



**HAL**  
open science

**A new risk of bias checklist applicable to randomized trials, observational studies, and systematic reviews was developed and validated to be used for systematic reviews focusing on drug adverse events**

Jean-Luc Faillie, Pili Ferrer, Amandine Gouverneur, Damien Driot, Shoma Berkemeyer, Xavier Vidal, Maria José Martinez-Zapata, Consuelo Huerta, Xavier Castells, Marietta Rottenkolber, et al.

► **To cite this version:**

Jean-Luc Faillie, Pili Ferrer, Amandine Gouverneur, Damien Driot, Shoma Berkemeyer, et al.. A new risk of bias checklist applicable to randomized trials, observational studies, and systematic reviews was developed and validated to be used for systematic reviews focusing on drug adverse events. *Journal of Clinical Epidemiology*, 2017, 86, pp.168–175. 10.1016/j.jclinepi.2017.04.023 . hal-01761705

**HAL Id: hal-01761705**

**<https://hal.science/hal-01761705v1>**

Submitted on 6 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **A new risk of bias checklist applicable to randomized trials, observational studies, and systematic reviews was developed and validated to be used for systematic reviews focusing on drug adverse events**

## **Authors**

Jean-Luc Faillie<sup>1,2</sup>, Pili Ferrer<sup>3</sup>, Amandine Gouverneur<sup>4</sup>, Damien Driot<sup>5</sup>, Shoma Berkemeyer<sup>6</sup>, Xavier Vidal<sup>3</sup>, Maria José Martínez-Zapata<sup>7</sup>, Consuelo Huerta<sup>8</sup>, Xavier Castells<sup>9</sup>, Marietta Rottenkolber<sup>10</sup>, Sven Schmiedl<sup>11,12</sup>, Mònica Sabaté<sup>3</sup>, Elena Ballarín<sup>3</sup> and Luisa Ibáñez<sup>3</sup>.

## **Affiliations**

<sup>1</sup>Laboratory of Biostatistics, Epidemiology and Public Health (EA2415), Faculty of Medicine, University of Montpellier, Montpellier, France

<sup>2</sup>Department of Medical Pharmacology and Toxicology, CHU Montpellier University Hospital, Montpellier, France

<sup>3</sup>Catalan Institute of Pharmacology Foundation (FICF), Department of Pharmacology Therapeutics and Toxicology, Autonomous University of Barcelona, Clinical Pharmacology Unit, Vall d'Hebron University Hospital, Barcelona, Spain

<sup>4</sup>Inserm U1219, University of Bordeaux, Department of Medical Pharmacology, CHU Bordeaux University Hospital, Bordeaux, France

<sup>5</sup>University of Toulouse, Department of Clinical and Medical Pharmacology, CHU Toulouse University Hospital, Toulouse, France

<sup>6</sup>Department of Community Health, Hochschule fuer Gesundheit, Bochum, Germany

<sup>7</sup>Iberoamerican Cochrane Centre. Institute of Biomedical Research (IIB Sant Pau), Barcelona. CIBER de Epidemiología y Salud Pública (CIBERESP), Spain

<sup>8</sup>Agencia Española de Medicamentos y Productos Sanitarios (AEMPS), Madrid, Spain.

<sup>9</sup>TransLab Research Group, Department of Medical Sciences, Girona University, Girona, Spain.

<sup>10</sup>Diabetes Research Group, Medizinische Klinik und Poliklinik IV, Klinikum der Universitaet, Munich, Germany

<sup>11</sup>Department of Clinical Pharmacology, School of Medicine, Faculty of Health, Witten/Herdecke University, Witten, Germany

<sup>12</sup>Philipp Klee-Institute for Clinical Pharmacology, HELIOS Clinic Wuppertal, Wuppertal, Germany

## **Corresponding author:**

Luisa Ibáñez, MD PhD

Fundació Institut Català de Farmacologia, Hospital Universitari Vall d'Hebron,

Pg Vall d'Hebron 119-129, Barcelona 08029

Phone: +34 93 489 41 05

E-mail: [li@icf.uab.es](mailto:li@icf.uab.es)

## **Abstract**

### **Objective**

To develop and validate an adequate tool to evaluate the risk of bias of randomized controlled trials, observational studies, and systematic reviews assessing drug adverse events.

### **Study Design and Setting**

We developed a structured risk of bias checklist applicable to randomized trials, cohort, case-control and nested case-control studies, and systematic reviews focusing on drug safety. Face and content validity was judged by three experienced reviewers. Inter- and intra-rater reliability was determined using 20 randomly selected studies, assessed by three other independent reviewers including one performing a 3-week retest.

### **Results**

The developed checklist examines eight domains: study design and objectives, selection bias, attrition, adverse events information bias, other information bias, statistical methods to control confounding, other statistical methods, and conflicts of interest. The total number of questions varied from ten to 32 depending on the study design. Inter- and intra-rater agreement were fair with Kendall's W of 0.70 and 0.74, respectively. Median time to complete the checklist was 8.5 minutes.

### **Conclusion**

The developed checklist showed face and content validity and acceptable reliability to assess the risk of bias for studies analyzing drug adverse events. Hence, it might be considered as a novel useful tool for systematic reviews and meta-analyses focusing on drug safety.

**Key words:** risk of bias; study quality; systematic review; meta-analysis; adverse drug events; checklist.

**Running title:** *The PROTECT Checklist*

**Word count (abstract): 203**

**Word count (text): 3424**

## What is new

### ***Key findings***

- In systematic reviews analyzing drug safety, assessing the risk of bias of included studies requires careful consideration.
- We developed and validated a structured risk of bias checklist applicable to randomized trials, cohort, case-control and nested case-control studies, and systematic reviews assessing drug safety.
- The developed checklist proved to be a valid and acceptably reliable instrument to assess the risk of bias.

### ***What this adds to what is known***

- The developed checklist is the first standard method that is specific to studies analyzing drug adverse events and applicable to all study designs.

### ***What is the implication, what should change now***

- The developed checklist might be considered as a novel useful tool for systematic reviews and meta-analyses focusing on adverse effects of drugs.

## 1. Introduction

Systematic reviews are considered the highest level of evidence in biomedical research, since they collate and synthesize results of all relevant studies on the effects of an intervention. However, the reliability of the evidence provided by systematic reviews depends on the assumption that the studies included are not biased. Thus, assessing the quality of the studies included in systematic reviews is essential to help interpreting the findings.

The concept of study quality suffers from the absence of a unanimously accepted definition. Different approaches are used to assess it: (i) by determining the completeness of reporting of different study features, according to reporting guidelines (e.g. CONSORT for randomized clinical trials) [1]; (ii) by determining the validity as the absence of threats to the internal and external validity of a study [2, 3]; or, (iii) by considering internal validity (susceptibility of bias) as the best way to assess the quality of a study [4, 5]. Report completeness is considered to be not sufficient to evaluate a study, hence, several tools have been developed to assess the quality/risk of bias for randomized clinical trials [6, 7], observational studies [8-11], both randomized trials and observational studies [3] and systematic reviews [12, 13].

When conducting a systematic review of drug adverse events, there are several specific factors related to the adverse event that needs to be considered as they may affect the quality of the assessment: the definition of the drug adverse event and its degree of severity; the explanation of how drug adverse events were identified and reported, including the temporal relationship between the ingestion of the drug and their onset; the registration frequency of adverse events during the study; the use of a validated instrument for reporting drug adverse events; and, the methodology to attribute the adverse events to the intervention [14]. Nevertheless, the risk of bias of included studies must be carefully considered in systematic reviews analyzing drug adverse events especially since such reviews frequently consider different study designs (randomized trials, observational studies and other systematic reviews). Currently, there exists no adequate and validated method to assess the risk of bias applicable to all the types of studies focusing on drug adverse events [14-16]. Several risk of bias tools have been used in this purpose but they did not take into account the specificity of drug adverse event assessment and were often focused on one specific study design. In 2008, Santaguida *et al.* developed a tool (the McHarm scale) to evaluate the quality of drug adverse event assessment in both randomized and observational studies. However, the scale consisted in a 15-items scale focusing principally in the reporting of drug adverse events (the original article by Santaguida *et al.* is no longer available but the scale has been reproduced by Chou *et al.* [17]).

In general, risk of bias assessment of studies included, in systematic reviews of adverse drug events, is poorly reported [18, 19]. When reported, the authors usually measure the quality of the studies by either using multiple tools (e.g. combining the Cochrane Risk of Bias Tool and the

McHarm scale) [20] or by building new and not formally validated tools to assess specific drug adverse effects [21-24]. Furthermore, instead of checklists, the use of rating scales has been shown to be not adequate for risk of bias assessment [25]. Recently, the Cochrane Collaboration developed a risk of bias tool for observational studies, however, although it has been used for studies of drug safety, it was not designed for the specific biases of drug adverse events assessment [11, 26].

The Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) project is an European collaborative project addressing limitations of current methods in the field of pharmacoepidemiology and pharmacovigilance [27]. As part of its objectives, systematic reviews have been conducted to assess the effects of antiepileptic drugs on suicide risk [28], macrolides and amoxicillin/clavulanate on acute liver injury [29], calcium channel blockers on malignancies, or inhaled beta2-agonists/muscarinic antagonists on acute myocardial infarction [30]. During these studies, PROTECT investigators faced the absence of an adequate tool to evaluate risk of bias and an *ad hoc* checklist was built and tested [28, 30]. Hence, the objective of the present work was to validate a checklist designed to assess the risk of bias of randomized trials and observational studies included in systematic reviews focused on drug adverse events.

## **2. Methods**

### **2.1 Development of the checklist**

#### **2.1.1 Construct definition**

In the present study, we examined the specific risk of bias in assessing drug adverse events and considered it as a synonym of internal validity, that is, according to the Cochrane group, “the extent to which the design and conduct of a study are likely to have prevented bias” [31]. As it is stated by the Agency for Healthcare Research and Quality, risk of bias assessment does not take into consideration the statistical power of the study nor the external validity of their results [32].

#### **2.1.2 Terminology**

In the checklist, the term “drug adverse events” is used as generic term to define undesirable outcomes that occurred during or after the use of a drug, whether the causal relationship with the drug had been established or not.

#### **2.1.3 Included study designs**

The developed checklist aims to assess the risk of bias for the different types of study used to assess drug safety and therefore potentially included in a systematic review: randomized controlled trials, cohort studies, case-control studies (including nested case-control studies) and systematic reviews. We included systematic reviews as they can be part of the studies used to perform or update systematic reviews [33] or perform overviews of reviews which, despite their relevance for practical and clinical issues, often lack risk of bias assessment [34-38].

#### **2.1.4 Format and structure**

We chose to develop a checklist composed of a list of items (questions) structured in several relevant domains. Depending on the nature of the item, the response alternatives vary from two (“Yes” or “No”) to four (“Not applicable”, “Yes”, “Unclear”, or “No”). The checklist includes one summary risk of bias assessment relative to each domain. At the end of the checklist, a final summary risk of bias assessment is asked for the whole study. The format used for the responses to assess risk of bias is: “Low”, “Unclear” or “High”. Since several items are only relevant for specific study designs, each item indicates which study design(s) it concerns. Consequently, the total number of items varies depending on the study design.

#### **2.1.5 Checklist content**

In the pilot version of the checklist, the items and domains were derived from several sources: previous quality assessment instruments (Downs and Black scale [3], Cochrane risk of bias tool [6], and AMSTAR scale [12]), reviews of the literature [16, 39], and recommendations from the Agency for Healthcare Research and Quality for assessing risk of bias [32] and drug adverse events [17]. Since secondary database studies are widely used in pharmacoepidemiology, we



adapted the existing items to their specificities. In addition, items relative to the quality of collecting information on adverse events in randomized clinical trials of efficacy were derived from Loke *et al.* [40].

## **2.2 Validity study**

### **2.2.1 Face and content validity**

Three senior clinical pharmacologists, experts in drug safety assessment (CH, MJMZ and XC), evaluated the face and content validity of the pilot version of the checklist. The number of raters who evaluated the face and content validity and reliability was chosen based on the methodology of the previous validity study of the Downs and Black scale [3]. On one hand, face validity indicates whether, on the surface, the items appear to be measuring what they actually are implying that every item is not ambiguous and its meaning and relevance are self-evident [41]. The reviewers were asked to examine every item of the checklist and indicate if the wording and meaning satisfy face validity. On the other hand, content validity consists of a judgement whether the checklist samples all the relevant or important content, *i.e.* the checklist has enough items and adequately covers the domain under investigation [41]. The first aspect of content validity was to make sure that every item of the developed checklist was essential to the domain [42]. Thus, the reviewers were first asked whether each item of the developed checklist was necessary in order to assess the risk of bias of studies included in a systematic review of drug adverse events. The second aspect of content validity implies that no important element was omitted in the developed tool. To be correctly assessed, this aspect necessitates a detailed description of the content domain [43]. Since no such description was available for the risk of bias of studies assessing drug adverse events for all selected study designs, we provided to the reviewers a comprehensive list of domains and elements believed to affect the risk of bias of studies in general (based on work by the Agency for Healthcare Research and Quality in 2002 [44]) and of pharmacoepidemiologic safety studies (based on work by Neyarapally *et al.* in 2012 [16]). For each element, the reviewers were asked to answer: (i) whether the item was relevant to the assessment of the risk of bias of studies included in a systematic review of drug adverse events, and (ii) if it was relevant, whether the item was addressed in the developed checklist. Finally, the reviewers were asked to mention whether they judged that other important elements should be added in the developed checklist.

### **2.2.2 Revised version of the checklist**

The checklist was revised according to the responses and commentaries of the expert reviewers. To resolve disagreements between two or three reviewers regarding face or content validity, a discussion between reviewers was planned until a consensus was reached.

### **2.2.3 Inter- and intra-rater reliability**

The instrument reliability measures whether a similar risk of bias assessment can be derived on the same study when the checklist is used either by different assessors (inter-rater reliability) or by the same assessor at different periods (intra-rater reliability).

Reliability was tested using random samples of four studies for each study design (20 studies in total: four RCTs, four cohort studies, four case-control studies, four nested case-control studies, and four systematic reviews) extracted from the PROTECT systematic reviews for key drug-adverse events including antiepileptic drugs and suicide, inhaled beta-2-agonists/muscarinic antagonists and acute myocardial infarction, antidepressants and hip fractures, benzodiazepines and hip fractures, macrolides and amoxicillin/clavulanate and liver injury and calcium channel blockers and risk of cancer. The list of the randomly selected articles is provided in Table 1. The criteria used to select drug-adverse events to be investigated by a systematic review in the PROTECT project have been described elsewhere [45]. Three independent reviewers with an advanced degree in epidemiology / clinical research assessed the 20 studies using the revised checklist. Although the risk of bias assessment is not strictly an ordinal scale, the responses (“Low”, “Unclear”, or “High”) can be interpreted as ordinal. Hence, Kendall's W coefficient of concordance for ordinal response was used as it can be used for the assessment of agreement between more than two raters [46]. Kendall's W coefficient ranges from 0 (no agreement) to 1 (complete agreement) and is provided with a statistical test of whether it is equal to zero. After a 3-week interval, one reviewer performed a second assessment of the 20 studies (*i.e.* a “retest”). Kendall's W coefficient was also used to measure agreement between test and retest (intra-rater reliability). Inter- and intra-rater reliability was also explored within the checklist domains and within the different study designs. According to Streiner *et al.* a sample of 20 studies appeared to be sufficient to measure acceptable reliability [41]. This was calculated based on an acceptable value of Cohen's kappa coefficient of 0.6 which is the threshold for “substantial” reliability according to Landis and Koch and “excellent” reliability according to Cicchetti and Sparrow [41, 47, 48]. As to evaluate the usability of the checklist, time to assess each study was measured for each rater and median times with quartiles were calculated.

### 3. Results

The pilot version of the checklist was revised according to the responses and commentaries of the expert reviewers assessing face and content validity. The term “study quality” has been replaced in the whole document by “risk of bias”; the terms “drug adverse event” and “drug safety outcome” has been chosen as synonymous generic terms and replaced the terms “adverse effect” or “adverse reaction”; an Introduction section has been added to precise the terminology and the instructions to use the checklist; several items or instructions were reformulated (B3, B5-6, C3, D1-2, D8, E2-3 to take into account drug use duration and recall bias, E6, F1 to precise exact, complete and over-matching, F5, G1-2) or moved from one domain to another (B8, C4); a domain previously named “Definition of adverse event” has been included in the domain relative to the information bias which has been separated into two domains (information bias regarding the harmful outcome and other information bias); the item relative to both allocation sequence and concealment has been separated into two items (B1 and B2); the item relative to blinding methods has been separated into three items concerning participants (D3), outcome assessors (D4) and care givers during follow-up (E1); 5 items have been added: B4 (“Were inclusion and exclusion criteria implemented uniformly across study groups?”), F4 about time-dependent confounders, G3 about adequate construction of composite outcomes, G6 about consistency of analyses, and H2 about conflicts of interest.. The revised checklist was then checked and all reviewers approved the final version as valid approach to measure the risk of bias. The final version of the checklist is composed of a list of items structured in eight domains (noted from A to H): A) Study design and objectives; B) Bias in selection of subjects and constitution of study groups; C) Bias due to withdrawals or loss of follow-up (attrition); D) Information bias regarding drug adverse events; E) Other information bias; F) Statistical methods to control confounding; G) Statistical methods excluding methods to control confounding; H) Conflicts of interest. Depending on the study design, the total number of items in the final version of the checklist is: 32 for RCTs, 32 for cohort studies, 24 for case-control studies, 25 for nested case-control studies, and 10 for systematic reviews. The final version of the checklist is available in supplementary content.

Among the selected studies used for reliability assessment, 9 (45%) were considered at low risk of bias, 9 (45%) at unclear risk and 2 (10%) at high risk by at least two reviewers. Inter- and intra-rater agreements were fair with Kendall’s W of 0.70 ( $p < 0.001$ ) and 0.74 ( $p < 0.001$ ), respectively. Inter- and intra-rater agreements according to risk of bias domains and study designs are presented in Table 2 and 3, respectively. The highest inter- and intra-rater agreements were found for systematic reviews (Kendall’s W 0.88 and 0.90, respectively), whereas lowest values were found for nested case-control studies (Kendall’s W 0.60 and 0.56, respectively). Median time to complete the checklist was 8.5 minutes (Q1-Q3: 6 min - 15 min) in general and, according to study type, 12.5 min, 13 min, 7.5 min, 6.75 min, and 4.5 min for

randomized controlled trials, cohort studies, case-control studies, nested case-control studies and systematic reviews, respectively.

## 4. Discussion

The present study shows that the developed checklist is a valid and acceptably reliable instrument to assess the risk of bias for studies analyzing adverse events of drugs. The developed checklist is the first to specifically address the issues of risk of bias assessment in the context of systematic reviews focusing on drug adverse events.

Existing risk of bias assessment tools are found either in form of scales or checklists with or without summary judgment. The choice of the checklist format with a summary judgment was based on recommendations [32], and limitations attributed to scales. In a checklist, the components are evaluated separately and do not have numerical scores attached to them, whereas, in a scale, each item is scored numerically and an overall quality score is generated [41]. In scales, high scores usually represent better quality, but no standardized scoring result is provided [49]. Using summary numerical scores to assess risk of bias has been considered hazardous [25]. In our tool, the format used for the responses to assess risk of bias is: “Low”, “Unclear” or “High”, for the different domains and for the general assessment. In the recent version of the Cochrane risk of bias tool for observational studies, a broader spectrum of responses are proposed (« Low », « Moderate », « Serious », « Critical » and « No information ») [11]. We did not consider that the difference with our 3-level assessment represents a major issue for the evaluation and the management of risk of bias in systematic reviews.

Regarding the item about comprehensive literature search for systematic reviews (B9), we did not consider other sources of information such as data from spontaneous reporting systems and data from regulatory agencies since the relevance of including such data in systematic reviews is controverted [14]. However, the corresponding item mentions that “Other sources (published and unpublished research) are considered” allowing the assessment of such data.

The validity of the checklist was investigated by assessing its face and content validity and its inter- and intra-rater reliability. It was not possible to measure the other parameters that are generally explored in validity studies of health measurements tools. Internal consistency, usually measured for quantitative or binary items (with Cronbach’s alpha and Kuder-Richardson formula 20, respectively) could not be examined because our items were neither quantitative nor all in a binary format. Criterion validity, defined as the extent to which the measurement correlates with an external criterion variable that can be measured objectively and independently (*i.e.* a “gold standard”), can neither be assessed since no acceptable gold standard was available for the risk of bias assessment for studies of drug adverse events. Construct validity would aim to insure that our construct (the specific risk of bias in assessing drug adverse events) is related to other indices

of that construct, or associated with related construct (such as the general assessment of risk of bias) [41]. It did not seem relevant to evaluate the construct validity by comparing the performance of our checklist with other general risk of bias tools since our aim was specifically to highlight the specificities of drug adverse event assessment that differ between our checklist and general risk of bias tools.

Reliability assessment was based on 20 studies extracted from previous systematic reviews that examined various drugs (CNS drugs, cardiovascular drugs, or antibiotics) and adverse events in various medical fields (psychiatric and cardiovascular effects, effects on bone, on liver, or on cancer risk). Although there were fewer studies at high risk of bias than studies at low or unclear risk of bias, this method allowed us to include studies that reflect the heterogeneous risk of bias in real systematic reviews of drug adverse events.

Although all agreement coefficients were statically significant, interpretation of the reliability assessment is limited by the value of coefficients of concordance which only correspond to “fair” agreements for inter- and intra-rater reliability. Reliability evaluation according to risk of bias domains and study designs mainly showed similar results than for the whole checklist. The lack of high agreement between raters may be due to differential practical experience in risk of bias assessment and uncertainties regarding some concepts used in the checklist. For example, concerning statistical methods to control confounding, the adequate methods are not clearly defined in the checklist and the rater’s interpretation may depend on his/her methodological background. Furthermore, a learning effect could affect the results as shown by the similar fair agreement between test and retest in the intra-rater reliability evaluation. Although reliability findings were acceptable, we recommend that the developed checklist be used by reviewers with a practical experience in risk of bias assessment and after having tested it in a few training studies. Despite the relatively high number of items (up to 32 for randomized controlled trials), time to complete the checklist by naïve reviewers was practicable (from around 5 min for systematic reviews to around 13 min to randomized controlled trials and cohort studies). Further researches that would implement the developed checklist are needed to better assess its feasibility in real settings.

In conclusion, the checklist presented in this manuscript is a comprehensive tool specifically designed to evaluate the risk of bias in the context of drug safety assessment. It might be considered as a novel useful tool for systematic reviews and meta-analyses focusing on adverse effects of drugs.

## **Competing interest statement**

The authors declare that they have no conflicts of interest.

## **Contributions**

JLF, PF, MS, EB, XV, MR, SS, and LI participated in the study design. JLF and PF wrote the initial version of the checklist. LI supervised the development and validity study of the checklist. AG, DD, SB, MJMZ, CH, and XC participated as expert reviewers for the validity study. JLF did the statistical analyses. JLF wrote the initial draft. PF, AG, DD, SB, MJMZ, CH, XC, SS, MS, EB and LI critically revised the manuscript.

## **Ethical considerations**

Since this study did neither include individual patients nor study individual patient data, ethical approval was not needed. In order to use the existing tools, permission from the original authors have been requested and granted by Professor Nick Black (Downs and Black scale) and Professor Julian Higgins (Cochrane risk of bias tool). No response was obtained from Doctor Beverley J Shea (AMSTAR scale) and Doctor Yoon Loke (reporting adverse events from randomized controlled trials of efficacy).

## **Funding**

The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsortTium, [www.imi-protect.eu](http://www.imi-protect.eu)) which is a public-private partnership coordinated by the European Medicines Agency. The PROTECT project has received support from the Innovative Medicine Initiative Joint Undertaking ([www.imi.europa.eu](http://www.imi.europa.eu)) under Grant Agreement n° 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. Dr. Maria José Martínez Zapata receives funds through a Miguel Servet research contract from the Instituto de Salud Carlos III (CP15/00116).

## References

- [1] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
- [2] Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv*. 1999;25:470-9.
- [3] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52:377-84.
- [4] Higgins JP, Altman DG, Sterne JAC. Assessing Risk of Bias in Included Studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: <http://handbook.cochrane.org/>.
- [5] Wells K, Littell JH. Study quality assessment in systematic reviews of research on intervention effects. *Res Soc Work Pract*. 2008;19:52–62.
- [6] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
- [7] Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1-12.
- [8] Wells G, Shea B, O'Connell J, Robertson J, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis. 2011. Available from: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.htm](http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm).
- [9] Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. *ANZ J Surg*. 2003;73:712-6.
- [10] Sterne JAC, Higgins JPT, Reeves BC on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014. Available from: <http://www.riskofbias.info>.
- [11] Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. 2016;355;i4919.
- [12] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009;62:1013-20.
- [13] Oxman AD. Checklists for review articles. *BMJ*. 1994;309:648-51.

- [14] Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol.* 2007;7:32.
- [15] Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med.* 2005;142:1090-9.
- [16] Neyarapally GA, Hammad TA, Pinheiro SP, Iyasu S. Review of quality assessment tools for the evaluation of pharmacoepidemiological safety studies. *BMJ Open.* 2012;2.
- [17] Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol.* 2010;63:502-12.
- [18] Cornelius VR, Perrio MJ, Shakir SA, Smith LA. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiol Drug Saf.* 2009;18:1223-31.
- [19] Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ.* 2014;348:f7668.
- [20] Tricco AC, Soobiah C, Berliner S, Ho JM, Ng CH, Ashoor HM, et al. Efficacy and safety of cognitive enhancers for patients with mild cognitive impairment: a systematic review and meta-analysis. *CMAJ.* 2013;185:1393-401.
- [21] Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. *CMAJ.* 2009;180:291-7.
- [22] Chou R, Fu R, Carson S, Saha S, Helfand M. Methodological shortcomings predicted lower harm estimates in one of two sets of studies of clinical interventions. *J Clin Epidemiol.* 2007;60:18-28.
- [23] Lawlor DA, Juni P, Ebrahim S, Egger M. Systematic review of the epidemiologic and trial evidence of an association between antidepressant medication and breast cancer. *J Clin Epidemiol.* 2003;56:155-63.
- [24] McDonagh M, Peterson K, Lee N, Thakurta S. Drug Class Review: Antiepileptic Drugs for Indications Other Than Epilepsy. Oregon Evidence-Based Practice Center. Portland, Oregon: Oregon Health and Science University; 2008. p. 117. Available from: <http://www.ohsu.edu/drugeffectiveness/reports/final.cfm>.
- [25] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *Jama.* 1999;282:1054-60.
- [26] Bilandzic A, Fitzpatrick T, Rosella L, Henry D. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med.* 2016;13:e1001987.
- [27] Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT). Available from: <http://www.imi-protect.eu/about.shtml>.



- [28] Ferrer P, Ballarin E, Sabate M, Vidal X, Rottenkolber M, Amelio J, et al. Antiepileptic drugs and suicide: a systematic review of adverse effects. *Neuroepidemiology*. 2014;42:107-20.
- [29] Ferrer P, Amelio J, Ballarin E, Sabate M, Vidal X, Rottenkolber M, et al. Systematic Review and Meta-Analysis: Macrolides- and Amoxicillin/Clavulanate-induced Acute Liver Injury. *Basic Clin Pharmacol Toxicol*. 2016;119:3-9.
- [30] Rottenkolber M, Rottenkolber D, Fischer R, Ibanez L, Fortuny J, Ballarin E, et al. Inhaled beta-2-agonists/muscarinic antagonists and acute myocardial infarction in COPD patients. *Respir Med*. 2014;108:1075-90.
- [31] Glossary of Terms in The Cochrane Collaboration Version 4.2.5. [updated May 2005]. The Cochrane Collaboration, 2005. Available from: [www.cochrane.org/sites/default/files/uploads/glossary.pdf](http://www.cochrane.org/sites/default/files/uploads/glossary.pdf).
- [32] Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, Santaguida PL, Shamliyan T, Singh K, Tsertsvadze A, Treadwell JR. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. March 2012. AHRQ Publication No. 12-EHC047-EF. Available from: [www.effectivehealthcare.ahrq.gov/](http://www.effectivehealthcare.ahrq.gov/).
- [33] Robinson KA, Chou R, Berkman ND, Newberry SJ, Fu R, Hartling L, et al. Integrating Bodies of Evidence: Existing Systematic Reviews and Primary Studies. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK216379/>.
- [34] Becker L, Oxman A. Chapter 22: Overviews of reviews. In: *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.10 (updated March 2011). Higgins JPT, Green S (editors); 2011. Available from: [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
- [35] Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. *PLoS One*. 2012;7:e49667.
- [36] Thomson D, Russell K, Becker L, Klassen T, Hartling L. The evolution of a new publication type: Steps and challenges of producing overviews of reviews. *Res Synth Methods*. 2010;1:198-211.
- [37] Hartling L, Vandermeer B, Fernandes RM. Systematic reviews, overviews of reviews and comparative effectiveness reviews: a discussion of approaches to knowledge synthesis. *Evid Based Child Health*. 2014;9:486-94.
- [38] Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol*. 2012;65:1267-73.
- [39] Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36:666-76.

- [40] Loke Y, Price D, Herxheimer A. Adverse effects. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: <http://handbook.cochrane.org/>.
- [41] Streiner DL, Norman CR. Health measurement scales: a practical guide to their development and use. Fifth edition. New York, Oxford University Press, 2015.
- [42] Lawshe CH. A quantitative approach to content validity. *Personnel Psychology* 1975;28:563-75.
- [43] Crocker L, Algina J. Introduction to Classical & Modern Test Theory. Fort Worth: Harcourt Brace Jovanovich College Publishers, 1986.
- [44] West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux K. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment No. 47 AHRQ Publication No. 02-E016. Rockville, MD. Agency for Healthcare Research and Quality, 2002. Available from: <http://archive.ahrq.gov/clinic/epcsums/strengthsum.htm>.
- [45] Abbing-Karahagopian V, Kurz X, de Vries F, van Staa T, Alvarez Y, Hesse U, et al. Bridging differences in outcomes of pharmacoepidemiological studies: designs and first results of the PROTECT Project. *Curr Clin Pharmacol* 2014; 9:130-138.
- [46] Kendall MG, Babington Smith B. The Problem of m Rankings. *Annals of Mathematical Statistics* 1939;10:275-87.
- [47] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- [48] Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86:127-37.
- [49] Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *J Clin Epidemiol.* 2011;64:79-89.

## Tables

**Table 1. Selected articles for the reliability assessment**

Study design	Citations
Randomized control trials	Tashkin DP, Celli B, Senn S, Burkhart D, Kesten S, Menjoge S, <i>et al.</i> A 4-year trial of tiotropium in chronic obstructive pulmonary disease. <i>N Engl J Med</i> 2008;359:1543-1554.
	Volgelmeier C, Hederer B, Glaab T, Schmidt H, Rutten-van Molken RM, Beeh KM, <i>et al.</i> Tiotropium versus salmeterol for the prevention of exacerbations of COPD. <i>N Engl J Med</i> 2011;364:1093-1103
	Kamada AK, Hill MR, Iklé DN, Brenner AM, Szeffler SJ. Efficacy and safety of low-dose troleandomycin therapy in children with severe, steroid-requiring asthma. <i>J Allergy Clin Immunol</i> 1993;91:873-82.
	Johnston SL, Blasi F, Black PN, Martin RJ, Farrell DJ, Nieman RB. The effect of telithromycin in acute exacerbations of asthma. <i>N Engl J Med.</i> 2006;354:1589-600.
Cohort studies	Andrade-Machado R, Espinosa AG, Melendrez D, González YR, García VF, Rodríguez YQ. Suicidal risk and suicide attempts in people treated with antiepileptic drugs for epilepsy. <i>Seizure</i> 2011; 20:280-284
	Sondergard L, López AG, Andersen PK, Kssing LV. Mood-stabilizing pharmacological treatment in bipolar disorders and risk of suicide. <i>Bipolar Disord</i> 2008;10:87-94.
	Ensrud KE, Blackwell T, Mangine CM, Bowman PJ, Bauer DC, Schwartz A, <i>et al.</i> Central nervous system active medications and risk for fractures in older women. <i>Arch Intern Med</i> 2003;163:949-957.
	Cohen HJ, Pieper CF, Hanlon JT, Wall WE, Burchett BM, Havilik RJ. Calcium channel blockers and cancer. <i>Am J Med</i> 2000;108:210-215.
Case-control studies	Ray WA, Griffin MR, Schaffner W, Baugh DK, Melton LJ 3rd. Psychotropic drug use and the risk of hip fracture. <i>N Engl J Med</i> 1987;316:363-369.
	Lichtenstein MJ, Griffin MR, Cornell JE, Malcolm E, Ray WA. Risk factors for hip fractures occurring in the hospital. <i>Am J Epidemiol</i> 1994;140:830-838
	Carson JL, Strom BL, Duff A, Gupta A, Das K, Shaw M, <i>et al.</i> Acute Liver Disease Associated with Erythromycins, Sulfonamides, and Tetracyclines. <i>Ann Intern Med.</i> 1993;119:576-83.
	Perron L, Bairati I, Harel F, Meyer F. Antihypertensive drug use and the risk of prostate cancer (Canada). <i>Cancer Causes Control</i> 2004;15:535-541
Nested case-control studies	Chang CM, Wu EC, Chang IS, Lin KM. Benzodiazepine and risk of hip fractures in older people: a nested case-control study in Taiwan. <i>Am J Geriatr Psychiatry</i> 2008;16:686-692
	Kaye JA, Castellsague J, Bui CL, Calingaert B, McQuay LJ, Riera-Guardia N, <i>et al.</i> Risk of Acute Liver Injury Associated with the Use of Moxifloxacin and Other Oral Antimicrobials: A Retrospective, Population-Based Cohort Study. <i>Pharmacotherapy</i> 2014;34:336-49.
	Assimes TL, Elstein E, Langleben A, Suissa S. Long-term use of antihypertensive drugs and risk of cancer. <i>Pharmacoepidemiol Drug Saf</i> 2008;17:1039-1049.
	Jick H, Jick S, Derby LE, Vasilakis C, Wald Myers M, Meier CR. Calcium-channel blockers and risk of cancer. <i>The Lancet</i> 1997;349:525-528.
Systematic reviews and meta-analyses	Van Lieshout RJ, Mac Queen GM. Efficacy and acceptability of mood stabilizers in the treatment of acute bipolar depression: systematic review. <i>Br J Psychiatry</i> 2010;196:266-273
	Premkumar TS, Pcik J. Lamotrigine for schizophrenia. <i>Cochrane Database Syst Rev</i> 2006;4: CD005962
	Barr RG, Bourbeau J, Camargo CA, Ram FS. Tiotropium for stable chronic obstructive pulmonary disease: a meta-analysis. <i>Thorax</i> 2006;61:854-862
	Celli B, Decramer M, Leimer I, Vogel U, Kesten S, Tashkin DP. Cardiovascular safety of tiotropium in patients with COPD. <i>Chest</i> 2010;137:20-30

**Table 2. Inter- and intra-rater agreements according to the eight risk of bias domains.**

<b>Domain</b>	<b>Kendall's W coefficient for inter-rater reliability*</b>	<b>Kendall's W coefficient for intra-rater reliability*</b>
A) Study design and objectives	0.75	0.88
B) Bias in selection of subjects and constitution of study groups	0.64	0.64
C) Bias due to withdrawals or loss of follow-up (attrition)	0.54	0.76
D) Information bias regarding drug adverse events	0.77	0.76
E) Other information bias	0.68	0.62
F) Statistical methods to control confounding	0.57	0.57
G) Statistical methods excluding methods to control confounding	0.63	0.75
H) Conflicts of interest	0.83	0.86

\* p<0.001 for all Kendall's coefficients

**Table 3. Inter- and intra-rater agreements according to study designs.**

<b>Study design</b>	<b>Kendall's W coefficient for inter-rater reliability*</b>	<b>Kendall's W coefficient for intra-rater reliability*</b>
Randomized controlled trial	0.79	0.80
Cohort study	0.63	0.73
Case-control study	0.65	0.68
Nested case-control study	0.60	0.56
Systematic review	0.88	0.90

\* p<0.001 for all Kendall's coefficients

## **RISK OF BIAS ASSESSMENT CHECKLIST FOR STUDIES INCLUDED IN SYSTEMATIC REVIEWS OF DRUG ADVERSE EVENTS**

### **I. Introduction**

#### **Terminology**

This checklist is an instrument for rating the risk of bias of studies included in systematic reviews of drug adverse events. We use the term “risk of bias” as a synonymous of internal validity *i.e.* “the extent to which the design and conduct of a study are likely to have prevented bias”.<sup>1</sup> Bias is defined as “a systematic error or deviation from the truth, in results or inferences.”<sup>2</sup> As it is recommended by the *Agency for Healthcare Research and Quality*,<sup>3</sup> this instrument does neither take precision/statistical power nor applicability/generalizability/external validity of the results into consideration for the risk of bias assessment of studies.

In this instrument, we use the term “drug adverse event” or “drug safety outcome” as generic terms to define undesirable outcomes that occur during or after the use of a drug whether the causal relationship with the drug has been established or not.

This instrument is aimed to assess the risk of bias of different types of study: randomized controlled trial (RCT), cohort and case-control studies, and systematic reviews. We include systematic reviews as they can be part of the studies used to perform a new systematic review or update an existing one. The type of the study concerned by each question is specified in each question of the checklist.

- “All studies” refers to all the study types: RCT, cohort and case-control studies, and systematic reviews.
- “RCT” refers to all the possible designs of randomized controlled trial.
- “Systematic reviews +/- meta-analyses” only refers to systematic review with or without meta-analyses
- “Secondary database studies” only refers to observational studies (cohort and case-control studies) using administrative and clinical databases containing secondary data.
- “Cohort” and “case control” studies refer to all the cohort and case control designs whether they are classical (“population”) studies or studies using a secondary database.
- Note that a nested case-control performed in a medico-administrative database will be concerned by the item for case-control studies, cohort and secondary database studies.
- N/A stands for “not applicable”.

---

<sup>1</sup> Cochrane Collaboration Glossary Version 4.2.5. 2005. Available at: [www.cochrane.org/sites/default/files/uploads/glossary.pdf](http://www.cochrane.org/sites/default/files/uploads/glossary.pdf); <http://effectivehealthcare.ahrq.gov/>

<sup>2</sup> Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. In: Higgins JPT, Green S, eds. The Cochrane Collaboration; 2011.

<sup>3</sup> Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, Santaguida PL, Shamliyan T, Singh K, Tsertsvadze A, Treadwell JR. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. March 2012. AHRQ Publication No. 12-EHC047-EF. Available at: [www.effectivehealthcare.ahrq.gov/](http://www.effectivehealthcare.ahrq.gov/)

## Instructions

This checklist is composed of a structured list of questions. Depending on the question, answers can include “Not applicable”, “Yes”, “Unclear” or “No”.

Each question indicates which type of studies is concerned. The number of questions depends on the study type: 32 for RCTs, 32 for cohort studies, 24 for case-control studies, 25 for nested case-control studies, and 10 for systematic reviews.

The questions are structured in 8 risk-of-bias domains (noted from A to H). For each domain, a summary risk of bias assessment is asked after the questions related to the domain. At the end of the checklist, a final summary risk of bias assessment is asked for the whole study (see table below for the definitions of the risk of bias categories).

**Summary assessment outcomes across domains and across the study** (according to the Cochrane risk of bias tool).

<b>Risk of bias</b>	<b>Within a domain</b>	<b>Within the study</b>
Low risk of bias	Plausible bias unlikely to seriously alter the results	Low risk of bias for all key domains
Unclear risk of bias	Plausible bias that raises some doubts about the results or when information on which to base risk of bias judgments is missing or poorly reported.	Unclear risk of bias for one or more key domain
High risk of bias	Plausible bias that seriously weakens confidence in the results	High risk of bias for one or more key domains

## II. RISK OF BIAS ASSESSMENT CHECKLIST FOR STUDIES INCLUDED IN SYSTEMATIC REVIEWS OF DRUG ADVERSE EVENTS

### A. STUDY DESIGN AND OBJECTIVES

**A1. All studies: Are study objectives clearly specified and appropriate?**

Yes

Study objectives clearly specified and appropriate.

No

Study objectives not clearly specified or not appropriate.

**A2. All studies: Is study design clearly specified and appropriate?**

Yes

Study design clearly specified and appropriate.

No

Study design not clearly specified or not appropriate.

**A3. RCT: Is the study design free of run-in/lead-in period before inclusion/randomization of participants?**

Yes

No run-in/lead-in period

Unclear

Not clear information

No

Presence of a run-in/lead-in period

**A4. RCT (cross-over designs): Is the study designed to adequately address carry-over effect?**

N/A

Yes

Carry-over effect absent or adequately addressed (randomized order and sufficiently long wash-out period).

Unclear

Not clear information

No

Carry-over effect not adequately addressed or susceptible to bias the results.

### **RISK OF BIAS ASSESSMENT FOR STUDY DESIGN AND OBJECTIVES**

Low

Unclear

High

Comments:



## B. BIAS IN SELECTION OF SUBJECTS AND CONSTITUTION OF STUDY GROUPS

**B1. RCT: Was the method used to generate the allocation sequence adequate as to produce comparable groups?**

Yes

Allocation methods are adequate to produce comparable groups.

Unclear

Allocation methods are not clearly reported.

No

Allocation methods are not adequate (e.g. assignment to treatment by birth date, week day, etc.), groups are not comparable.

**B2. RCT: Was the method used to conceal the allocation sequence adequate as to produce comparable groups?**

Yes

Concealment is adequate.

Unclear

Concealment methods are not clearly reported and groups may not be comparable.

No

Concealment methods are not adequate, groups are not comparable.

**B3. RCT, Cohort and case-control studies: Are all the subjects recruited from the same source population?**

Yes

All the subjects recruited from the same source population.

Unclear

Unclear if all the subjects recruited from the same source population.

No

Not all the subjects are recruited from the same source population

**B4. RCT, Cohort and case-control studies: Were inclusion and exclusion criteria implemented uniformly across study groups?**

Yes

Selection criteria uniformly implemented.

Unclear

Unclear if selection criteria are uniformly implemented.

No

Selection criteria not uniformly implemented.

**B5. Case-control studies: Is the origin of controls clearly specified?**

Yes

Origin of controls clearly specified.

No

Origin of controls not clearly specified.

**B6. Secondary databases studies: Are the characteristics of the database clearly described?**

Yes

The authors describe the database or explain (or give a reference to explain) why the database was initially created

No

Information about the database is missing or not clear.

**B7. Cohort studies: Are exposed subjects new users of the drug?**

Yes

A new-user design is used and the risk of prevalent users is low

Unclear

It is not clear whether exposed subjects are new users or prevalent users

No

A new-user design is not used or the risk of prevalent users is substantial

**B8. RCT, cohort studies: Are baseline characteristics and prognostic factors comparable between different groups?**

Yes

Unclear

No

**RCT:** Groups are comparable at baseline.

No description of baseline characteristics or only significance tests.

The groups are unbalanced at baseline.

**Cohort studies:** Groups are comparable at baseline or matched for the main prognostic factors

**B9. Systematic review +/- meta-analyses: Was a comprehensive literature search conducted and were studies adequately selected?**

Yes

Unclear

No

Comprehensive literature search:  $\geq 2$  electronic databases, without dates or language restriction. Search terms (MeSH and free text words) and strategy are reported. Other sources (published and unpublished research) are considered. Inclusion and exclusion criteria are specified.

The search process is partly reported, or any of the steps in conducting the search of articles is not reported.

No comprehensive literature search: Only one electronic bibliographic database or English language restriction (without adequate justification) or inadequate dates. Poor search strategy: one single word as search term (no synonyms or MeSH terms included).

**RISK OF BIAS ASSESSMENT FOR SELECTION OF PARTICIPANTS AND CONSTITUTION OF STUDY GROUPS**

Low

Unclear

High

Comments:

### C. BIAS DUE TO WITHDRAWALS OR LOSS OF FOLLOW-UP (ATTRITION)

**C1. RCT, cohort studies: Are the number of participants clearly reported throughout the study?**

Yes

Numbers of participants throughout the study are reported. Complete flow chart.

No

Numbers of patients at every stage is not clearly reported. Confusing information is reported regarding the number of participants. No or incomplete flow chart.

**C2. RCT: Is the number of drop-outs/withdrawals due to drug safety outcome clearly stated for each treatment arm?**

Yes

The number of drop-outs due to drug safety outcome is specified.

No

The number of drop-outs due to drug safety outcome is not specified, unclear or combined.

**C3. RCT, cohort studies: Does the study adequately address biased loss to follow-up?**

Yes

Complete follow-up or drop-outs unlikely to introduce bias or adequately controlled.

Unclear

Drop-outs/withdraws due to drug safety outcome are not clearly reported.

No

Loss to follow-up affects the safety outcome and is not adequately controlled.

**C4. RCT: Are the results based on an intention-to-treat analysis?**

Yes

Results are based on a strict intention-to-treat analysis.

Unclear

Not clear if an intention-to-treat analysis is performed.

No strict intention-to-treat analysis.

No

Results are not based on intention-to-treat analysis (not done or not possible).

**RISK OF BIAS DUE TO WITHDRAWALS OR LOSS OF FOLLOW-UP (ATTRITION)**

Low

Unclear

High

Comments:

## D. INFORMATION BIAS REGARDING THE DRUG SAFETY OUTCOME

### D1. All studies: Is the definition of the drug safety outcome clearly stated?

Yes

**RCT, cohort studies:** clear / standardized definition of the drug safety outcome (e.g. diagnostic codes, clinical and laboratory data).

**Case-control studies:** clear definition of cases.

No

Definition of the drug safety outcome not reported or that leads to confusion. Terms not well-constructed, wrong definition.

### D2. All studies: If applicable, is the severity of the drug safety outcome clearly stated?

N/A

Yes

Self evident severity (e.g. death) Detailed degree of severity or reference to a known scale of severity or a new scale developed for the study.

No

Unclear degrees of severity or without clear boundaries between them.

### D3. RCT: Was the blinding methods of participants regarding the intervention appropriate considering the nature of the drug safety outcome?

Yes

Blinding ensured (and unlikely broken) or outcome not likely to be influenced by lack of blinding.

Unclear

There is no sufficient information regarding the process of blinding or the outcome assessment.

No

No blinding (or incomplete blinding or risk of broken blinding) and outcome likely to be influenced by lack of blinding.

### D4. RCT, cohort, case-control studies: Was the blinding method of drug safety outcome assessment appropriate considering the nature of the adverse event?

Yes

Blinding ensured (and unlikely broken) or outcome assessment not likely to be influenced by lack of blinding.

Unclear

There is no sufficient information regarding the process of blinding of outcome assessment.

No

No blinding (or incomplete blinding or risk of broken blinding) and outcome likely to be influenced by lack of blinding.

**D5. RCT, cohort studies: Was the duration of follow-up adequate to assess the drug safety outcome?**

Yes

Sufficient duration of follow-up to assess the outcome

Unclear

It is unclear whether the duration of follow-up is adequate

No

Too short duration of follow-up

**D6. RCT, cohort, case-control studies: Was the method for ascertaining the drug safety outcome adequately constructed and equal for all participants?**

Yes

Adequate or validated methods of outcome measurement for all participants.

Clinical reactions medically confirmed by a physician. Minimized risk of misclassification or differential assessment, reporting or detection.

**RCT:** Active drug safety outcome surveillance (prospective/retrospective case-record review, questionnaires, patient's diary/checklist...).

Unclear

There is no or not sufficient information to clearly determine how information on drug safety outcome is collected or the process of minimizing misclassification.

No

Substantial risk of misclassification of outcome or differential assessment, reporting or detection.

Clinical reactions not medically confirmed.

**RCT:** Passive drug safety outcome surveillance (patient's volunteer reporting).

**D7. RCT: Are the number of drug adverse events and the number of patients with a drug adverse event reported in both treatment arms?**

Yes

Numbers are reported. It is possible to calculate the rates of drug adverse events.

Unclear

Confusion between the number of drug adverse events or the number of patients with a drug adverse event, or general statements such as "*5% of patients developed a drug adverse event*".

No

Neither the number of drug adverse events nor the number of patients with a drug adverse event is reported. Or numbers are combining both treatment arms.

**D8. RCT, cohort studies: Is the time frequency of drug safety outcome assessment during the follow-up period appropriate?**

Yes

For all study groups, the time frequency at which the drug safety outcome is assessed is appropriate.

Unclear

General statements such as "*patients were routinely assessed for drug safety outcomes*".

No

There is no regular collection of data on drug safety outcomes during the study.

**D9. RCT, cohort: Was the time between the exposure to a drug and the onset of the adverse event reported?**

Yes

The time between the drug exposure to the onset of adverse event is specified.

Unclear

The authors do not report a clear time frame between drug exposure and adverse event.

No

The authors do not report the time between the drug exposure to the onset of adverse event.



**D10. RCT: Was the process of determining that the adverse event is linked to the drug appropriate? Was the process blinded to the assigned treatment?**

Yes

Methods for causality assessment are appropriate and, if applicable, made by investigators blinded to the intervention.

Unclear

Unclear how the causality attribution is made. It is not clear who make the assessment or whether it is blinded to the assigned treatment.

No

Causality assessment is made by investigators not blinded to the intervention, or by participants or sponsors, or unblinding of treatment assignment precedes the decision to withdraw.

**RISK OF BIAS ASSESSMENT FOR INFORMATION BIAS REGARDING THE DRUG SAFETY OUTCOME**

Low

Unclear

High

Comments:

## E. OTHER INFORMATION BIAS

**E1. RCT: Is blinding of healthcare staff during follow-up adequately performed in order to avoid differential care between study groups (performance bias)?**

Yes

Unclear

No

There is no risk of differential care or it is adequately addressed.

Unclear risk of bias due to differential care

The bias due to differential care is not controlled

**E2. Cohort, case-control studies: Does drug exposure assessment appear free of time-related bias?**

Yes

Unclear

No

**Case-control studies:** similar length of time to measure exposure for cases and controls (no time-window bias).

Not clear information to determine the presence of time-related bias or error in drug use duration measurement

Substantial risk of time-related bias or error in drug use duration measurement.

**Cohort studies:** if present, the period from cohort entry to the first prescription is considered unexposed (no immortal time).

**E3. Cohort, case-control studies: Was the method for ascertaining drug use and drug use duration adequately constructed, and equal for all participants?**

Yes

Unclear

No

Adequate or validated methods of drug use and drug use duration measurement for all participants.

No sufficient information is provided regarding the process of minimizing misclassification.

Substantial risk of misclassification of exposure or differential assessment

**Case-control studies:** minimized risk of recall bias

**E4. Cohort, case-control studies: Was the method for ascertaining confounders adequately constructed, and equal for all participants?**

Yes

Unclear

No

Adequate or validated methods of confounders measurement for all participants.

No sufficient information is provided regarding the process of minimizing misclassification.

Substantial risk of misclassification of confounders or differential assessment.

Minimized risk of misclassification of confounders.

**E5. RCT, cohort, case-control studies: Does the study appear free of other information bias (competing risk/informative censoring, immeasurable time, reverse causality/protopathic bias, etc.)?**

Yes

Unclear

No

The study appears to be free of other information bias

Unclear presence of other information bias

Additional source of other information bias

**E6. Systematic review +/- meta-analyses: Is data extraction from included studies to perform the meta-analysis objective and reliable?**

Yes

Unclear

No

Protocol developed with standard data extraction form. At least 2 independent reviewers and a method to resolve disagreements.

Insufficient information to know who extracted the information or how disagreements were resolved.

No protocol developed, or only one reviewer retrieved the information from the articles included.

**RISK OF BIAS ASSESSMENT FOR OTHER INFORMATION BIAS**

Low

Unclear

High

Comments (other biases):

## F. STATISTICAL METHODS TO CONTROL CONFOUNDING

**F1. Matched case-control studies: Were cases and controls adequately matched for potential confounding factors?**

<input type="checkbox"/> Not applicable	<input type="checkbox"/> Yes	<input type="checkbox"/> Unclear	<input type="checkbox"/> No
Complete and exact matching is performed or nearly reached. Overmatching is avoided.	It is unclear whether cases and controls are adequately matched.	Cases and controls are not or inadequately matched.	

**F2. Cohort, case-control studies: Does the study adequately address confounding by indication or by severity of the disease, channeling or healthy user bias?**

<input type="checkbox"/> Yes	<input type="checkbox"/> Unclear	<input type="checkbox"/> No
Subjects do not use the study drugs according to differential risks of drug safety outcome. If they do, adequate methods are used to control this bias (disease risk scores, propensity scores, instrumental variables, etc.)	Unclear risk of confounding by indication or by severity of the disease, channeling or healthy user bias.	Subjects use the study drugs according to differential risks of drug safety outcome and adequate methods are not used to control this bias.

**F3. Cohort, case-control studies: Does the analysis adequately adjust for identified confounding factors?**

<input type="checkbox"/> Yes	<input type="checkbox"/> Unclear	<input type="checkbox"/> No
The effects of the main confounders are demonstrated and adjustment for these confounders is made in the final analysis.	It is not clear if confounders are included in the final analysis.	The effect of the main confounders was not investigated or confounding was demonstrated but no adjustment was made in the final analyses.

**F4. Cohort, case-control studies: Does the analysis address time-dependent confounders?**

<input type="checkbox"/> Yes	<input type="checkbox"/> Unclear	<input type="checkbox"/> No
Time-dependent confounders are not present or adequately adjusted.	It is not clear if time-dependent confounders are present or adequately adjusted.	Time-dependent confounders are present and not adequately adjusted.

**F5. RCT, cohort, case-control studies: Does the study adequately address residual or unmeasured confounding?**

Yes

Unclear

No

The study adequately addresses residual or unmeasured confounding.

Unclear presence of residual or unmeasured confounding.

Residual or unmeasured confounding is likely to be important.

**RISK OF BIAS ASSESSMENT FOR STATISTICAL METHODS TO CONTROL CONFOUNDING:**

Low

Unclear

High

Comments:

**G. STATISTICAL METHODS EXCLUDING METHODS TO CONTROL CONFOUNDING**

**G1. All studies: Are the statistical methods used to analyze the drug safety outcome appropriate?**

Yes

Statistical techniques are appropriate to the data. If the distribution of the data (normal or not) is not described, it must be assumed that the estimates used were appropriate.

*Meta-analysis*: Homogeneity is stated. If heterogeneity exists, appropriate statistical methods are used and/or clinical appropriateness of combining the studies should be considered.

Unclear

There is no description of the statistical techniques used, or the description is vague and not understandable.

No

The statistical techniques used are not appropriate.

**G2. RCT and cohort studies: Is a survival analysis performed when there are individual differences in length of follow-up?**

Yes

Follow-up is the same for all study patients, if not survival analysis is performed.

Unclear

Unclear whether there are different lengths of follow-up or whether they are taken into account.

No

Differences of follow up were ignored.

**G3. RCT and cohort studies: If applicable, is composite outcome of drug safety adequately constructed?**

N/A

Yes

Composite outcome appropriate.

Unclear

Unclear whether composite outcome is appropriate.

No

Construction of composite not described or not appropriate.

**G4. Cohort, case-control studies: Is drug exposure analyzed as a time-dependent variable?**

Yes

Appropriate methods have been used to take into account changes of exposure over time (time-dependent Cox model, nested case control analysis...).

Unclear

Not sufficient information on whether time-dependant variables are considered in the final analytic model.

No

Time-dependent variables have not been considered.

**G5. Cohort, case-control studies: Do sensitivity analyses account for different exposure windows, induction/lag periods?**

Yes

Sensitivity analyses are conducted regarding different window exposures (definitions of drug use duration, latency...) and induction or lag periods preceding the drug safety outcome.

Unclear

There is not sufficient information on sensitivity analyses.

No

No sensitivity analyses have been conducted for different time-windows of exposure and induction time.

**G6. RCT, cohort, case control studies: Are the results consistent in primary and secondary analyses? Are confounding effects consistent with known associations?**

Yes

Unclear

No

Consistency of primary, secondary analyses and consistency of confounding effects with known associations

Not sufficient information to determine consistency.

No consistency of primary, secondary analyses or no consistency of confounding effects with known associations

**G7. Systematic review +/- meta-analyses: Is there a clear flow chart of the studies?**

Yes

No

Complete flow chart.

No or incomplete flow chart.

**G8. Systematic review +/- meta-analyses: Is publication bias assessed?**

Yes

Unclear

No

Assessment of publication bias includes a combination of graphical aids (e.g. funnel plot, or other available tests) and/or statistical tests (e.g. Egger regression test).

Not sufficient information to determine whether the authors assessed publication bias or not.

Publication bias was not assessed.

**G9. Systematic review +/- meta-analyses: Is the quality of studies being assessed and taken into account in the statistical analyses?**

Yes

Unclear

No

Quality of the primary studies is checked and considered either in a narrative synthesis or incorporated in statistical tests. A priori methods of assessment are provided

The authors do not provide sufficient information on how the quality of studies was assessed.

Quality of studies is not assessed or the items assessed are not appropriate. The results of the quality assessment are neither considered in a narrative synthesis nor in the statistical analysis.

**RISK OF BIAS ASSESSMENT FOR STATISTICAL METHODS EXCLUDING METHODS TO CONTROL CONFOUNDING:**

Low

Unclear

High

Comments:



## H. CONFLICT OF INTEREST

**H1. All studies: Were the conflict of interest or sources of funding clearly acknowledged?**

Yes

Potential sources of support are acknowledged.

No

No sources of funding reported or not sufficient information

**H2. All studies: Does the study appear free of conflicts of interest susceptible to have influenced design, analysis or reporting (selective reporting of outcome or analysis)?**

Yes

No conflicts of interest or not susceptible to have influenced design, analysis or reporting

Unclear

It is unclear if there are conflicts of interest or if they are susceptible to have influenced design, analysis or reporting.

No

Conflicts of interest susceptible to have influenced design, analysis or reporting.

**RISK OF BIAS ASSESSMENT FOR CONFLICT OF INTEREST**

Low

Unclear

High

Comments:

**SUMMARY RISK-OF-BIAS ASSESSMENT FOR THE STUDY**

Low risk of bias

Unclear risk of bias

High risk of bias

Comments: