Original Article

Harms Reporting Is Inadequate in Systematic Reviews Regarding Hip Arthroscopy

Caleb Peters, B.S., Matthew Chancellor, B.S., Holly Flores, B.A., Audrey Wise, B.A., B.S., Morgan Garrett, B.S., Jake Checketts, D.O., Chad Hanson, D.O., and Matt Vassar, Ph.D.

Purpose: To investigate the quality of harms reporting in systematic reviews (SRs) regarding hip arthroscopy in the current literature. **Methods:** In May 2022, an extensive search of 4 major databases was performed identifying SRs regarding hip arthroscopy: MEDLINE (PubMed and Ovid), EMBASE, Epistemonikos, and Cochrane Database of Systematic Reviews. A cross-sectional analysis was conducted, in which investigators performed screening and data extraction of the included studies in a masked, duplicate fashion. AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews-2) was used to assess the methodologic quality and bias of the included studies. The corrected covered area was calculated for SR dyads. **Results:** A total of 82 SRs were included in our study for data extraction. Of these SRs, 37 reported under 50% of the harms criteria (37 of 82, 45.1%) and 9 did not report harms at all (9 of 82, 10.9%). A significant relation was found between completeness of harms reporting and overall AMSTAR appraisal ($P = .0261$), as well as whether a harm was listed as a primary or secondary outcome ($P = .0001$). Eight SR dyads had corrected covered areas of 50% or greater and were compared for shared harms reported. **Conclusions:** In this study, we found inadequate harms reporting in most SRs concerning hip arthroscopy. **Clinical Relevance:** With the magnitude of hip arthroscopic procedures being performed, adequate reporting of harms-related information in the research surrounding this treatment is essential in assessing the efficacy of the treatment. This study provides data in relation to harms reporting in SRs regarding hip arthroscopy.

Hip arthroscopy is a rapidly growing discipline within orthopaedic surgery. Between 2011 and 2018, the incidence of arthroscopic hip procedures increased to greater than 85%.[1] These procedures are most commonly performed for femoroacetabular impingement syndrome (FAIS). One study found that treatment of FAIS with hip arthroscopy had a mean aggregate of increased productivity of nearly $10,000 per patient.[2] Furthermore, through cost analysis, this study found a mean cumulative 10-year societal monetary savings of nearly $70,000 and over a 2-year gain of quality-adjusted life-years per patient.[2] Because of the quickly increasing use of hip arthroscopy in orthopaedic surgery and especially owing to the positive outcomes it provides for patients, it is essential that research conducted on hip arthroscopy be of sound methodologic quality because this research is building the foundation for this rapidly evolving modality.

Several tools exist to assess the methodologic quality of a study. However, one of the most common methods involves using standardized reporting guidelines. These methodologic safeguards have been developed to address most study types and have even been adapted for clinical practice guidelines and other patient care documents.[3-5] One recently developed reporting guideline, Consolidated Standards of Reporting Trials—Extension for Harms (CONSORT-Harms), was developed as a tool for standardizing the quality of reporting when it comes to adverse patient events or

From the Office of Medical Student Research, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma, U.S.A. (C.P., M.C., H.F., A.W., M.G., M.V.); Department of Orthopaedic Surgery, Oklahoma State University Medical Center, Tulsa, Oklahoma, U.S.A. (J.C., C.H.); and Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma, U.S.A. (M.V.).

The authors report the following potential conflicts of interest or sources of funding: M.V. reports receipt of funding from the National Institute on Drug Abuse, the National Institute on Alcohol Abuse and Alcoholism, the US Office of Research Integrity, Oklahoma Center for Advancement of Science and Technology, and internal grants from Oklahoma State University Center for Health Sciences, all outside of the present work. Full ICMJE author disclosure forms are available for this article online, as supplementary material.

Received August 2, 2022; accepted October 5, 2022.

Address correspondence to Caleb Peters, B.S., Oklahoma State University Center for Health Sciences, 1111 W 17th St, Tulsa, OK 74107, U.S.A. E-mail: caleb.petersosu@gmail.com

2666-061X/22844

https://doi.org/10.1016/j.asmr.2022.10.010

complications. The harms extension includes an additional 10 items specifically tailored for the reporting of harms.[6] Furthermore, this extension addresses instances wherein authors only report positive outcomes while lacking in their reporting of harms resulting from studied interventions.[7-10] There have been several investigations using the harms checklist that suggest that the percentage of adherence to these standards is suboptimal. For example, a 2021 *JAMA Ophthalmology* study investigating the quality of reporting of harms-related data within a sample of retinal detachment clinical trials revealed that harms data are infrequently quantified or reported.[11] Furthermore, a study that adapted the harms checklist for systematic reviews (SRs) found that the reporting in trials cited within SRs is unreliable and lacks standardization.[12] These findings come with concern because SRs are regarded as the highest quality of evidence available in the medical literature.

SRs have been considered the cornerstone of evidence-based medicine.[13] Because of the importance of SRs, many clinical practice guidelines and other practice-influencing documents use SRs as their source of evidence. Owing to the weight that SRs hold in modern-day evidence-based medicine, the reporting of harms within SRs must be objectively and clearly reported in a standardized manner, especially for rapidly developing and evolving disciplines such as hip arthroscopy. The purpose of this study was to investigate the quality of harms reporting in SRs regarding hip arthroscopy in the current literature. Our hypothesis was that harms-related data would be incompletely reported, as seen in previous studies of other disciplines.

Methods

Study Design

Following the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) guidelines, we performed a cross-sectional analysis that investigated the reporting of harms in SRs related to hip arthroscopy.[14,15] Human subjects were not involved in this study; therefore, it was not subject to institutional review board approval.

Harms Terminology

This investigation adhered to PRISMA harms terminology to classify harms. A glossary of these terms can be found in Figure 1.

Search Strategy

The following 4 databases were searched using a search strategy created by a SR librarian: MEDLINE (PubMed and Ovid), EMBASE, Epistemonikos, and Cochrane Database of Systematic Reviews. The search returns were uploaded into the SR screening platform Rayyan (https://rayyan.qcri.org/). Two investigators (C.P. and M.C.) separately screened titles and abstracts, removed duplicates, and determined the studies that met the inclusion criteria outlined in the "Eligibility Criteria" section. After completing the initial screening, these investigators were unmasked to resolve any disagreements. In the event that they could not reach an agreement, a third-party adjudicator (H.F.) was available to resolve any discrepancies; however, further adjudication was not needed.

Search String

The search string was uploaded to the Open Science Framework (OSF).[16]

Eligibility Criteria

To be included in our sample, studies had to meet the following inclusion criteria: SR with or without a meta-analysis designed to evaluate hip arthroscopy for any indication, written in English, including only human subjects. Studies were excluded from our sample for the following reasons: SRs not related to hip arthroscopy, SRs that evaluated hip arthroscopy as one of many interventions in the same treatment group, animal studies, duplicates, withdrawn or retracted studies, clinical trials, narrative reviews, letters to the editor, observational studies (including cohort studies, case-control studies, cost-effective studies, and cross-sectional studies), and literature reviews, as well as any remaining study that did not meet the inclusion criteria.

Training

The investigators (C.P. and M.C.) completed the Johns Hopkins SR course provided by the Coursera platform (Mountain View, CA)[17] before working through several harms training exercises. Using a pilot-tested Google form (Alphabet, Mountain View, CA), the authors were trained on how to complete the extraction process for each data item as it pertains to harms. After harms training, the authors were trained to use AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews-2) to assess each SR for methodologic quality in video and lecture format.[18] By use of another pilot-tested Google form, each data item of the AMSTAR-2 tool was recorded to save individual responses to items. With an additional investigator (H.F.) present to provide assistance, the investigators (C.P. and M.C.) performed data extraction for both harms reporting and AMSTAR-2 on 3 example SRs in a masked, duplicate fashion. The investigators were then unmasked to review and compare responses and discuss any disagreements. The authors repeated the process until they were able to reach agreement on all items. All training was led by a senior investigator (M.V.) who has published several articles regarding SR methodology.

Glossary of terms*

Adverse effect **-** An unfavorable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it

Adverse drug reaction**-** An adverse effect specific to a drug

Adverse event **-** An unfavorable outcome that occurs during or after the use of a drug or other intervention and the causal relation between the intervention and the event is at least a reasonable possibility

Complication **-** An adverse event or effect following surgical and other invasive intervention

Harm **-** The totality of possible adverse consequences (if single or multiple) of an intervention or therapy; harms are the direct opposite of benefits

Safety **-** Substantive evidence of an absence of harm. The term is often misused when there is sample absence of evidence of harm

Side effect **-** Any unintended effect, adverse or beneficial, of a drug that occurs at doses normally used for treatment

Toxicity **-** Drug related harm. The term may be most appropriate for laboratory determined measurements, although it is also used in relation to clinical events

*Adapted from Zorzela L, Loke YK, Ioannidis JP, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. BMJ. 2016;352:i157.

Fig 1. Glossary of terms. (Adapted from Zorzela L, Loke YK, Ioannidis JP, et al. PRISMA harms checklist: Improving harms reporting in systematic reviews. *BMJ* 2016;352:i157.)

Data Extraction

Two investigators (C.P. and M.C.) extracted the following study characteristics from each of the included SRs using a pilot-tested Google form: title; Rayyan identification number; journal name; conflict-of-interest statement; funding source; funding statement; indications of the intervention; whether a meta-analysis was performed; whether the SR found the intervention favorable; whether a harm was evaluated as an outcome and, if so, whether the outcome was primary or secondary; whether a patient-reported outcome measure (PROM) was used and, if so, whether the PROM specified harms; and whether the SR mentioned adherence to PRISMA guidelines.[15] Using similar methodology to Mahady et al.,[19] the 2 aforementioned investigators extracted harms data items from each SR. These data items can be found in Table 1. Each of these items was coded as "yes" or "no." Furthermore, using similar methods to Qureshi et al,[20-22] 2 authors (C.P. and M.C.) extracted additional harms items from each included SR. These data items can be found in Table 2. Items 1, 4, and 7 were coded as "yes" or "no." Items 2 and 6 were coded as a free response. Item 3 was coded as "yes," "no", "there was a protocol available but it did not address harms," or "could not find protocol." Because item 5 was dependent on whether a given SR contained a meta-analysis, this item was coded as "qualitatively only," "quantitatively only," "both qualitatively and quantitatively," or "not applicable." Regarding all of the previously described items, data extraction and coding were performed in a masked, duplicate fashion with a

Table 1. Assessment for Completion of Harms Reporting (N = 82) According to Mahady et al.[19]

	Frequency (%)		Total Systematic Reviews, n (%)
	Yes	No	
Harms assessment			
1. Are harms stated in the title or abstract?	52 (63.4)	30 (36.6)	
2. Are harms presented in the introduction?	45 (54.9)	37 (45.1)	
3. Are harms listed and separately defined in the methods?	18 (22.0)	64 (78.1)	
4. Are grades and/or severity scales used to classify harms in the methods?	8 (9.8)	74 (90.2)	
5. Is there a method of harms data collection stated in the methods?	47 (57.3)	35 (42.7)	
6. Is there a planned statistical analysis for harms stated in the methods?	33 (40.2)	49 (59.8)	
7. Is the number of patients available for harms analyses stated in the results?	67 (81.7)	15 (18.3)	
8. Is the number of treatment discontinuations in each arm reported in the results?	0 (0.0)	82 (100.0)	
9. Are absolute figures for each harm in treatment and control groups presented in the results?	53 (64.6)	29 (35.4)	
10. Are limitations of harms analyses discussed?	39 (47.6)	43 (52.4)	
11. Is a balanced discussion of harms and benefits provided?	55 (67.1)	27 (32.9)	
12. Did the authors discuss what future research would be needed to better clarify harms?	24 (29.3)	58 (70.7)	
Harms items completed			
0% of harms items			9 (11.0)
1%-50% of harms items			28 (34.1)
>50% of harms items			45 (54.9)

third-party adjudicator (H.F.) available to resolve any discrepancies.

We additionally assessed the overlapping use of primary studies in each SR using the corrected covered area (CCA).[23] The CCA is a mathematical equation designed to determine the number of overlapping primary studies between 2 or more SRs. The equation is written according to a table in which each SR is compared against the other SRs in the sample. The equation is as follows:

$$CCA = \frac{C - U}{(U * R) - U}$$

in which C is the total number of citations across the included SRs, U is the total number of unique citations, and R is the number of SRs included in the sample. To calculate the CCA, the total number of unique citations (U) is subtracted from the total number of citations (C), giving the total number of non-unique citations. This value is then divided by the total number of unique citations (U) multiplied by the total number of SRs in the sample (R) minus the total number of unique citations (U). The result provides the amount of citation overlap between 2 or more SRs. We considered a CCA of 50% or greater to indicate high overlap; between 20% and 50%, moderate overlap; and less than 20%, minimal overlap. If overlap between 2 SRs was found to be high, we compared the extracted harms from each SR to determine similarities or differences in the reporting of harms.[23] In this study, the term "dyad" is used when referring to a pair of SRs.

Additionally, 2 investigators (C.P. and M.C.) performed a quality appraisal of each SR using the AMSTAR-2 instrument.[24] AMSTAR-2 is an appraisal tool that is widely accepted as a valid method of evaluating SR quality.[25] Each of the 16 items was scored as "yes" if all criteria were met, "partial yes" if only some of the criteria were met, and "no" if no criteria were met. Items 11, 12, and 15 pertain to SRs that contained a meta-analysis. Therefore, if an SR did not contain a meta-analysis, the SR received a score out of 13 rather than 16. Each SR—with or without a meta-analysis—was given a quality rating of "high," "moderate," "low," or "critically low" based on the AMSTAR-2 quality assessment generator.

Data Analysis

Individual item completion of general characteristics, reporting of harms, and AMSTAR-2 for all SRs in our study was reported using percentages and frequencies. We performed a bivariate analysis between variables such as quality rating, general characteristics, and harms reporting to determine whether any relations exist. The choice of statistical test depended on data characteristics (e.g., statistical assumptions and distributional qualities). We considered $P \leq .05$ to represent a statistically significant relation. Regarding the CCA, we reported the overall number of primary studies across all SRs in our sample; the range of primary studies used by a single SR; and how many primary studies were reported in only 1 SR, in 2 to 4 SRs, and in 5 or more SRs.[22] We also calculated the overall CCA across all SRs. Furthermore, we compared individual harms and results in all pairs of reviews with a CCA of 50% or greater, indicating very high overlap of primary studies.[23] Microsoft Excel (Microsoft, Redmond, WA) was used for data cleaning, and all data analyses were conducted with Stata (version 16.1; StataCorp, College Station, TX).

Table 2. Assessment for Completion of Harms Reporting (N = 82) According to Qureshi et al[20-22]

Harms Assessment	n (%)
1. Did the study pre-specify any harms?	
Yes	53 (64.6)
No	29 (35.4)
2a. What were the types of harms assessed?	Uploaded to OSF
2b. What language was used to describe those types of harms?	Uploaded to OSF
2c. What were the effect estimates used to assess harms?	
Mean difference	11 (13.3)
Odds ratio	9 (10.8)
Relative risk	1 (1.2)
Risk ratio	1 (1.2)
Interclass correlation coefficient	2 (2.4)
Not applicable	58 (69.9)
3. Was a prespecified protocol available that addressed harms?	
Yes	1 (1.2)
No	59 (72.0)
Could not find protocol	2 (2.4)
Available protocol did not address harms	19 (23.2)
4. Were any specific harms or harms language included in the search strategy?	
Yes	19 (23.2)
No	63 (76.8)
5. Was a given harm assessed qualitatively or quantitatively (i.e., within a meta-analysis)?	
Both quantitative and qualitative	22 (26.8)
Only quantitative	39 (47.6)
Only qualitative	3 (3.7)
Not applicable	17 (20.7)
6. If a given harm was assessed quantitatively, what models and assumptions were used?	
Fixed effects	4 (4.9)
Random effects	9 (11.0)
Fixed effects and random effects	1 (1.2)
Not applicable	68 (82.9)
7. Did the authors apply selection criteria to reported harms?	
Yes	0 (0.0)
No	82 (100.0)

OSF, Open Science Framework.

Reproducibility

In the interest of promoting transparency and reproducibility, the study protocol, search string, raw data, analysis scripts, data dictionaries, and extraction forms were uploaded to the Open Science Framework (OSF).[26] This study was performed in conjunction with other studies investigating different interventions using similar methodology.

Results

Study Selection Process

The initial search string provided 2,299 records, of which 2,179 were excluded after title and abstract screening. An additional 38 studies were excluded during full-text screening and data extraction. In total, 82 SRs were included in our study for further analysis. A flow diagram of our screening process and reasons for exclusion is presented in Figure 2.

Characteristics of Included Studies

For the included SRs, the date of publication ranged from 2009 to 2022. A total of 62 SRs found hip arthroscopy to be a favorable intervention (62 of 82, 75.6%). The most frequently reported indication for hip arthroscopy was femoroacetabular impingement (34 of 82, 41.5%). Of the evaluated SRs, 57 reported harms as a primary outcome (57 of 82, 69.5%), 13 reported harms as a secondary outcome (13 of 82, 15.9%), and 12 did not report harms as an outcome of interest (12 of 82, 14.6%). General characteristics of the included studies are shown in Table 3.

Harms Extraction

Of the 82 included SRs, 52 reported harms either in the title or in the abstract (52 of 82, 63.4%). A total of 64 SRs did not list or separately define harms in the methods (64 of 82, 78.1%). A method for harms data collection was stated in the methods of 47 SRs (47 of 82, 57.3%). Twenty-four studies discussed, in some way, what future research would be needed to better clarify harms (24 of 82, 29.3%). A prespecified protocol that addressed harms was found in 1 SR (1 of 82, 1.2%). PROMs were used in 68 of the SRs (68 of 82, 82.9%). Four SRs defined the measurement scales (and subscales) of the scoring systems used and specified harms included in these (4 of 68, 5.9%). A total of 37 SRs reported under 50% of the harms criteria (37 of 82, 45.1%). and 9 SRs did not report harms at all (9 of 82, 10.9%). A complete report of results regarding harms reporting can be found in Tables 1, 2, and 3.

AMSTAR-2 Assessment

By use of the AMSTAR-2 appraisal tool, 60 SRs were rated as critically low (60 of 82, 73.2%); 17, low (17 of 82, 20.7%); 4, moderate (4 of 82, 4.9%); and 1, high (1 of 82, 1.2%).

Associations

The Kruskal-Wallis test showed significant relations between completeness of harms reporting (via the methodology of Mahady et al.[19]) and (1) a critically low AMSTAR appraisal ($P = .0261$) and (2) whether a harm was listed as a primary or secondary outcome ($P = .0001$). No significant relation was found between completeness of harms reporting (Mahady) and whether the SR reported adherence to PRISMA guidelines. Among the SRs in our sample, there was a statistically significant association between completeness of harms reporting (Mahady) and whether a PROM scoring system was used that did not specify harms ($P = .0012$).

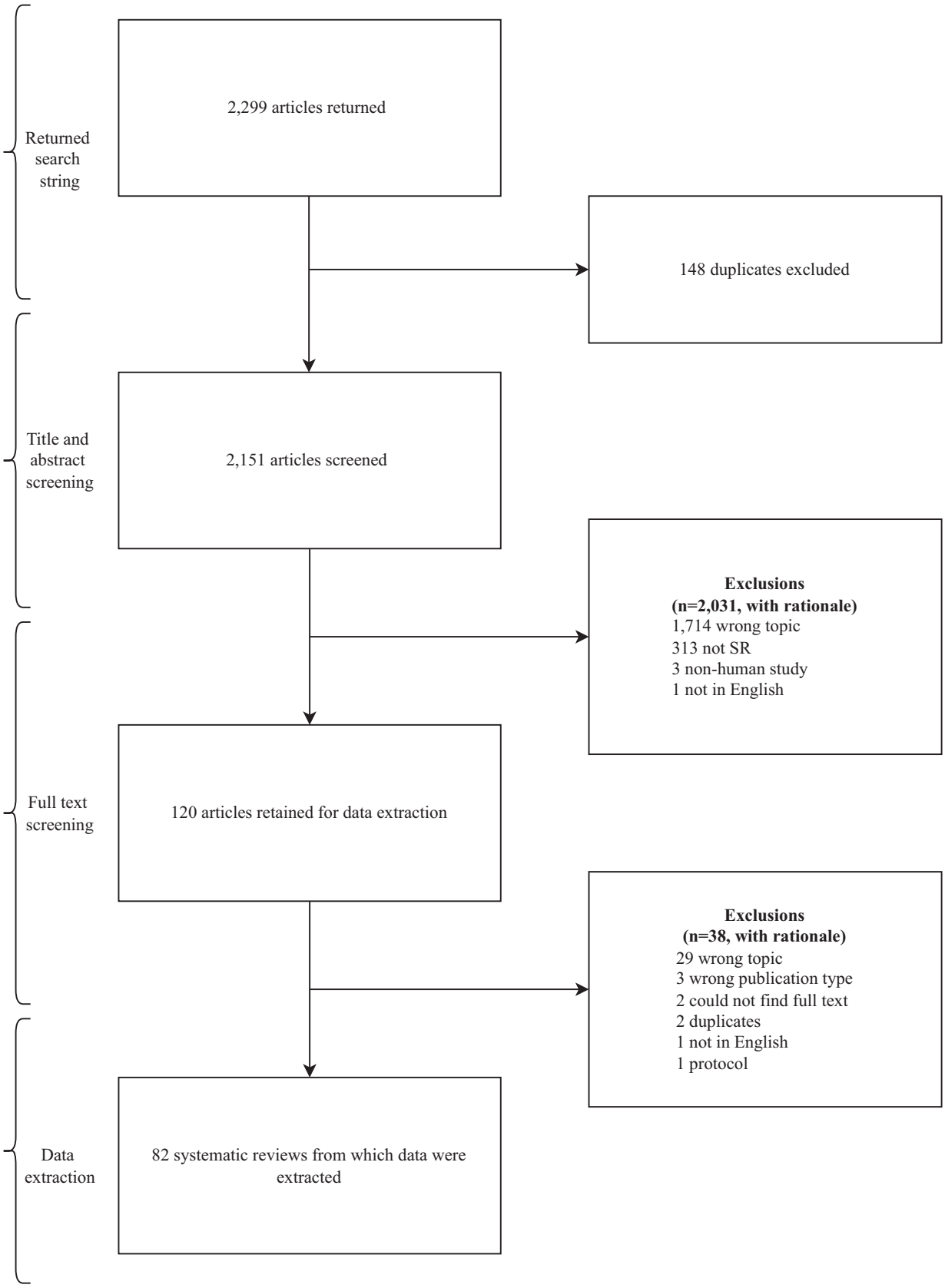

Fig 2. Flow diagram of study selection. (SR, systematic review.)

Table 3. Summary of Characteristics of Included Studies (N = 82)

Review Characteristic	n (%)
Indication	
Femoroacetabular impingement	34 (41.5)
Multiple indications	17 (20.7)
Acetabular labral tear and chondral lesion	5 (6.1)
Acetabular labral tear and femoroacetabular impingement	5 (6.1
Hip dysplasia	4 (4.9)
Acetabular labral tear	2 (2.4)
Hip capsule laxity or instability	2 (2.4)
Coxa saltans	2 (2.4)
Acetabular dysplasia	1 (1.2)
Acetabular labral tear and ligamentum teres injury	1 (1.2)
Acetabular retroversion	1 (1.2)
Femoroacetabular impingement and hip dysplasia	1 (1.2)
Femoroacetabular impingement and mild acetabular dysplasia	1 (1.2)
Hip osteoarthritis	1 (1.2)
Legg-Calvé-Perthes disease	1 (1.2)
Ligamentum teres injury	1 (1.2)
Septic arthritis	1 (1.2)
Borderline developmental dysplasia of hip	1 (1.2)
Traumatic hip dislocation	1 (1.2)
Adherence to PRISMA guidelines mentioned in study	
Yes	60 (73.2)
No	22 (26.8)
Intervention deemed favorable	
Yes	62 (75.6)
No	20 (24.4)
Classification of harms as primary or secondary outcome or neither	
Primary outcome	57 (69.5)
Secondary outcome	13 (15.9)
Neither	12 (14.6)
Use of scoring system (PROMs)	
Yes with harms specified	4 (4.9)
Yes with harms not specified	64 (78.0)
No	14 (17.1)
Conflicts of interest	
Yes	48 (58.5)
No	23 (28.5)
Funding source	
Not funded	14 (17.1)
Not mentioned	36 (43.9)
Private	35 (42.7)
Public	8 (9.8)
AMSTAR-2 rating	
High	1 (1.2)
Moderate	4 (4.9)
Low	17 (20.7)
Critically low	60 (73.2)

AMSTAR-2, A Measurement Tool to Assess Systematic Reviews-2; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses; PROM, patient-reported outcome measure.

Corrected Covered Area

In the 75 SRs available for CCA analysis, 1,096 unique primary studies were cited. Of the 82 SRs included in this review, 7 were excluded from CCA analysis because the primary articles used in the SRs could not be obtained. The calculated CCA for all 75 SRs was 0.6%. The most citations of a given study was 273 whereas the least was 3. Of the 2,775 dyads—pairs of SRs with overlap—8 had a CCA value that was considered high overlap (>50%) and 12 had a CCA value that was considered moderate overlap (20%-50%); the remaining 2,755 dyads had a CCA value that was considered minimal overlap (<20%). The reporting of harms items between each of the 8 dyads with high overlap was compared. Of the primary studies in our sample, 23 were included in 5 or more SRs, 315 were included in 2 to 4 SRs, and 758 were included in 1 SR. A complete list of all harms items reported in each SR, along with the CCA value, can be found in Table 4.

Discussion

Through our cross-sectional analysis of harms reporting in SRs on hip arthroscopy, we found that approximately half of our sample reported less than 50% of the harms criteria and 11% of SRs did not report harms at all. Furthermore, although harms were listed as the primary outcome in nearly 70% of SRs, fewer than one-fourth separately defined harms in the methods and only 1 SR included a protocol addressing harms. This information suggests that investigators often introduce harms as an outcome of interest but fail to assess and adequately report them in their studies.

Incomplete reporting of harms presents a problem in fully understanding the potential consequences of a procedure. Among the SRs in our sample, the SR by Horner et al.[35] reported harms in the abstract and as a primary outcome in the study, stating that the purpose of the study was to report clinical outcomes and complication rates as they relate to hip arthroscopy. However, the authors failed to define harms in both the methods and the protocol. The harms of interest included deep vein thrombosis, perineal numbness, superficial wound infection, minor scrotal skin burn, transient paresthesia in the ipsilateral foot, and deep infection requiring open irrigation. Despite these harms of interest, the "Results" section of this article did not sufficiently describe these complications and there was not a balanced discussion regarding the harms.

We compared our findings with the broader medical literature on harms reporting and found similar results. A study conducted by Saini et al.[36] looked at clinical studies in a cohort of 92 Cochrane SRs to determine whether there was selective non-reporting of harms data. They found harms to be under-reported in 86% of their sample. In another study, Stubenrouch et al.[37] found that despite the CONSORT (Consolidated Standards of Reporting Trials) guidelines, current reporting of harms in surgical trials failed to communicate clear treatment outcomes with patients. Finally, in a recent publication in the *Journal of Arthroplasty*, harms reporting in trials supporting orthopaedic surgery clinical practice guidelines was found to be suboptimal.[38]

Table 4. Hip Arthroscopy Harms Reported by Paired Reviews With Corrected Cover Area of 50% or Greater (8 Dyads)

	Harms Reported	
	Study 1	Study 2
Dyad 1,180	Yeung et al.[27] (2016)	Duplantier et al.[28] (2016)
	Hip instability	
	Hip dislocation	Hip dislocation
		Hip subluxation
		Continued pain
		Repeated fall
		Femoral neck stress fracture
	Percentage of harms mentioned in study of Duplantier et al. (2016): 50.0% (1 of 2)	Percentage of harms mentioned in study of Yeung et al. (2016): 20.0% (1 of 5)
Dyad 1,480	Zhang et al.[29] (2016)	Qiao et al.[30] (2020)
	Nerve damage	Nerve paresis
	Wound infection	Superficial infection
	Wound dehiscence	
	Reoperation	
		Hematoma
		Heterotopic ossification
		Deep venous thrombosis
		Heterotopic ossification
		Transient paresthesia of the pudendal nerve
		Persistent pain
		New-onset symptomatic internal snapping
		Reversible pudendal nerve paresis
		Perineal cutaneous necrosis
		Compartment syndrome
	Percentage of harms mentioned in study of Qiao et al. (2020): 50.0% (2 of 4)	Percentage of harms mentioned in study of Zhang et al. (2016): 16.7% (2 of 12)
Dyad 2,314	Casartelli et al.[31] (2021)	Ferreira et al.[32] (2020)
	Slipped capital femoral epiphysis	
	Proximal femoral physeal separation	
	Growth disturbance	
	Neurapraxia	
	Infection	Superficial wound infection
		Injuries to lateral cutaneous nerve of thigh
		Chronic pain
		Muscle soreness
		Numbness in groin
		Proximal thigh numbness
		Swelling and infection
	Percentage of harms mentioned in study of Ferreira et al. (2020): 20.0% (1 of 5)	Percentage of harms mentioned in study of Casartelli et al. (2021): 14.3% (1 of 7)
Dyad 2,322	Casartelli et al.[31] (2021)	Schwabe et al.[33] (2020)
	Slipped capital femoral epiphysis	
	Proximal femoral physeal separation	
	Growth disturbance	
	Neurapraxia	
	Infection	
		Thigh numbness
		Hip infection leading to arthroplasty
		Hip osteoarthritis
		Muscle soreness or spasms
		Hip pain or stiffness
		Unscheduled hospital appointments
		Heterotopic ossification
		Fracture
		Reoperation rate
		Superficial wound infection
		Temporary lateral femoral cutaneous neurapraxia
	Percentage of harms mentioned in study of Schwabe et al. (2020): 0.0% (0 of 5)	Percentage of harms mentioned in study of Casartelli et al. (2021): 0.0% (0 of 11)
Dyad 2,333	Casartelli et al.[31] (2021)	Mok et al.[34] (2021)
	Slipped capital femoral epiphysis	
	Proximal femoral physeal separation	
	Growth disturbance	
	Neurapraxia	
	Infection	
	Percentage of harms mentioned in study of Mok et al. (2021): 0.0% (0 of 5)	Percentage of harms mentioned in study of Casartelli et al. (2021): 0.0% (0 of 0)
Dyad 2,432	Ferreira et al.[32] (2020)	Schwabe et al.[33] (2020)
	Muscle soreness	Muscle soreness or spasms
	Numbness in groin	
	Proximal thigh numbness	Thigh numbness
	Swelling and infection	
	Superficial wound infection	Superficial wound infection
	Injuries to lateral cutaneous nerve of thigh	Temporary lateral femoral cutaneous neurapraxia
	Chronic pain	
		Hip osteoarthritis
		Heterotopic ossification
		Fracture
		Reoperation rate
		Hip infection leading to arthroplasty
		Hip pain or stiffness
		Unscheduled hospital appointments

(continued)

Table 4. Continued

	Harms Reported	
	Study 1	Study 2
	Percentage of harms mentioned in study of Schwabe et al. (2020): 57.1% (4 of 7)	Percentage of harms mentioned in study of Ferreira et al. (2020): 36.4% (4 of 11)
Dyad 2,443	Ferreira et al.[32] (2020) Muscle soreness Numbness in groin Proximal thigh numbness Swelling and infection Superficial wound infection Injuries to lateral cutaneous nerve of thigh Chronic pain	Mok et al.[34] (2021)
	Percentage of harms mentioned in study of Mok et al. (2021): 0.0% (0 of 7)	Percentage of harms mentioned in study of Ferreira et al. (2020): 0.0% (0 of 0)
Dyad 2,615	Schwabe et al.[33] (2020) Heterotopic ossification Fracture Reoperation rate Superficial wound infection Temporary lateral femoral cutaneous neurapraxia Thigh numbness Hip infection leading to arthroplasty Hip osteoarthritis Muscle soreness or spasms Hip pain or stiffness Unscheduled hospital appointments	Mok et al.[34] (2021)
	Percentage of harms mentioned in study of Mok et al. (2021): 0.0% (0 of 11)	Percentage of harms mentioned in study of Schwabe et al. (2020): 0.0% (0 of 0)

We conducted a CCA analysis of our sample based on the methods developed by Qureshi et al.[22] The CCA accounted for overlap of primary studies that were shared between the SRs included in our sample. Overlap between the SRs could potentially lead to overestimation of our results owing to reporting of the same data. However, regarding the SRs with high overlap, our data suggest the opposite. Of the dyads in our sample, 8 (0.28%) had a high amount of overlap. Furthermore, our findings suggest that the dyads with a high amount of overlap failed to sufficiently report on the same harms, although their data were obtained from more than 50% of the same primary studies. For example, dyads 5, 7, and 8 reported 0% of the same harms.

In our sample, most SRs used a wide variety of PROM tools, such as the Harris Hip Score (HHS), Oxford Hip Score (OHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). These scoring systems are commonly used in the orthopaedic community to assess patient mobility and health. Kunze et al.[39] conducted an SR on the quantification methods within the current hip arthroscopy literature and discovered 19 unique outcome measures within their sample, stating that this finding led to diverse and inconsistent methods of quantifying the minimal clinically important difference. The prevalence and variety of PROM reporting raise a question as to whether studies are specifically addressing the harms established in PROMs. To determine whether a PROM does specifically address a harm, the overall scoring system and specific values have to be clearly defined. This issue was evident in our study because over 80% of the included SRs contained PROMs. Of these, only 5.9% described the scoring systems of the PROMs used and specified harms. For example, de Sa et al.[40] conducted an SR to evaluate indications, outcomes, and complications in the surgical management of FAIS. In this study, more than 7 PROMs were used but were not specifically defined, and only the final score was given. This information suggests that heterogeneous measurement systems such as PROMs, along with the lack of harms reporting in SRs, may lead to unclear communication of harms and research waste.

Harms reporting in SRs is likely to guide clinicians toward making more effective decisions for their patients; reduce financial costs relating to unprecedented complications; and save time, money, and energy regarding research. We suggest holding authors, peer reviewers, and journals to higher standards when looking to publish SRs. Through standardization of guidelines with the use of tools such as CONSORT-Harms and PRISMA, SRs may be held to higher standards of methodologic quality and validity. It has previously been elucidated that endorsement of reporting guidelines by journals leads to greater adherence by authors of articles published in those journals.[4,41-43] Thus, it is reasonable to assume that if journals, funding agencies, and specialty societies endorse thorough reporting of harms, greater adherence to this methodologic safeguard will be noticed. One concern is that the use of PROMs seems to be unregulated. Many of these outcome scales consist of subscales that are not explicitly described. The use of many different PROMs may lead to difficulty interpreting results, including harms. We suggest limiting the use of various PROMs to decrease heterogeneity or standardizing a set of PROMs that are easily interpreted.

Limitations

This study is not without limitations. One limitation was the subjectivity of the investigators in determining and/or classifying harms language included in the texts. For example, when multiple terms were used to describe the same pathology, subjective determination was required to identify whether harms were being reported. Furthermore, when clear details were not provided, subjective determination was required in the evaluation of whether reoperation rates were directly correlated with complications relating to a prior hip arthroscopic procedure. Another limitation to our study is the limited availability of similar studies for comparison; therefore, our results should be interpreted while taking this into consideration. Finally, we were unable to locate full reference lists for 7 of the SRs included in our study. We were unsuccessful in our attempts to contact the authors via email prior to publication.

Conclusions

In this study, we found inadequate harms reporting in most SRs concerning hip arthroscopy.

Acknowledgment

The authors are grateful to Dr. Courtney Kennedy, who assisted in the development of their search strategy, and to the Oklahoma State University medical library for its procurement of relevant literature. They are also grateful for the corrected covered area (CCA) guidance and code provided by Dr. Riaz Qureshi.

References

1. Zusmanovich M, Haselman W, Serrano B, Banffy M. The incidence of hip arthroscopy in patients with femoroacetabular impingement syndrome and labral pathology increased by 85% between 2011 and 2018 in the United States. *Arthroscopy* 2022;38:82-87.
2. Mather RC III, Nho SJ, Federer A, et al. Effects of arthroscopy for femoroacetabular impingement syndrome on quality of life and economic outcomes. *Am J Sports Med* 2018;46:1205-1213.
3. Scott J, Checketts JX, Cooper CM, Boose M, Wayant C, Vassar M. An evaluation of publication bias in high-impact orthopaedic literature. *JB JS Open Access* 2019;4:e0055.
4. Sims MT, Checketts JX, Wayant C, Vassar M. Requirements for trial registration and adherence to reporting guidelines in critical care journals: A meta-epidemiological study of journals' instructions for authors. *Int J Evid Based Healthc* 2018;16:55-65.
5. Batioja KLE, Nguyen T, Anderson JM, et al. An analysis of the evidence underpinning the American Academy Orthopaedic Surgery pediatrics clinical practice guidelines. *J Pediatr Orthop* 2022;42:e218-e223.
6. CONSORT. Harms. http://www.consort-statement.org/extensions/overview/harms. Accessed February 8, 2022.
7. Breau RH, Gaboury I, Scales CD Jr, Fesperman SF, Watterson JD, Dahm P. Reporting of harm in randomized controlled trials published in the urological literature. *J Urol* 2010;183:1693-1697.
8. Cuervo LG, Clarke M. Balancing benefits and harms in health care. *BMJ* 2003;327:65-66.
9. Haddad C, Sigha OB, Lebrun-Vignes B, Chosidow O, Fardet L. Reporting of harm and safety results in randomized controlled trials published in 5 dermatology journals. *J Am Acad Dermatol* 2017;77:98-104.e1.
10. Haidich AB, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: Survey of major general medical journals. *J Clin Epidemiol* 2011;64:124-135.
11. Xu ZY, Azuara-Blanco A, Kadonosono K, et al. Reporting of complications in retinal detachment surgical trials: A systematic review using the CONSORT extension for harms *[published online June 17, 2021]. JAMA Ophthalmol* doi:10.1001/jamaophthalmol.2021.1836.
12. Qureshi R, Mayo-Wilson E, Rittiphairoj T, McAdams-DeMarco M, Guallar E, Li T. Harms in systematic reviews paper 2: Methods used to assess harms are neglected in systematic reviews of gabapentin. *J Clin Epidemiol* 2022;143:212-223.
13. Lai N. Evidence based medicine series: Evidence based medicine: An overview. *Malays Fam Physician* 2009;4:19-22.
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 2009;339:b2535.
15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
16. OSF. https://osf.io/. Accessed June 1, 2022.
17. Deo SV, Deo VS. Systematic review and meta-analysis: A brief introduction. *Indian J Thorac Cardiovasc Surg* 2022;38:342-345.
18. *National Collaborating Centre for Method and Tools. Spotlight on Methods and Tools: AMSTAR 2* 2018. https://www.youtube.com/watch?v=gdbIFo7FCVE. Accessed June 1, 2022.
19. Mahady SE, Schlub T, Bero L, et al. Side effects are incompletely reported among systematic reviews in gastroenterology. *J Clin Epidemiol* 2015;68:144-153.
20. Qureshi R, Mayo-Wilson E, Li T. Summaries of harms in systematic reviews are unreliable paper 1: An introduction to research on harms *[published online November 3, 2021]. J Clin Epidemiol* doi:10.1016/j.jclinepi.2021.10.023.
21. Qureshi R, Mayo-Wilson E, Rittiphairoj T, McAdams-DeMarco M, Guallar E, Li T. Harms in systematic reviews paper 2: methods used to assess harms are neglected in systematic reviews. *J Clin Epidemiol* 2022;143:212-223.
22. Qureshi R, Mayo-Wilson E, Rittiphairoj T, McAdams-DeMarco M, Guallar E, Li T. Summaries of harms in systematic reviews are unreliable paper 3: Given the same data sources, systematic reviews of gabapentin have different results for harms *[published online November 3, 2021]. J Clin Epidemiol* doi:10.1016/j.jclinepi.2021.10.025.

23. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: Guidance for using the corrected covered area index. *Res Synth Methods* 2020;11:134-145.
24. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
25. AMSTAR. *Assessing the methodological quality of systematic reviews*. https://amstar.ca/. Accessed June 1, 2022.
26. Wise A, Kee MD, Rucker B, et al. *Hip arthroscopy. Open Science Framework Published May 31, 2022*, https://osf.io/yrc9z. Accessed June 23, 2022.
27. Yeung M, Memon M, Simunovic N, Belzile E, Philippon MJ, Ayeni OR. Gross instability after hip arthroscopy: an analysis of case reports evaluating surgical and patient Factors. *Arthroscopy* 2016;32:1196-1204.e1.
28. Duplantier NL, McCulloch PC, Nho SJ, Mather RC 3rd, Lewis BD, Harris JD. Hip dislocation or subluxation after hip arthroscopy: a systematic review. *Arthroscopy* 2016;32:1428-1434.
29. Zhang D, Chen L, Wang G. Hip arthroscopy versus open surgical dislocation for femoroacetabular impingement: A systematic review and meta-analysis. *Medicine* 2016;95:e5122.
30. Qiao HY, Zhang YH, Ren YM, Tian MQ. Arthroscopic versus open treatment for femoroacetabular impingement: A systematic review and meta-analyses. *Medicine* 2020;99:e23364.
31. Casartelli NC, Valenzuela PL, Maffiuletti NA, Leunig M. Effectiveness of hip arthroscopy on treatment of femoroacetabular impingement syndrome: a meta-analysis of randomized controlled trials. *Arthritis Care Res* 2021;73:1140-1145.
32. Ferreira GE, O'Keeffe M, Maher CG, et al. The effectiveness of hip arthroscopic surgery for the treatment of femoroacetabular impingement syndrome: A systematic review and meta-analysis. *J Sci Med Sport* 2021;24:21-29.
33. Schwabe MT, Clohisy JC, Cheng AL, et al. Short-term clinical outcomes of hip arthroscopy versus physical therapy in patients with femoroacetabular impingement: a systematic review and meta-analysis of randomized controlled trials. *Orthop J Sports Med* 2020;8:2325967120968490.
34. Mok TN, He QY, Teng Q, et al. Arthroscopic hip surgery versus conservative therapy on femoroacetabular impingement syndrome: a meta-analysis of RCTS. *Orthop Surg* 2021;13:1755-1764.
35. Horner NS, Ekhtiari S, Simunovic N, Safran MR, Philippon MJ, Ayeni OR. Hip Arthroscopy in patients age 40 or older: A systematic review. *Arthroscopy* 2017;33:464-475.e3.
36. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: Findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.
37. Stubenrouch FE, Cohen ES, Bossuyt PMM, Koelemay MJW, van der Vet PCR, Ubbink DT. Systematic review of reporting benefits and harms of surgical interventions in randomized clinical trials. *BJS Open* 2020;4:171-181.
38. Anderson JM, Howard C, Staggs J, et al. Inadequate reporting of complications in randomized controlled trials cited as supporting evidence underpinning AAOS CPG recommendations for hip and knee osteoarthritis: Application of the CONSORT Harms Checklist *[published online June 13, 2022]. J Arthroplasty* doi:10.1016/j.arth.2022.06.004.
39. Kunze KN, Bart JA, Ahmad M, Nho SJ, Chahla J. Large heterogeneity among minimal clinically important differences for hip arthroscopy outcomes: A systematic review of reporting trends and quantification methods. *Arthroscopy* 2021;37:1028-1037.e6.
40. de Sa D, Cargnelli S, Catapano M, et al. Femoroacetabular impingement in skeletally immature patients: A systematic review examining indications, outcomes, and complications of open and arthroscopic treatment. *Arthroscopy* 2015;31:373-384.
41. Cook C, Checketts JX, Atakpo P, Nelson N, Vassar M. How well are reporting guidelines and trial registration used by dermatology journals to limit bias? A meta-epidemiological study. *Br J Dermatol* 2018;178:1433-1434.
42. Checketts JX, Cook C, Imani S, Duckett L, Vassar M. An evaluation of reporting guidelines and clinical trial registry requirements among plastic surgery journals. *Ann Plast Surg* 2018;81:215-219.
43. Checketts JX, Sims MT, Detweiler B, Middlemist K, Jones J, Vassar M. An evaluation of reporting guidelines and clinical trial registry requirements among orthopaedic surgery journals. *J Bone Joint Surg Am* 2018;100:e15.