Risk of Bias and Quality of Reporting in Colon and Rectal Cancer Systematic Reviews Cited by National Comprehensive Cancer Network Guidelines

C Wayant, BS[1] , L Puljak, PhD[2], M Bibens, MS[3], and M Vassar, PhD[3]

[1]Department of Biomedical Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, USA; [2]Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Zagreb, Croatia; [3]Department of Psychiatry and Behavioral Health, Oklahoma State University Center for Health Sciences, Tulsa, OK, USA.

INTRODUCTION: Given the changing landscape of colorectal cancer, systematic reviews are likely to play a key role in advancing the understanding of prevention, diagnosis, and treatment.

METHODS: We conducted a cross-sectional investigation of the risk of bias and reporting quality of systematic reviews referenced by colon and rectal cancer National Comprehensive Cancer Network (NCCN) guidelines. We used two widely accepted tools: Risk of Bias in Systematic reviews (ROBIS) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

RESULTS: Using ROBIS, only 3 (4.8%) systematic reviews were judged with low risk of bias, 35 (55.6%) systematic reviews were judged with unclear risk of bias, and 25 (39.7%) systematic reviews were judged with high risk of bias. Across all systematic reviews, the individual bias domains at the highest risk of bias were domains 1 (protocol and eligibility criteria) and 2 (methods to identify and select studies). Across all studies, the median adherence to PRISMA was 74.1% (IQR 69.2–80.0%), corresponding to approximately 20 of 27 items.

CONCLUSIONS: Systematic reviews cited in NCCN guidelines for colon and rectal cancer are frequently at unclear or high risk of bias and do not report key systematic review items that are important for the critical appraisal of results.

KEY WORDS: systematic review; bias; oncology; colon cancer; rectal cancer.

Key Message We found that systematic reviews cited by colorectal cancer NCCN guidelines were frequently at high or unclear risk of bias. High risk of bias means poor methods were reported; unclear risk of bias means there was incomplete reporting, making bias judgments difficult. The implication for readers is that the NCCN guideline authors do not have robust systematic review evidence at their disposal.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11606-020-05639-y) contains supplementary material, which is available to authorized users.

INTRODUCTION

Systematic reviews (SRs) combine results from similar, individual studies in an attempt to provide a reliable answer to a healthcare question.[1] Previous SRs have demonstrated benefit to both physicians and patients. An iconic example of how SRs have influenced clinical practice concerns antenatal corticosteroid use in women at risk for preterm birth.[2] This SR demonstrated a survival benefit for preterm infants and resolved unanswered clinical questions, such as the long-term effects of corticosteroids on surviving infants. The authors of this SR reported their methodology and conducted their study in a manner that promotes reproducibility and trustworthiness. Examples of such practices include publishing the search strategy used to identify included studies, assessing the included studies for risk of bias, and using robust statistical methods to combine these studies for determining the pooled treatment effect. These practices, however, are not common as previous studies suggest that SRs often fail to report detailed search strategies or evaluate for risk of bias using valid tools.[3, 4] Such incomplete SR methodology may lead to biased results, the consequences of which are far-reaching, including spurious alterations to clinical practice and future research questions. These consequences are especially harmful if the SR is cited to support clinical practice guideline (CPG) recommendations.

Clinical practice guidelines are consensus documents developed by a group of experts that are designed to guide patient care.[5] Systematic reviews are often used, alongside robust clinical trials, to assign level 1 evidence to CPG recommendations.[6] However, for a SR to be a trustworthy and accurate source of clinical information, its methods and reporting must first be robust. Previous investigations of SRs underpinning CPG recommendations have identified suboptimal methodology and reporting.[7–9] Such SRs may be irreproducible, and the critical appraisal of their summary effects by CPG development groups may be compromised.

Systematic reviews are also critically important to the prevention, diagnosis, and treatment of colon and rectal cancer. Currently, colorectal cancer is being diagnosed in patients under age 50 at an increasing rate.[10] Even worse, there is evidence that colon cancer in younger patients may differ from colon cancer in older adults with respect to clinical

presentation, pathologic findings, and tumor biology.[11] Therefore, there is a fundamental need for robust research based on rigorous methodology to continue the advancements in understanding preventing, diagnosing, and treating colorectal cancer. Systematic reviews are likely to play a key role in these advancements.

Therefore, the primary objective of this study is to assess the risk of bias and reporting quality in SRs cited in the National Comprehensive Cancer Network (NCCN) guidelines for the treatment of colon (Version 2.2018)[12] and rectal (Version 1.2018)[13] cancer, since NCCN guidelines are heavily used by physicians to guide patient care[14] and SRs are the highest level of medical evidence. To do so, we applied the novel Risk of Bias in Systematic Reviews (ROBIS) tool[15] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.[16]

METHODS

In this review, we adhered to PRISMA guidance where possible and applicable,[17] despite this study not being a systematic review. Our choice to adhere to PRISMA was made because no validated reporting checklist for cross-sectional or meta-epidemiological studies exists. We defined a SR according to the PRISMA-P (PRISMA for protocols) definition: articles that explicitly stated methods to identify studies (i.e., a search strategy), explicitly stated methods of study selection (e.g., eligibility criteria and selection process), and explicitly described methods of synthesis (or other type of summary).[17] Systematic reviews were gathered from a previous study whose protocol is available via the Open Science Framework.[18] To identify SRs in the previous study, one author (CW) manually screened the reference list of and Discussion (main narrative) section of the NCCN colon and rectal cancer guidelines. Keywords searches were conducted for studies referenced as a "systematic review," "meta-analysis," "review," or "metaanalysis." Any referenced papers published in the Cochrane Database for Systematic Reviews were also extracted. All extracted references were added to a PubMed collection and exported to Rayyan[19] for title and abstract screening. In accordance with the previous study from which our sample is derived, a SR was included if it had at least one meta-analysis that included at least one randomized-controlled trial. Further, each SR that met these criteria must have been published after 2011 to allow uptake of the PRISMA statement (published in 2009). To extract data for this study, we developed Google Forms based on the ROBIS and PRISMA statements. Two authors extracted data in duplicate with masking for ROBIS (CW, LP) and for PRISMA (CW, MB). All discrepancies were resolved between the authors, with the availability of a third-party adjudicator (MV).

The ROBIS statement assesses whether a SR is at risk of bias based on its methods and conduct. ROBIS includes 3 phases: (1) assess relevance (optional), (2) identify concerns

with the SR process, and (3) judge risk of bias of the SR. We opted to exclude the first phase of assessing relevance, since all SRs from the NCCN colorectal guidelines are relevant to our research question. In the second phase, signaling questions are asked to guide an investigator through 4 bias domains: (1) study eligibility criteria, (2) identification and selection of studies, (3) data collection and study appraisal, and (4) synthesis and findings. Signaling questions are answered as "yes," "probably yes," "no," "probably no," and "no information." We followed the guidance of the ROBIS statement manual when answering signaling questions. Based on the answers to the signaling questions in each domain, each domain is assigned a risk of bias grade. Potential grades include "high," "low," or "unclear." In the third phase, signaling questions are again asked, except these questions relate to the overall reliability of the SR findings. If limitations are identified in phase 2, SR authors will be required to address these limitations and interpret the findings accordingly. Further, SR authors will be required to not emphasize findings based on statistical significance alone. After completing phase 3, a summary judgment (e.g., high, low, or unclear) regarding the risk of bias for the SR will be rendered. For this study, we distinguished between high and unclear risk of bias based on the completeness of SR reporting. For example, to be judged high risk of bias, the SR would have to report the use of flawed methods, such as a flawed risk of bias scale, use of only one database to gather studies, or single-author data extraction. If an SR did not report enough information for us to determine whether the methods were at high or low quality, we judged that SR with unclear risk of bias.

Contrary to ROBIS, PRISMA assesses reporting quality, so rather than asking if an item was conducted adequately, PRISMA asks whether an item was reported. For example, whereas ROBIS may ask about the adequacy and comprehensiveness of the search strategy, PRISMA asks if a search strategy was reported. This distinction is important for complementing the assessment of risk of bias in SR methodology measured using ROBIS. The PRISMA checklist contains 27 items divided into 7 domains: Title, Abstract, Introduction, Methods, Results, Discussion, and Funding. For each item, we judged whether a SR fully reported what was required by PRISMA and scored each item with a 1 (yes—fully reported) or 0 (not reported or partially reported). Our rationale for partial reported being grouped with "not reported" is that PRISMA does only asks whether an item is mentioned, not that it was methodologically robust. So, failure to completely report an item indicates that a key piece of that item is not available to readers. For example, item 5 requires SR authors to indicate if a protocol exists and direct readers to it with a citation or registration number. Failure to direct readers means readers are unable to access the protocol, just as if the SR authors did not mention a protocol at all. After scoring each PRISMA item, we summed the adherence across each article and each item. It should be noted that PRISMA is not a measurement tool, but a reporting checklist. Despite that fact,
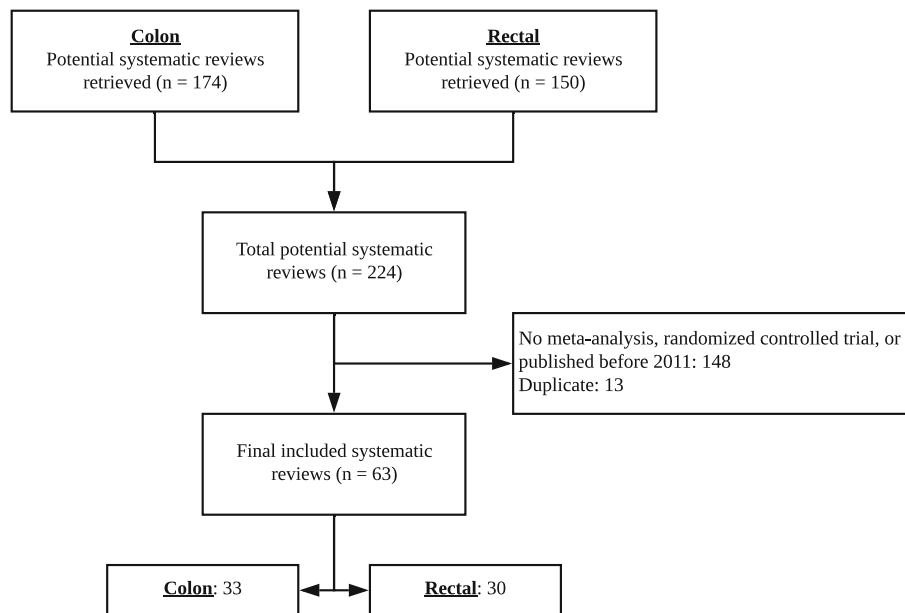
Fig. 1 Flow diagram of included and excluded systematic reviews.

PRISMA has been used in numerous previous studies as a measurement tool, since no other validated option to assess reporting quality exists.

We used Google Sheets for all summary statistics and measures of central tendency (medians and interquartile ranges (IQR)).

RESULTS

Sixty-three SRs (33 colon, 30 rectal) were included in this study (Fig. 1). The 63 SRs included a median of 10 (IQR 7–16) studies and a median of 3160 (IQR 1270–5825) patients. Twenty-four (38.1%) SRs stated that they adhered to PRISMA guidelines. The included SRs were cited a total of 76 times, overwhelmingly for support of NCCN committee recommendations (56/76, 73.7%). SRs were also cited as evidence of harm for available therapies (10/76, 13.2%), as evidence that contradicts the committee recommendations (5/76, 6.6%), and as background evidence where no recommendation was given (5/76, 6.6%). All primary data and the protocol from this investigation are available via the Open Science Framework.[20]

Using ROBIS, only 3 (4.8%) SRs were judged with low risk of bias, 35 (55.6%) SRs were judged with unclear risk of bias, and 25 (39.7%) SRs were judged with high risk of bias (Table 1). Across all SRs, the individual bias domains at the highest risk of bias were domains 1 (protocol and eligibility criteria) and 2 (methods to identify and select studies). Twenty-eight (44.4%) SRs were at high risk of bias for domain 1 and 26 (41.3%) were at high risk of bias for domain 2. Specific areas of concern in these two domains were the lack of information about publication of an SR protocol, language restrictions, choice of bibliographic databases, and searches for gray literature. Domains 3 (data collection and appraisal)

and 4 (synthesis of findings) were predominantly judged as unclear risk of bias, corresponding to a frequent lack of critical information that would have aided our assessments. Individual study scores are shown in eTable 1.

Across all studies, the median adherence to PRISMA was 74.1% (IQR 69.2–80.0%), corresponding to approximately 20 of 27 items (eTable 2). Two items had 100% adherence: item 3 (rationale for SR) and item 21 (presentation of results with measures of precision). Thirteen additional items had adherence greater than 75%, with 7 items maintaining adherence greater than 90%. Only 3 items had adherence lower than 25%: item 8 (search strategy), item 5 (protocol and registration), and item 4 (provision of PICO-format research question). There was no difference between SRs that adhered to PRISMA ($n = 24$) and did not adhere to PRISMA ($n = 39$) in terms of number of fully reported items (20 PRISMA vs. 20.5 no PRISMA).

DISCUSSION

Our investigation found that SRs cited in colorectal guidelines are frequently at unclear or high risk of bias and do not report key SR items that are important for the critical appraisal of results. Specifically, that our predominant risk of bias judgment was unclear signals that much of the critical SR methodological items were missing or poorly described. Our finding—that SRs adhered to a median of 20/27 PRISMA items—may appear at odds with our risk of bias findings. However, the difference in these two findings highlights our key takeaway: a SR item may be reported but still represent a flawed method, thus placing the SR at risk of bias. Thus, our findings identify two key action items for future and ongoing SRs in colorectal cancer: ensure SRs report all items from

Table 1 Summary of Risk of Bias Judgments Across All Studies (*n* = 63)

	Domain 1	Domain 2	Domain 3	Domain 4	Consensus
	Protocol and eligibility criteria	Methods to identify and/or select studies	Data collection and appraisal	Synthesis of findings	
Low	6 (9.5%)	5 (7.9%)	5 (7.9%)	4 (6.3%)	3 (4.8%)
Unclear	29 (46.0%)	32 (50.8%)	43 (68.3%)	46 (73.0%)	35 (55.6%)
High	28 (44.4%)	26 (41.3%)	15 (23.8%)	13 (20.6%)	25 (39.7%)

PRISMA and ensure SRs describe methods in enough detail to facilitate critical appraisal of results.

Two key examples of how missing or poorly described information may affect the critical appraisal of an SR relate to study protocols and risk of bias evaluations. In our sample, SRs rarely directed the reader to a publicly available, a priori protocol (2/63, 3.2%). It has been shown that SRs, like randomized controlled trials,[21, 22] exhibit significant rates of selective outcome reporting—defined as the selective inclusion, omission, or alteration of study outcomes, often due to statistical significance.[23] Thus, the lack of a publicly available protocol leaves the possibility that SR results are published at the author's discretion, rather than at the behest of a prespecified protocol. Similarly, a lack of detail regarding risk of bias evaluations may compromise the validity of meta-analytic effects in an SR. In our study, authors often reported that a risk of bias evaluation was conducted (46/63, 73.0%), but further inspection of the risk of bias methods showed that many authors used outdated, flawed tools. For example, authors frequently used the Jadad scale for assessing risk of bias of included clinical trials. The Jadad scale is notorious for its omission of allocation concealment as a bias domain, and according to the Cochrane Handbook, use of the Jadad scale is "explicitly discouraged."[24] Thus, the use of the Jadad scale leaves the possibility that interventional effects shown in the included colorectal SRs are confounded by bias that is undetected by SR authors. Furthermore, even if authors used Cochrane risk of bias tool, they often reported only judgment for individual risk of bias domains, without an accompanying comment that explained the judgment. It has been shown previously that authors frequently make erroneous judgments (i.e., judgments that were not in line with the accompanying comment), and thus, not in line with recommendations available in the Cochrane Handbook.[25–27] Therefore, inadequate reporting of Cochrane risk of bias tool prevents readers to verify accuracy of authors' judgments.

The cohort of SRs we analyzed is unique since these SRs informed the evidence base of NCCN colorectal guidelines. However, this sample of SRs is likely not the only, or even the primary, source of evidence for most NCCN recommendations, since the field of oncology relies heavily on randomized controlled trial data. Indeed, the NCCN categories of recommendations simply state that "high-level evidence" and "uniform NCCN consensus" are necessary to achieve level 1 evidence status[28]. Nonetheless, the findings from our study warrant concern due to the predominance of unclear or high risk of bias judgments and variability in reporting quality. For example, in the NCCN rectal cancer guidelines, seven SRs were cited in the discussion of laparoscopy vs. open resection (Jiang et al. 2015; Zhao et al. 2016; Zhang et al. 2014; Xiong et al. 2012; Vennix et al. 2014; Arezzo et al. 2013; Trastulli et al. 2012). Five of these SRs were at high or unclear risk of bias, while 2 were at low risk, including the only Cochrane review. There was no discussion of the risk of bias for any of these SRs. This oversight may be reasonable in this case because of the dearth of other data available and cited for laparoscopy, all pointing to a fairly certain conclusion of its risks and benefits. Moreover, in this case, the low risk of bias SRs had similar findings as the high and unclear risk of bias SRs. However, even this scenario highlights an important point—risk of bias assessments is crucial to reasoned discussions and serves to augment the ongoing, skillful clinical appraisal inherent to CPG panel discussions. In this case, where the benefits and risks of laparoscopy are fairly well-established, the harm of omitting risk of bias from a CPG discussion may be benign, but for emerging therapies with less certain benefit, risk of bias evaluations are necessary because the risk of false positive or negative results may have a broad impact of CPG recommendations and clinical practice. This study has several key limitations. First, our findings may not be generalizable to all colorectal SRs, since we only evaluated SRs cited by the NCCN rather than all colorectal SRs available. Next, we discourage the interpretation of our findings to mean that NCCN recommendations are at risk of bias, since the NCCN recommendations rely on other robust research, such as clinical trials, that we did not include in our investigation. Any judgments about the quality of NCCN recommendations would need to be supported with thorough assessment of all evidence included and validated tools for assessment of clinical guidelines. Moreover, the included NCCN guidelines included 1698 total references, so our 63 included SRs represent only a small fraction of the cited evidence. Finally, this study is limited by investigating only guidelines written for healthcare professionals, rather than NCCN guidelines for patients. In conclusion, our investigation of the risk of bias and quality of reporting of SRs referenced by the NCCN guidelines for colon and rectal cancer found that SRs are commonly at high risk of bias and do not fully report key items. Specifically, we found that a SR item may be mentioned, but may report a flawed method or incomplete report all aspects of the item. The implication for the treatment and management of colon and rectal cancer, which relies on high-

quality evidence for demographically diverse patients, is that summary effects may not exemplify the trust normally imputed on systematic reviews and meta-analyses. Further, even though the objective of our investigation is not to question the strength of NCCN guideline recommendations, our findings may be of concern to oncologists who heavily rely on NCCN recommendations. The NCCN developers use what literature is available to formulate recommendations, and thus, we recommend more stringent SR methodology and reporting be enforced in journal publications. When readers or guideline developers encounter a biased SR, we recommend careful critical appraisal of the results and conclusions, since bias may result in false positive or false negative results.

Corresponding Author: *C Wayant, BS; Department of Biomedical Sciences Oklahoma State University Center for Health Sciences, 1111 West 17th Street, Tulsa, OK 74104, USA (e-mail: cole. wayant@okstate.edu).*

REFERENCES

1. Higgins JPT GS, ed. Cochrane Handbook for Systematic Reviews of Interventions: online version (5.1.0, March 2011). *The Cochrane Collaboration.* 2011. www.cochrane-handbook.org.
2. **Roberts D**, **Dalziel S**. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database Syst Rev.* 2006;(3):CD004454.
3. **Roundtree AK**, **Kallen MA**, **Lopez-Olivo MA**, et al. Poor reporting of search strategy and conflict of interest in over 250 narrative and systematic reviews of two biologic agents in arthritis: a systematic review. *J Clin Epidemiol.* 2009;62(2):128–137.
4. **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.
5. Institute of Medicine Board on Health Care Services Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical Practice Guidelines We Can Trust.* National Academies Press; 2011.
6. **Schünemann H**, **Brożek J**, **Guyatt G**, Oxman A, ed. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations.; 2013.
7. **Nissen T**, **Wayant C**, **Wahlstrom A**, et al. Methodological quality, completeness of reporting and use of systematic reviews as evidence in clinical practice guidelines for paediatric overweight and obesity. *Clin Obes.* 2017;7(1):34–45.
8. **Ross A**, **Rankin J**, **Beaman J**, et al. Methodological quality of systematic reviews referenced in clinical practice guidelines for the treatment of opioid use disorder. *PLoS One.* 2017;12(8):e0181927.
9. **Scott J**, **Howard B**, **Sinnett P**, et al. Variable methodological quality and use found in systematic reviews referenced in STEMI clinical practice guidelines. *Am J Emerg Med.* June 2017. doi:https://doi.org/10.1016/j. ajem.2017.06.010
10. **Bailey CE**, **Hu C-Y**, **You YN**, et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg.* 2015;150(1):17–22.
11. **Weinberg BA**, **Marshall JL**, **Salem ME**. The Growing Challenge of Young Adults With Colorectal Cancer. *Oncology* . 2017;31(5):381–389.
12. National Comprehensive Cancer Network (NCCN). Colon Cancer. https:// www.nccn.org/professionals/physician_gls/pdf/colon.pdf. Published October 19, 2018. Accessed January 29, 2019.
13. National Comprehensive Cancer Network (NCCN). Rectal Cancer. https://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf. Published August 7, 2018. Accessed January 29, 2019.
14. **Jagsi R**, **Huang G**, **Griffith K**, et al. Attitudes toward and use of cancer management guidelines in a national sample of medical oncologists and surgeons. *J Natl Compr Canc Netw.* 2014;12(2):204–212.
15. **Whiting P**, **Savović J**, **Higgins JPT**, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol.* 2016;69:225–234.
16. **Moher D**, **Liberati A**, **Tetzlaff J**, **Altman DG**, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med.* 2009;151(4):264–269, **W64**.
17. **Moher D**, **Shamseer L**, **Clarke M**, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1.
18. **Cole Wayant**, Matthew J Page, Matt Vassar. Reproducibility of Oncology Meta-Analyses. https://osf.io/kxj9z/. Published May 23, 2018.
19. **Ouzzani M**, **Hammady H**, **Fedorowicz Z**, **Elmagarmid A**. Rayyan—a web and mobile app for systematic reviews. *Syst Rev.* 2016;5(1):210.
20. **Wayant C**, **Puljak L**, **Bibens M**, **Vassar M**. Protocol: Risk of bias and reporting quality in systematic reviews underpinning colorectal cancer guidelines. https://osf.io/a24bu/. Published April 29, 2019. .
21. **Wayant C**, **Scheckel C**, **Hicks C**, et al. Evidence of selective reporting bias in hematology journals: A systematic review. *PLoS One.* 2017;12(6):e0178379.
22. **Raghav KPS**, **Mahajan S**, **Yao JC**, et al. From Protocols to Publications: A Study in Selective Reporting of Outcomes in Randomized Trials in Oncology. *J Clin Oncol.* 2015;33(31):3583–3590.
23. **Chan A-W**, **Hróbjartsson A**, **Haahr MT**, **Gøtzsche PC**, **Altman DG**. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291(20):2457–2465.
24. **Higgins J**, **Altman D**, **Sterne J**. 8.3.3 Quality scales and Cochrane reviews. Cochrane Handbook for Systematic Reviews of Interventions [Version 5.1.0]. https://handbook-5-1.cochrane.org/chapter_8/8_3_3_ quality_scales_and_cochrane_reviews.htm. Published March 2011. .
25. **Propadalo I**, **Tranfic M**, **Vuka I**, **Barcot O**, **Pericic TP**, **Puljak L**. In Cochrane reviews, risk of bias assessments for allocation concealment were frequently not in line with Cochrane's Handbook guidance. *J Clin Epidemiol.* 2019;106:10–17.
26. **Saric F**, **Barcot O**, **Puljak L**. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. *J Clin Epidemiol.* April 2019. doi:https://doi.org/10.1016/j.jclinepi.2019.04. 007
27. **Babic A**, **Tokalic R**, Amílcar Silva Cunha J, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Med Res Methodol.* 2019;19(1):76.
28. **National Comprehensive Cancer Network**. NCCN Categories of Evidence and Consensus. https://www.nccn.org/professionals/physician_ gls/categories_of_consensus.aspx.