


COMMENTARY

Open Access



Conducting separate reviews of benefits and harms could improve systematic reviews and meta-analyses

Evan Mayo-Wilson^{1*} , Riaz Qureshi² and Tianjing Li²

Abstract

Guidance for systematic reviews of interventions recommends both benefits and harms be included. Systematic reviews may reach conclusions about harms (or lack of harms) that are not true when reviews include only some relevant studies, rely on incomplete data from eligible studies, use inappropriate methods for synthesizing data, and report results selectively. Separate reviews about harms could address some of these problems, and we argue that conducting separate reviews of harms is a feasible alternative to current standards and practices. Systematic reviews of potential benefits could be organized around the use of interventions for specific health problems. Systematic reviews of potential harms could be broader, including more diverse study designs and including all people at risk of harms (who might use the same intervention to treat different health problems). Multiple reviews about benefits could refer to a single review of harms. This approach could improve the reliability, completeness, and efficiency of systematic reviews.

Commentary

Because potential benefits and harms are important to patients and providers, guidance recommends that both types of outcomes be included in systematic reviews of interventions [1–6]. In Cochrane systematic reviews of interventions, “considering potential adverse effects” is mandatory [7]. Yet, highly cited reporting guidelines include few recommendations for reporting harms [8, 9], which are addressed in a lesser-known extension [10], and most systematic reviews report harms poorly [11, 12]. We argue that guidance for systematic reviews may overlook important differences between benefits and harms, resulting in the production of incomplete reviews

and unreliable conclusions about harms. In many cases, separate systematic reviews would be preferable to combined reviews of both benefits and harms.

Harms may be described using many different terms with different and overlapping meanings, such as “adverse events” and “side effects.” [13] Information about harms might include different dimensions such as timing, duration, and severity. Here, we use “harms” broadly to refer to outcomes that negatively affect individuals receiving interventions, which may be assessed in randomized trials and other studies to investigate whether harms are causally related to interventions.

Systematic reviews including both benefits and harms can be done rigorously. For example, one review about spinal fusion included 17 randomized trials and 35 other studies [14]; using individual patient data meta-analysis, the reviewers found important differences between published and unpublished data about harms [15]. Such rigorous reviews are noteworthy because they include data that are difficult to access, apply expert knowledge of harms ascertainment, use appropriate statistical

*Correspondence:

Evan Mayo-Wilson
evan.mayo-wilson@unc.edu

¹ Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC 27599, USA

² Department of Ophthalmology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

methods, and require considerable investigator time and funding.

Challenges in systematic reviews of harms arise from methods used to assess harms in primary studies [6, 16]. For example, some harms can be assessed systematically like benefits (e.g., in the same way for all participants in a trial) [17]; however, many harms are assessed non-systematically in response to open-ended questions [18] or using unclear methods [19]. Because dozens or hundreds of different non-systematic harms might be identified in a single study, non-systematic harms are often reported based on study results using post hoc criteria. For example, authors might report harms that occurred in 5% or more of patients, or they might report harms occurring twice as frequently in the intervention group compared with the comparison group [20–22]. Such criteria have at least two implications for systematic reviews and meta-analyses. First, data for uncommon and rare events may be unavailable for synthesis. Second, syntheses may be biased for harms because estimates from primary studies will be included or excluded based on observed results [23, 24].

Systematic reviews of harms can compound problems arising from primary studies [25, 26]. Because many non-systematic harms might be identified in individual studies, different non-systematic harms might be identified across studies included in a systematic review. Consequently, reviewers apply selection criteria when choosing which harms to extract, synthesize, and report. For example, reviewers might choose to analyze the 10 most common harms, or reviewers might analyze harms reported in at least half of the included studies. Moreover, it is often unclear what selection criteria were applied in systematic reviews; for example, reviews often fail to define thresholds for the “most common” harms [22, 27]. Combined with trial-level selection criteria, review-level selection criteria increase the likelihood that systematic reviews will omit some harms, and that estimates for other harms will be biased. Moreover, inconsistent review-level selection criteria contribute to conflicting conclusions across reviews of the same interventions [27, 28].

Even for included studies, systematic reviews rarely use complete data about harms. Because published reports typically include less information about harms compared with benefits, unpublished reports and databases are preferred sources for information about harms in primary studies [29–31]. For example, journal articles about randomized trials typically include quantitative information about potential benefits in structured text, tables, and figures (e.g., primary and secondary outcomes). Potential benefits (“efficacy outcomes”) are defined before analysis begins, and these are often published in protocols and

trial registrations. By contrast, reviewers are advised to look for other data sources about harms such as clinical study reports (CSRs), case report forms, trial registers, and individual participant data (IPD) [31–39]. Locating and reviewing multiple data sources requires more time and resources than reviewing journal articles alone [40, 41], and sometimes, trial investigators do not provide requested data, so many reviews are limited to published reports [42–44]. This is an important limitation because conclusions can change when different data sources are used, as in reviews examining the relative harms and benefits of antidepressants for young people [45, 46].

Suboptimal methods for extracting and synthesizing data also lead to incorrect results. Data extraction errors are common in meta-analyses of harms [47], especially when harms need to be coded and standardized across studies before synthesis, as is the case when primary studies do not report harms using common classification systems. For example, the Medical Dictionary for Regulatory Activities (MedDRA), Common Terminology Criteria for Adverse Events (CTCAE), and Systematized Nomenclature of Medicine Clinical Terms (SNOMED) are hierarchical systems used for recording and reporting harms in different areas, but they are not used in all primary studies or reviews [28]. Additionally, many systematic reviews do not describe plans for handling rare events in their methods [48], and many reviews deal with missing information inappropriately in their results [49]. For example, it would be informative if 80% of trials in a review failed to report the proportion of people experiencing a given harm; however, many systematic reviews omit studies with zero events from meta-analyses [50]. Common methods to handle rare events, and ignoring zero events, can lead to biased results [51–53].

Systematic reviews may come to incorrect conclusions about harms when they exclude relevant study designs. For example, systematic reviews often include only randomized trials; many reviews are designed to assess potential benefits and consider harms as secondary outcomes, if they consider harms at all [54–57]. Even systematic reviews and meta-analyses of multiple randomized trials are typically underpowered to detect uncommon and rare events [58]. Non-randomized studies can provide better estimates of differences in uncommon harms, harms that occur after prolonged exposure, harms with long latency (e.g., occurring after acute treatment or after treatment discontinuation), and harms that occur in target populations who use interventions outside of trials [59–64]. “Real-world evidence” is used increasingly to evaluate the benefits and harms of interventions, including data from electronic health records, claims, and surveillance systems [64–69]. For new drugs, pre-clinical studies also provide valuable evidence about harms

that are difficult to observe in people, such as effects on developing embryos and drug interactions. For policy interventions and public health interventions, multiple studies might be needed to evaluate possible psychological harms, social harms, and effects on equity [70, 71]. Separate reviews of benefits and harms could include different types of evidence that address different types of questions about intervention effects.

There are pervasive and consequential imitations in syntheses of harms. For example, a study of cancer screening guidelines found that most guidelines failed to include information about harms. Some guidelines also compared harms associated with a single procedure with benefits that accrued through multiple procedures and treatments that resulted from cancer screening [72]. As another example, we compared systematic reviews of the drug gabapentin and found that different types of harms and different effect estimates were reported across reviews that included the same primary studies [28]. Lack of reliability is an indication that most of the reviews we assessed could be misleading.

Conclusions about harms in systematic reviews may be unreliable when reviews are limited to subsets of the at-risk population. For example, an anticonvulsant drug might be used to treat epilepsy, postherpetic neuralgia, or bipolar disorder. It might not be possible for participants with different health problems to experience the same benefits, so it would be sensible to conduct three separate reviews of: reduction in seizures for people with epilepsy, reduction in pain for people with postherpetic neuralgia, and reduction in depression for people with bipolar disorder. By contrast, participants with different health problems might be at risk of the same harms. Therefore, a comprehensive review of an intervention's harms would include all users who are at risk of harm, regardless of the health conditions for which they use the intervention. For example, people can experience dizziness whether they take anticonvulsants to treat epilepsy, postherpetic neuralgia, or bipolar disorder. A review examining the likelihood of dizziness for people with postherpetic neuralgia should not necessarily exclude studies of people with epilepsy or bipolar disorder. When limited to studies of people with a single health problem, systematic reviews might miss relevant evidence about harms.

Overviews can combine evidence about harms across systematic reviews [73], but combining systematic reviews does not address limitations in the study types included in those reviews (e.g., randomized trials), nor can overviews resolve methodological heterogeneity across reviews (e.g., different selection criteria for reporting harms). Because it would be redundant to synthesize the same evidence about harms in different systematic

reviews, we argue that only one review of harms might be needed for a given intervention.

For the reasons above, regulators including the US Food and Drug Administration (FDA) consider harms in multiple populations, animal studies, and clinical pharmacology when developing “prescribing information” for patients and clinicians [74]. Prescribing information applies to all users who might be at risk of an intervention's potential harms. Conclusions in systematic reviews might differ from regulatory guidance when reviews are restricted to subsets of the at-risk population or when reviews are otherwise incomplete.

Conducting separate reviews about benefits and harms is a feasible alternative to the current paradigm. Multiple systematic reviews about potential benefits in different subpopulations could reference or incorporate a comprehensive review about potential harms in the entire population of users. Where applicable, heterogeneity across subpopulations could be explored in reviews of harms. This approach to systematic reviews would have implications for other research such as clinical trials and guideline development. For example, core outcome sets typically describe the minimum benefits and harms to include in all studies of a particular health problem [75]. Instead, core outcome sets for benefits could be organized around health problems (e.g., depression, anxiety), while core outcome sets for harms could be organized around different types of interventions (e.g., antidepressants, antipsychotics) [76].

In conclusion, many studies show that systematic reviews use suboptimal methods to assess harms, and that conclusions in systematic reviews are unreliable (which implies that some of their conclusions are wrong). Mandating the inclusion of harms in all systematic reviews of interventions might exacerbate rather than solve known problems; this sort of tokenism should be replaced by focused reviews of harms conducted by teams with appropriate expertise. Reviewing harms and benefits separately could minimize redundant work across reviews of the same interventions, and it could reduce the likelihood that reviews of the same interventions reach inconsistent (incorrect) conclusions.

Acknowledgements

We thank Sean Grant for comments on the manuscript.

Authors' contributions

Conceptualization, EMW, RQ, and TL. EMW wrote the first draft of the manuscript. RQ and TL reviewed and edited the manuscript. The authors read and approved the final manuscript.

Funding

No specific funding was received for this report.

Availability of data and materials

Not applicable.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 2 November 2022 Accepted: 10 April 2023

Published online: 15 April 2023

References

- Peryer G, Golder S, Junqueira D, et al. Chapter 19: Adverse effects. In: Higgins J, Thomas J, Chandler J, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 63. 2022.
- Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol*. 2010;63(5):502–12. <https://doi.org/10.1016/j.jclinepi.2008.06.007>.
- Institute of Medicine. *Finding what works in health care: standards for systematic reviews*. Washington: The National Academies Press; 2011.
- Chou R, Aronson N, Atkins D, et al. Assessing harms when comparing medical interventions. In: *Methods guide for effectiveness and comparative effectiveness reviews: Agency for Healthcare Research and Quality (US)*. 2008.
- Thomas J, Kneale D, McKenzie J, et al. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins JTJ, Chandler J, Cumpston M, Li T, Page M, Welch V, editors., et al., *Cochrane Handbook for Systematic Reviews of Interventions* version 63. 2022.
- Bennetts M, Whalen E, Ahadiet S, et al. An appraisal of meta-analysis guidelines: how do they relate to safety outcomes? *Res Synth Methods*. 2017;8(1):64–78. <https://doi.org/10.1002/jrsm.1219>.
- Higgins J, Lasserson T, Chandler J, et al. *Methodological expectations of Cochrane Intervention Reviews*. London: Cochrane; 2022.
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10(1):89. <https://doi.org/10.1186/s13643-021-01626-4>.
- Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160. <https://doi.org/10.1136/bmj.n160>.
- Zorzela L, Loke YK, Ioannidis JP, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ*. 2016;352:i157. <https://doi.org/10.1136/bmj.i157>.
- Hammad TA, Neyarapally GA, Pinheiro SP, et al. Reporting of meta-analyses of randomized controlled trials with a focus on drug safety: an empirical assessment. *Clin Trials*. 2013;10(3):389–97. <https://doi.org/10.1177/1740774513479467>.
- Zorzela L, Golder S, Liu Y, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ*. 2014;348:f7668. <https://doi.org/10.1136/bmj.f7668>.
- Qureshi R, Mayo-Wilson E, Li T. Harms in systematic reviews paper 1: an introduction to research on harms. *J Clin Epidemiol*. 2022;143:186–96. <https://doi.org/10.1016/j.jclinepi.2021.10.023>.
- Simmonds MC, Brown JV, Heirs MK, et al. Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. *Ann Intern Med*. 2013;158(12):877–89. <https://doi.org/10.7326/0003-4819-158-12-201306180-00005>.
- Rodgers MA, Brown JV, Heirs MK, et al. Reporting of industry funded study outcome data: comparison of confidential and published data on the safety and effectiveness of rhBMP-2 for spinal fusion. *BMJ*. 2013;346:f3981. <https://doi.org/10.1136/bmj.f3981>.
- Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol*. 2007;7:32. <https://doi.org/10.1186/1471-2288-7-32>.
- Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported inadequately part 1: systematic adverse events. *J Clin Epidemiol*. 2019;113:20–7. <https://doi.org/10.1016/j.jclinepi.2019.04.022>.
- Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported inadequately part 2: nonsystematic adverse events. *J Clin Epidemiol*. 2019;113:11–9. <https://doi.org/10.1016/j.jclinepi.2019.04.020>.
- Papaioannou D, Cooper C, Mooney C, et al. Adverse event recording failed to reflect potential harms: a review of trial protocols of behavioral, lifestyle and psychological therapy interventions. *J Clin Epidemiol*. 2021;136:64–76. <https://doi.org/10.1016/j.jclinepi.2021.03.002>.
- Mayo-Wilson E, Fusco N, Hong H, et al. Opportunities for selective reporting of harms in randomized clinical trials: selection criteria for non-systematic adverse events. *Trials*. 2019;20(1):553. <https://doi.org/10.1186/s13063-019-3581-3>.
- Pitrou I, Boutron I, Ahmad N, et al. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med*. 2009;169(19):1756–61. <https://doi.org/10.1001/archinternmed.2009.306>.
- Cornelius VR, Sauzet O, Aiy S, et al. Adverse event reporting in randomised trials of neuropathic pain: challenges for clinical usefulness of safety data. *Trials*. 2011;12(S1):1–2. <https://doi.org/10.1186/1745-6215-12-s1-a12>.
- Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*. 2013;8(7):e66844. <https://doi.org/10.1371/journal.pone.0066844>.
- Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349:g6501. <https://doi.org/10.1136/bmj.g6501>.
- Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews. *Stat Med*. 2015;34(20):2781–93. <https://doi.org/10.1002/sim.6525>.
- Hammad TA, Pinheiro SP, Neyarapally GA. Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clin Trials*. 2011;8(5):559–70. <https://doi.org/10.1177/1740774511419165>.
- Qureshi R, Mayo-Wilson E, Rittiphairoj T, et al. Harms in systematic reviews paper 2: methods used to assess harms are neglected in systematic reviews of gabapentin. *J Clin Epidemiol*. 2022;143:212–23. <https://doi.org/10.1016/j.jclinepi.2021.10.024>.
- Qureshi R, Mayo-Wilson E, Rittiphairoj T, et al. Harms in systematic reviews paper 3: given the same data sources, systematic reviews of gabapentin have different results for harms. *J Clin Epidemiol*. 2022;143:224–41. <https://doi.org/10.1016/j.jclinepi.2021.10.025>.
- Tang E, Ravaud P, Riveros C, et al. Comparison of serious adverse events posted at ClinicalTrials.gov and published in corresponding journal articles. *BMC Med*. 2015;13:189. <https://doi.org/10.1186/s12916-015-0430-4>.
- Golder S, Loke YK, Wright K, et al. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. *PLoS Med*. 2016;13(9):e1002127. <https://doi.org/10.1371/journal.pmed.1002127>.
- Hodkinson A, Dietz KC, Lefebvre C, et al. The use of clinical study reports to enhance the quality of systematic reviews: a survey of systematic review authors. *Syst Rev*. 2018;7(1):117. <https://doi.org/10.1186/s13643-018-0766-x>.
- Golder S, Loke Y, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Med Res Methodol*. 2006;6:3. <https://doi.org/10.1186/1471-2288-6-3>.
- Jefferson T, Doshi P, Boutron I, et al. When to include clinical study reports and regulatory documents in systematic reviews. *BMJ Evid Based Med*. 2018;23(6):210–7. <https://doi.org/10.1136/bmjebm-2018-110963>.
- Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS Med*. 2012;9(4):e1001201. <https://doi.org/10.1371/journal.pmed.1001201>.
- Wieseler B, Wolfram N, McGauran N, et al. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished

- clinical study reports with publicly available data. *PLoS Med.* 2013;10(10):e1001526. <https://doi.org/10.1371/journal.pmed.1001526>.
36. Schroll JB, Penninga EI, Gotsche PC. Assessment of adverse events in protocols, clinical study reports, and published papers of trials of orlistat: a document analysis. *PLoS Med.* 2016;13(8):e1002101. <https://doi.org/10.1371/journal.pmed.1002101>.
 37. Council for International Organizations of Medical Sciences (CIOMS). Evidence synthesis and meta-analysis for drug safety. Geneva: WHO Press; 2016.
 38. Doshi P, Jones M, Jefferson T. Rethinking credible evidence synthesis. *BMJ.* 2012;344:d7898. <https://doi.org/10.1136/bmj.d7898>.
 39. U.S. Food and Drug Administration. Meta-analyses of randomized controlled clinical trials to evaluate the safety of human drugs or biological products guidance for industry (FDA-2018-D-3710) Silver Spring, MD. 2018. (<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/meta-analyses-randomized-controlled-clinical-trials-evaluate-safety-human-drugs-or-biological>).
 40. Mayo-Wilson E, Li T, Fusco N, et al. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). *Res Synth Methods.* 2018;9(1):2–12. <https://doi.org/10.1002/jrsm.1277>.
 41. Li T, Mayo-Wilson E, Fusco N, et al. Caveat emptor: the combined effects of multiplicity and selective reporting. *Trials.* 2018;19:497. <https://doi.org/10.1186/s13063-018-2888-9>.
 42. Mayo-Wilson E, Doshi P, Dickersin K. Are manufacturers sharing data as promised? *BMJ.* 2015;351:h4169. <https://doi.org/10.1136/bmj.h4169>.
 43. Golder S, Loke YK, Wright K, et al. Most systematic reviews of adverse effects did not include unpublished data. *J Clin Epidemiol.* 2016;77:125–33. <https://doi.org/10.1016/j.jclinepi.2016.05.003>.
 44. Veroniki AA, Ashoor HM, Le SPC, et al. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *J Clin Epidemiol.* 2019;113:176–88. <https://doi.org/10.1016/j.jclinepi.2019.05.031>.
 45. Whittington CJ, Kendall T, Fonagy P, et al. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *Lancet.* 2004;363(9418):1341–5. [https://doi.org/10.1016/S0140-6736\(04\)16043-1](https://doi.org/10.1016/S0140-6736(04)16043-1).
 46. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry.* 2006;63(3):332–9. <https://doi.org/10.1001/archpsyc.63.3.332>.
 47. Xu C, Yu T, Furuya-Kanamori L, et al. Validity of data extraction in evidence synthesis practice of adverse events: reproducibility study. *BMJ.* 2022;377:e069155. <https://doi.org/10.1136/bmj-2021-069155>.
 48. Zhou Y, Zhu B, Lin L, et al. Protocols for meta-analysis of intervention safety seldom specified methods to deal with rare events. *J Clin Epidemiol.* 2020;128:109–17. <https://doi.org/10.1016/j.jclinepi.2020.09.023>.
 49. Cornelius VR, Perrio MJ, Shakir SA, et al. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiol Drug Saf.* 2009;18(12):1223–31. <https://doi.org/10.1002/pds.1844>.
 50. Xu C, Li L, Lin L, et al. Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. *J Clin Epidemiol.* 2020;123:91–9. <https://doi.org/10.1016/j.jclinepi.2020.03.020>.
 51. Ren Y, Lin L, Lian Q, et al. Real-world performance of meta-analysis methods for double-zero-event studies with dichotomous outcomes using the Cochrane Database of Systematic Reviews. *J Gen Intern Med.* 2019;34(6):960–8. <https://doi.org/10.1007/s11606-019-04925-8>.
 52. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med.* 2004;23(9):1351–75. <https://doi.org/10.1002/sim.1761>.
 53. Bradburn MJ, Deeks JJ, Berlin JA, et al. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med.* 2007;26(1):53–77. <https://doi.org/10.1002/sim.2528>.
 54. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med.* 2016;13(5):e1002028.
 55. Golder S, Loke YK, Zorzela L. Some improvements are apparent in identifying adverse effects in systematic reviews from 1994 to 2011. *J Clin Epidemiol.* 2013;66(3):253–60. <https://doi.org/10.1016/j.jclinepi.2012.09.013>.
 56. Hopewell S, Wolfenden L, Clarke M. Reporting of adverse events in systematic reviews can be improved: survey results. *J Clin Epidemiol.* 2008;61(6):597–602. <https://doi.org/10.1016/j.jclinepi.2007.10.005>.
 57. Ernst E, Pittler MH. Assessment of therapeutic safety in systematic reviews: literature review. *BMJ.* 2001;323(7312):546.
 58. Papanikolaou PN, Ioannidis JP. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med.* 2004;117(8):582–9. <https://doi.org/10.1016/j.amjmed.2004.04.026>.
 59. Kaufman DW, Shapiro S. Epidemiological assessment of drug-induced disease. *Lancet.* 2000;356(9238):1339–43. [https://doi.org/10.1016/S0140-6736\(00\)02826-9](https://doi.org/10.1016/S0140-6736(00)02826-9).
 60. Vandembroucke JP. Benefits and harms of drug treatments. *BMJ.* 2004;329(7456):2–3. <https://doi.org/10.1136/bmj.329.7456.2>.
 61. Jick H. The discovery of drug-induced illness. *N Engl J Med.* 1977;296(9):481–5. <https://doi.org/10.1056/NEJM197703032960904>.
 62. Dieppe P, Bartlett C, Davey P, et al. Balancing benefits and harms: the example of non-steroidal anti-inflammatory drugs. *BMJ.* 2004;329(7456):31–4. <https://doi.org/10.1136/bmj.329.7456.31>.
 63. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med.* 2005;142(12 Pt 2):1090–9. https://doi.org/10.7326/0003-4819-142-12-part_2-200506211-00009.
 64. Saldanha I, Skelly A, Vander Ley K, et al. Inclusion of nonrandomized studies of interventions in systematic reviews of intervention effectiveness: an update. methods guide for comparative effectiveness reviews. Rockville: Agency for Healthcare Research and Quality; 2022. <https://doi.org/10.23970/AHRQEPCCMETHODSGUIDENRSI>. (Prepared by the Scientific Resource Center under Contract No. 290-2017-00003-C. AHRQ Publication No. 22-EHC033).
 65. US Food and Drug Administration (2018). Framework for FDA's real-world evidence program. Silver Spring, MD. <https://www.fda.gov/media/120060/download>.
 66. Desai RJ, Matheny ME, Johnson K, et al. Broadening the reach of the FDA Sentinel system: a roadmap for integrating electronic health record data in a causal analysis framework. *NPJ Digit Med.* 2021;4(1):170. <https://doi.org/10.1038/s41746-021-00542-0>.
 67. Shimabukuro TT, Nguyen M, Martin D, et al. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). *Vaccine.* 2015;33(36):4398–405. <https://doi.org/10.1016/j.vaccine.2015.07.035>.
 68. European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU). <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu> (Accessed 23 Feb 2023).
 69. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the Joint ISPOR-ISPE Special Task Force on Real-World Evidence in Health Care Decision Making. *Value Health.* 2017;20(8):1003–8. <https://doi.org/10.1016/j.jval.2017.08.3019>.
 70. Lorenz T, Oliver K. Adverse effects of public health interventions: a conceptual framework. *J Epidemiol Community Health.* 2014;68(3):288–90. <https://doi.org/10.1136/jech-2013-203118>.
 71. Bonell C, Jamal F, Melendez-Torres GJ, et al. 'Dark logic': theorising the harmful consequences of public health interventions. *J Epidemiol Community Health.* 2015;69(1):95–8. <https://doi.org/10.1136/jech-2014-204671>.
 72. Kamineni A, Doria-Rose VP, Chubak J, et al. Evaluation of harms reporting in U.S. cancer screening guidelines. *Ann Intern Med.* 2022. <https://doi.org/10.7326/M22-1139>.
 73. Sachse T, Kanji S, Thabet P, et al. Clinical utility of overviews on adverse events of pharmacological interventions (preprint). *Res Square.* 2023. <https://doi.org/10.21203/rs.3.rs-2057583/v1>.
 74. Gassman AL, Nguyen CP, Joffe HV. FDA regulation of prescription drugs. *N Engl J Med.* 2017;376(7):674–82. <https://doi.org/10.1056/NEJMr1602972>.
 75. Clarke M, Williamson PR. Core outcome sets and systematic reviews. *Syst Rev.* 2016;5:11. <https://doi.org/10.1186/s13643-016-0188-6>.
 76. Cornelius VR, Sauzet O, Williams JE, et al. Adverse event reporting in randomised controlled trials of neuropathic pain: considerations for future practice. *Pain.* 2013;154(2):213–20. <https://doi.org/10.1016/j.pain.2012.08.012>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.