommended but not required, not applicable for analysis of secondary dataset)
		17c	Description of primary and secondary outcome results
		17d	Precision of reporting of outcomes (e.g. 95% confidence interval) [as applicable, see above]
		17e	Description of whether primary outcome(s) exceeded a specified success threshold (as applicable, see above)
		17f	Results of any other analyses performed, including subgroup analyses, interactions and sensitivity analyses, distinguishing pre-specified from exploratory, identification of unintended impact of the risk minimization intervention or the evaluation study

Table 3 Inter-rater reliability testing: percentage agreement, Kappa and AC1 statistics by item

Subscale	Item	Rater 1		Rater 2		Percentage agreement	Kappa	AC1 statistic
Key information	1	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	2a	Y: 2	N: 8	Y: 2	N: 8	80	0.38	0.71
	2b	Y: 9	N: 1	Y: 10	N: 0	90	0	0.89
	3a	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	3b	Y: 9	N: 1	Y: 10	N: 1	90	0	0.89
	3c	Y: 4	N: 6	Y: 3	N: 7	70	0.35	0.45
	3d	Y: 2	N: 8	Y: 1	N: 9	90	0.62	0.87
	4	Y: 9	N: 1	Y: 10	N: 0	90	0	0.89
Intervention description	5a	Y: 10	N: 0	Y: 8	N: 2	80	0	0.76
	5b	Y: 9	N: 1	Y: 7	N: 3	80	0.41	0.71
	5c	Y: 0	N: 10	Y: 0	N: 10	100	1.00	1.00
	6	Y: 3	N: 7	Y: 1	N: 9	80	0.41	0.71
	7a	Y: 8	N: 2	Y: 7	N: 3	70	0.21	0.52
	7b	Y: 8	N: 2	Y: 9	N: 1	90	0.62	0.87
	7c	Y: 0	N: 10	Y: 0	N: 10	100	1.00	1.00
	7d	Y: 2	N: 8	Y: 1	N: 9	90	0.62	0.87
	7e	Y: 6	N: 4	Y: 9	N: 1	70	0.29	0.52
	7f	Y: 6	N: 4	Y: 6	N: 4	60	0.17	0.23
	8	Y: 0	N: 10	Y: 2	N: 8	80	0	0.76
Implementation	9a	Y: 2	N: 8	Y: 8	N: 2	40	0.12	− 0.20
	9b	Y: 2	N: 8	Y: 1	N: 9	70	− 0.15	0.60
	9c	Y: 0	N: 10	Y: 0	N: 10	100	1.00	1.00
	9d	Y: 4	N: 6	Y: 1	N: 9	70	0.29	0.52
	10a	Y: 1	N: 9	Y: 0	N: 10	90	0	0.89
	10b	Y: 3	N: 7	Y: 2	N: 8	70	0.21	0.52
Evaluation	12	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	13a	Y: 7	N: 3	Y: 10	N: 0	70	0	0.59
	13b	Y: 5	N: 5	Y: 4	N: 6	70	0.40	0.41
	14a	Y: 3	N: 7	Y: 3	N: 7	80	0.52	0.66
	14b	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	14c	Y: 7	N: 3	Y: 7	N: 3	100	1.00	1.00
	14d	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	15a	Y: 3	N: 7	Y: 1	N: 9	80	0.41	0.71
	15b	Y: 8	N: 2	Y: 9	N: 1	90	0.62	0.87
	15c	Y: 2	N: 8	Y: 1	N: 9	90	0.62	0.87
	16a	Y: 5	N: 5	Y: 3	N: 7	80	0.60	0.62
	16b	Y: 2	N: 8	Y: 5	N: 5	70	0.40	0.45
	17a	Y: 7	N: 3	Y: 6	N: 4	70	0.35	0.45
	17b	Y: 3	N: 7	Y: 6	N: 4	70	0.44	0.41
	17c	Y: 10	N: 0	Y: 10	N: 0	100	1.00	1.00
	17d	Y: 8	N: 2	Y: 8	N: 2	80	0.38	0.71
	17e	Y: 1	N: 9	Y: 3	N: 7	80	0.41	0.71
	17f	Y: 9	N: 1	Y: 9	N: 1	80	− 0.11	0.71

The RIMES statement is intended for use by a range of audiences, including regulatory, industry, academic evaluators and journal editors. Standardized reporting of risk minimization evaluation studies, such as that provided by the RIMES statement, can facilitate systematic reviews and data synthesis, including meta-analyses. This is a

Table 4 Inter-rater reliability summary statistics

	Kappa	AC1
Overall	0.65	0.65
By subscale		
Key information	0.73	0.80
Intervention design	0.64	0.64
Implementation	0.17	0.60
Evaluation	0.66	0.69

particularly important feature given that, to date, pharmaceutical risk minimization has been singularly uninformed by research findings from other relevant sciences, including public health, communication, behavioral medicine and health services research. In addition, the checklist can guide research planning and manuscript development in the first instance, and serve as a platform for bridging pharmaceutical risk minimization science with other relevant fields. Not least, it can also assist sponsors in designing higher-quality risk minimization evaluation studies, and through learning from this evidence may potentially enhance the quality and effectiveness of risk minimization programs themselves.

The main limitation of this study related to the relatively small size (n = 10) of the sample of articles reviewed by only two reviewers. With a larger sample of articles or additional reviewers, both the kappa and AC1 statistics would have been more reliable and the rates of discrepancy between them would have been reduced. However, as noted previously, we were limited in our sample size by the relative paucity of articles on risk minimization evaluation studies that have been published in the peer-reviewed literature to date.

5 Conclusions

Results of preliminary reliability testing show that the RIMES statement has good inter-rater reliability among a small sample of articles. Important next steps in its development would include conducting testing among a larger sample to confirm item reliability, particularly for items in this analysis that have low kappa coefficients. It is possible some items are underperforming and may require adjustment. In addition, formal usability testing and an examination of both the content and construct validity of the checklist based on a more comprehensive and systematic assessment of relevant publications, including those found in the grey literature. To enhance the checklist's practicality, future research should also assess ways to streamline it further, potentially via factor analytic methods, and to explore possible approaches to item weighting. Not least, to aid in standardizing the interpretation of checklist items, a user manual should be developed.

Although additional methodological work is planned, the current version of the checklist showed good reliability when tested by two raters among a small sample of articles. As such, this checklist represents an important step forward in improving the quality of reporting of risk minimization evaluation studies, one that can benefit both the science of pharmaceutical risk minimization and, ultimately, patient safety.

Acknowledgements The authors would like to thank Han-Yao Huang, Dimitri Bennett and Douglas Watson for their review and comment on the final draft of this manuscript. All authors (with the exception of AR) collaborated on this manuscript as members of the Benefit-Risk Assessment, Communication and Evaluation (BRACE) SIG for the International Society of Pharmacoepidemiology (ISPE). The views expressed herein are the authors' personal views and not necessarily those of ISPE or any of its SIGs, and may also not be understood or quoted as being made on behalf of or reflecting the position of the organizations with which the authors are affiliated.

Compliance with Ethical Standards

Conflict of interest The following authors are employed by pharmaceutical companies: Meredith Y. Smith (Amgen), Sarah Frise (AstraZeneca), and Emily Freeman (AbbVie). Andrea Russell receives a graduate fellowship stipend from Amgen. Co-authors Priya Bahri, and Peter Mol have no conflicts of interest to declare. Elaine Morrato has received consulting fees and travel reimbursement from the American Academy of Pediatrics, Amgen, and the Merck Foundation. She has received a speaking honorarium, a manuscript preparation honorarium and travel reimbursement from the PhRMA Foundation, and is a Special Government Employee and advises the FDA on issues of drug safety and risk management.

Funding No sources of funding were used in the preparation of this study and manuscript.

References

1. Quintiles IMS Institute. Global medicines use in 2020: outlook and implications. Parsippany, NJ: IMS Institute for Healthcare Informatics. 2015. https://s3.amazonaws.com/assets.fiercemarkets.net/public/005-LifeSciences/imsglobalreport.pdf. Accessed 6 Jul 2017.
2. US Food and Drug Administration. Guidance for Industry Format and Content of Proposed Risk Evaluation and Mitigation Strategies (REMS), REMS Assessments, and Proposed REMS Modifications. Rockville, MD: US Department of Health and

Human Services. 2009. https://www.fda.gov/downloads/Drugs/…/Guidances/UCM184128.pdf. Accessed 15 May 2017.
3. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module XVI–Risk minimisation measures: selection of tools and effectiveness indicators (Rev 2). London: European Medicines Agency. 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/02/WC500162051.pdf. Accessed 15 May 2017.
4. Cepeda MS, Coplan PM, Kopper NW, Maziere JY, Wedin GP, Wallace LE. ER/LA opioid analgesics REMS: overview of ongoing assessments of its progress and its impact on health outcomes. Pain Med. 2017;18(1):78–85.
5. Kraus CN, Baldwin AT, McAllister RG Jr. Improving the effect of FDA-mandated drug safety alerts with Internet-based continuing medical education. Curr Drug Saf. 2013;8(1):11–6.
6. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.
7. Cargo M, Stankov I, Thomas J, Saini M, Rogers P, Mayo-Wilson E, et al. Development, inter-rater reliability and feasibility of a checklist to assess implementation (Ch-IMP) in systematic reviews: the case of provider-based prevention and treatment programs targeting children and youth. BMC Med Res Methodol. 2015;15:73.
8. Regulation (EC) No 726/2004 of the European Parliament and of the Council, Article 26(1)(h)2004. https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/reg_2004_726/reg_2004_726_en.pdf. Accessed 15 May 2017.
9. Bahri P, Dodoo AN, Edwards BD, Edwards IR, Fermont I, Hagemann U, et al. The ISoP CommSIG for improving medicinal product risk communication: A new special interest group of the international society of pharmacovigilance. Drug Saf. 2015;38(7):621–7.
10. US Food and Drug Administration. Approved Risk Evaluation and Mitigation Strategies (REMS). Silver Spring, MD: US Food and Drug Administration [date unknown]. https://www.accessdata.fda.gov/scripts/cder/rems/index.cfm. Accessed 1 May 2017.
11. Smith MY, Morrato E. Advancing the field of pharmaceutical risk minimization through application of implementation science best practices. Drug Saf. 2014;37(8):569–80.
12. Gridchyna I, Cloutier AM, Nkeng L, Craig C, Frise S, Moride Y. Methodological gaps in the assessment of risk minimization interventions: a systematic review. Pharmacoepidemiol Drug Saf. 2014;23(6):572–9.
13. Armstrong R, Waters E, Jackson N, Oliver S, Popay J, Shepherd J, et al. Guidelines for systematic reviews of health promotion and public health interventions. Australia: Melbourne University. 2007. https://ph.cochrane.org/sites/ph.cochrane.org/files/public/uploads/Guidelines%20HP_PH%20reviews.pdf. Accessed 2 July 2017.
14. Schulz KF, Altman DG, Moher D, Group C. Consort 2010 statement. Updated guidelines for reporting parallel group randomized trials. Ann Intern Med. 2010;152(11):726–32.
15. Dreyer NA, Bryant A, Velentgas P. The GRACE checklist: a validated assessment tool for high quality observational studies of comparative effectiveness. J Manag Care Spec Pharm. 2016;22(10):1107–13.
16. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806–8.
17. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–6.
18. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf. 2016;25(12):986–92.
19. Albrecht L, Archibald M, Arseneau D, Scott SD. Development of a checklist to assess the quality of reporting of knowledge translation interventions using the workgroup for intervention development and evaluation research (WIDER) recommendations. Implement Sci. 2013;16(8):52.
20. Phillips AC, Lewis LK, McEvoy MP, Galipeau J, Glasziou P, Moher D, et al. Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET). BMC Med Educ. 2016;16(1):237.
21. Council for International Organizations of Medical Sciences (CIOMS). Practical approaches to risk minimisation for medicinal products: Report of CIOMS Working Group IX. Geneva, Switzerland: Council for International Organizations of Medical Sciences (CIOMS); 2014. p. 200.
22. Klesges LM, Estabrooks PA, Dzewaltowski DA, Bull SS, Glasgow RE. Beginning with the application in mind: designing and planning health behavior change interventions to enhance dissemination. Ann Behav Med. 2005;29(2):66–75.
23. Radawski C, Morrato E, Hornbuckle K, Bahri P, Smith M, Juhaeri J, et al. Benefit-Risk Assessment, Communication, and Evaluation (BRACE) throughout the life cycle of therapeutic products: overall perspective and role of the pharmacoepidemiologist. Pharmacoepidemiol Drug Saf. 2015;24(12):1233–40.
24. Asnani MR, Quimby KR, Bennett NR, Francis DK. Interventions for patients and caregivers to improve knowledge of sickle cell disease and recognition of its related complications. Cochrane Database Syst Rev. 2016;6(10):CD011175.
25. Glanz K, Rimer BK, Viswanath K. Health behavior and health education: theory, research, and practice. San Francisco, CA: Jossey-Bass. 2002. http://riskybusiness.web.unc.edu/files/2015/01/Health-Behavior-and-Health-Education.pdf. Accessed 10 Jan 2017.
26. Bartholomew LKPG, Kok G, Gottlieb N. Intervention mapping: designing theory and evidence-based health promotion programs. Mountain View: Mayfield Publishing; 2001.
27. Fischhoff B, Brewer NT, Downs JS. Communicating risks and benefits: an evidence-based user's guide. Silver Spring: US Dept. of Health and Human Services, Food and Drug Administration; 2011.
28. Kotler P, Lee N. Social marketing: influencing behaviors for good. 3rd ed. Los Angeles: Sage Publications; 2008.
29. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 3). London: European Medicines Agency, 2014. Report No: EMA/95098/2010. http://www.encepp.eu/standards_and_guidances. Accessed 26 Jun 2017.
30. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ. 2015;350:h2147.
31. Prieto L, Spooner A, Hidalgo-Simon A, Rubino A, Kurz X, Arlett P. Evaluation of the effectiveness of risk minimization measures. Pharmacoepidemiol Drug Saf. 2012;21(8):896–9.
32. Chen HT. Practical program evaluation: assessing and improving planning, implementation, and effectiveness. Thousand Oaks: Sage Publications; 2005.
33. Banerjee AK, Mayall SJ, Smith MY. Therapeutic risk management of medicines, Chapter 11: evaluating the effectiveness of risk minimization. 1st ed. Cambridge: Woodhead Publishing; 2014. p. 241–76.
34. Smith MY. Benefit-risk management: current trends and future directions. In: Sietsema WK, Sprafka JM, editors. Risk

management principles for devices and pharmaceuticals: global perspectives on the benefit: risk assessment of medicinal products, chapter 6, 2nd ed. Rockville: Regulatory Affairs Professionals Society; 2018 **(in press)**.
35. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350:h1258.
36. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci. 2013;8(1):117.
37. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. Am J Public Health. 2013;103(6):e38–46.
38. Ishihara L, Beck M, Travis S, Akintayo O, Brickel N. Physician and pharmacist understanding of the risk of urinary retention with retigabine (ezogabine): A REMS assessment survey. Drugs Real World Outcomes. 2015;2(4):335–44.
39. Hagiwara H, Nakano S, Ogawa Y, Tohkin M. The effectiveness of risk communication regarding drug safety information: a nationwide survey by the Japanese public health insurance claims data. J Clin Pharm Ther. 2015;40(3):273–8.
40. Sumi E, Yamazaki T, Tanaka S, Yamamoto K, Nakayama T, Bessho K, et al. The increase in prescriptions of bisphosphonates and the incidence proportion of osteonecrosis of the jaw after risk communication activities in Japan: a hospital-based cohort study. Pharmacoepidemiol Drug Saf. 2014;23(4):398–405.
41. Piening S, de Graeff PA, Straus SM, Haaijer-Ruskamp FM, Mol PG. The additional value of an e-mail to inform healthcare professionals of a drug safety issue: a randomized controlled trial in the Netherlands. Drug Saf. 2013;36(9):723–31.
42. Kimura T, Shiosakai K, Takeda Y, Takahashi S, Kobayashi M, Sakaguchi M. Quantitative evaluation of compliance with recommendation for sulfonylurea dose co-administered with DPP-4 inhibitors in Japan. Pharmaceutics. 2012;4(3):479–93.
43. Arana A, Allen S, Burkowitz J, Fantoni V, Ghatnekar O, Rico MT, et al. Infliximab paediatric Crohn's disease educational plan: a European, cross-sectional, multicentre evaluation. Drug Saf. 2010;33(6):489–501.
44. Chol C, Guy C, Jacquet A, Castot-Villepelet A, Kreft-Jais C, Cambazard F, et al. Complications of BCG vaccine SSI(R) recent story and risk management plan: the French experience. Pharmacoepidemiol Drug Saf. 2013;22(4):359–64.
45. Enger C, Younus M, Petronis KR, Mo J, Gately R, Seeger JD. The effectiveness of varenicline medication guide for conveying safety information to patients: a REMS assessment survey. Pharmacoepidemiol Drug Saf. 2013;22(7):705–15.
46. Hollingsworth K, Romney MC, Crawford A, McAna J. The impact of the risk evaluation mitigation strategy for erythropoiesis-stimulating agents on their use and the incidence of stroke in Medicare subjects with chemotherapy-induced anemia with lung and/or breast cancers. Popul Health Manag. 2016;19(1):63–9.
47. Smith MY, Attig B, McNamee L, Eagle T. Tuberculosis screening in prescribers of anti-tumor necrosis factor therapy in the European Union. Int J Tuberc Lung Dis. 2012;16(9):1168–73.
48. Bester N, Di Vito-Smith M, McGarry T, Riffkin M, Kaehler S, Pilot R, et al. The effectiveness of an educational brochure as a risk minimization activity to communicate important rare adverse events to health-care professionals. Adv Ther. 2016;33(2):167–77.
49. Brody RS, Liss CL, Wray H, Iovin R, Michaylira C, Muthutantri A, et al. Effectiveness of a risk-minimization activity involving physician education on metabolic monitoring of patients receiving quetiapine: results from two postauthorization safety studies. Int Clin Psychopharmacol. 2016;31(1):34–41.
50. Tong K, Nicandro JP, Shringarpure R, Chuang E, Chang L. A 9-year evaluation of temporal trends in alosetron postmarketing safety under the risk management program. Ther Adv Gastroenterol. 2013;6(5):344–57.
51. DiSantostefano RL, Yeakey AM, Raphiou I, Stempel DA. An evaluation of asthma medication utilization for risk evaluation and mitigation strategies (REMS) in the United States: 2005–2011. J Asthma. 2013;50(7):776–82.
52. Blanchette CM, Nunes AP, Lin ND, Mortimer KM, Noone J, Tangirala K, et al. Adherence to risk evaluation and mitigation strategies (REMS) requirements for monthly testing of liver function. Drugs Context. 2015;4:212272.
53. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
54. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543–9.
55. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(Pt 1):29–48.
56. Gwet KL. Handbook of inter-rater reliability. Gaithersburg: Advanced Analytics, LLC; 2014.
57. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61.
58. Borek AJ, Abraham C, Smith JR, Greaves CJ, Tarrant M. A checklist to improve reporting of group-based behaviour-change interventions. BMC Public Health. 2015;15:963.