The Journal of Arthroplasty xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Evaluation of Quality of Lower Limb Arthroplasty Observational Studies Using the AQUILA Checklist

James B. Cowan, MD [a], Ryan A. Mlynarek, MD [a], Rob G.H.H. Nelissen, MD, PhD [b], Bart G.C.W. Pijls, MD, PhD [b], Joel J. Gagnier, ND, MSc, PhD [c]

[a] Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, Michigan
[b] Department of Orthopaedics, Leiden University Medical Center, RC Leiden, The Netherlands
[c] Departments of Orthopaedic Surgery and Epidemiology, University of Michigan, Ann Arbor, Michigan

ARTICLE INFO

Article history:
Received 8 August 2014
Accepted 24 March 2015
Available online xxxx

Keywords:
lower limb arthroplasty
observational studies
AQUILA checklist
research methodology

ABSTRACT

This study used the assessment of quality in lower limb arthroplasty (AQUILA) checklist to assess the quality of lower limb arthroplasty observational studies. Among 132 studies the mean reporting quality score was 5.4 (SD = 1.2) out of 8 possible points. Most studies adequately reported reasons for revisions (98%) and prosthesis brand and fixation (95%) in sufficient detail. Only 3% of studies adequately reported the number of patients unwilling to participate, 15% stated a clear primary research question or hypothesis, 11% reported a worst-case analysis or competing risk analysis for endpoints, and 42% reported more than 5% of patients were lost to follow-up. There is significant room for improvement in the reporting and methodology of lower limb arthroplasty observational studies. Level of evidence: Level III.

Proper clinical research methodology with complete and transparent reporting is required to ensure validity and to allow for the applicability of findings. For example, numerous studies have shown that results from lower quality studies may potentially exaggerate treatment effects [1–5]. Assessments of the quality of meta-analyses in the orthopedic literature have found that up to half of these studies have extensive or major flaws in methodology and that up to 43% demonstrate poor quality reporting [6,7]. Study quality should be assessed by meticulous, critical analysis on a study-by-study basis of numerous factors. Even relying on well-established external assessment tools such as levels of evidence may be misleading as a high level of evidence does not guarantee high-quality research nor does it guarantee the transparency or completeness of reporting of clinical research [8–10].

Numerous checklists and assessment tools are available for evaluating study methodology and reporting [11]. However many of them are designed for evaluating meta-analyses and randomized controlled trials (RCTs) and are not specific to subdisciplines such as orthopedic surgery. Surveyed orthopedic surgeons mentioned observational studies as being second only to personal experience as their source of evidence for clinical decision-making [12]. Checklists such as the STROBE statement are useful for evaluating the reporting quality of observational studies, however neither this checklist, nor any other checklist, addresses methodological quality for cohort studies [13]. Despite clear guidelines for how to use the STROBE statement, it is often used inappropriately to assess methodological quality [14]. A thorough assessment of reporting (which addresses the written content of a manuscript/report) and methodology/risk of bias (which addresses aspects of the study design and implementation) is necessary when critically evaluating the overall quality of a study.

The assessment of quality in lower limb arthroplasty (AQUILA) checklist was created through three cycles of an international Delphi panel of 37 experts (of 17 nationalities across 5 continents) in the fields of hip surgery, knee surgery, and epidemiology. Its purpose is to assess the methodology, reporting, and generalizability of case series and cohort studies in the total hip and total knee arthroplasty literature [15]. As the Delphi approach is an accepted research method for forming consensus among experts, a checklist designed using this approach has immediate content and face validity [16]. While primarily designed to investigate revision rate for aseptic loosening in total hip arthroplasty (THA) and total knee arthroplasty (TKA), the authors sought to create a checklist that was useful for case series and cohort studies with other endpoints (Fig. 1). The AQUILA checklist has been used previously to assess the quality of lower limb arthroplasty studies or to inform the methodology of lower limb arthroplasty systematic reviews [17–22].

While the reporting and methodological flaws among higher levels of evidence (RCTs) have been well studied, given the aforementioned surgeon reliance on observation studies for clinical decision-making, we believe that it is important to evaluate this issue for these studies as well. The purpose of our study was to use this AQUILA checklist to assess the reporting and methodological quality of lower limb arthroplasty observational studies from 2010 through 2011. We hypothesize that there is a high rate of reporting and methodological flaws in lower limb arthroplasty observational studies.

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to http://dx.doi.org/10.1016/j.arth.2015.03.020.

Reprint requests: Joel J. Gagnier, ND, MSc, PhD, Departments of Orthopaedic Surgery and Epidemiology, University of Michigan, 24 Frank Lloyd Wright Drive, Ann Arbor, MI 48106–0391.

http://dx.doi.org/10.1016/j.arth.2015.03.020
0883-5403/

Reporting Quality Item

1. Are the inclusion and exclusion criteria clearly reported?
2. Is information regarding the number of patients who did not give informed consent or who were not willing to participate adequately reported?
3. Are the baseline characteristics of included patients reported?
4. Is the surgical technique adequately reported?
5. Are the prosthesis brand and fixation reported with enough detail?
6. Are the reasons or definitions for revision adequately reported?
7. Are the number of revisions (N) and revision rates regarding aseptic loosening (either Kaplan-Meier or life table or revisions per 100 observed component years) adequately reported?
8. Are the number of deaths, lost-to-follow-up (e.g., no show at clinic or emigration), amputations, and revisions other than the primary endpoint adequately reported?

Methodological Quality Item

1. Is there a clear primary research question/hypothesis?[a]
2. How were the cohorts constructed?
   a. Consecutively[b]
   b. Non-consecutively
   c. Unknown
3. How adequate was the follow-up (FU)?
   a. Fully completed FU
   b. 5% or less lost-to-FU or FU quotient[c] is 1 or less
   c. More than 5% lost-to-FU or FU quotient is more than 1
   d. Unknown
4. How was the FU performed?
   a. Predefined, e.g., yearly
   b. When patients had complaints or chart review (of non-predefined FU)
   c. Unknown
5. How many arthroplasties are at risk at the FU of interest?
   a. 20 or more
   b. Less than 20
   c. Unknown
6. Has a worst case analysis or competing risk analysis for competing endpoints been performed?

[a] In cases of aseptic loosening: Does the research question or hypothesis include revision of the component due to aseptic loosening?

[b] Consecutively is defined as all patients receiving an arthroplasty (TKA or THA) in a defined period of time have also received the arthroplasty of interest. The following situation is therefore non-consecutive: patients receiving prosthesis X while prosthesis Y has also been used for the same indication during the specified period.

[c] FU quotient = Number lost to follow-up/number of failures.

Fig. 1. The assessment of quality in lower limb arthroplasty (AQUILA) checklist.

Materials and Methods

Investigators identified the top seven journals in orthopedic surgery relative to their impact factor from the Thomson Reuters 2012 Journal Citation Reports [23]. Included journals were Acta Orthopaedica (AOS); BMC Musculoskeletal Disorders (BMC); Clinical Orthopaedics and Related Research (CORR); Journal of Bone and Joint Surgery (JBJS); Journal of Bone and Joint Surgery—British Volume (JBJS (Br)); Journal of Arthroplasty (JOA); and Knee Surgery, Sports Traumatology, Arthroscopy (KSSTA). Studies included in our analysis were case series and cohort studies including male and female patients undergoing THA or TKA for degenerative disease processes, including inflammatory conditions, and reporting outcomes related to arthroplasty revision for any reason. To reduce heterogeneity, studies were excluded from analysis if they included patients undergoing total hip or knee arthroplasty after acute traumatic injuries. Two individuals (JC and RM) with expertise in orthopedics searched all issues of all journals published in 2010 and 2011, separately and independently applying the inclusion and exclusion criteria above, and met to discuss the inclusions. A third party (JG or BP) was consulted to resolve disagreements. Next, the two investigators (JC and RM) independently and separately applied the AQUILA checklist to all included papers. All items were scored using the AQUILA methods [15]. The investigators then compared their results and any differences in the scoring of items were discussed and resolved by consensus. A third party (JG or BP), with expertise in both research methodology and reporting, was consulted to resolve disagreements. Regarding reproducibility, we calculated the interclass correlation coefficient (ICC) for inter-rater reproducibility of the two raters who had not previously used the AQUILA checklist and who were not involved with the development of the AQUILA. The ceiling effect for both the reporting and methodological score was also calculated and defined as the percentage of studies that reached the maximum score [24].

We calculated summed reporting quality item scores across articles of each journal and across journals. We also summarized the methodological quality items by summing the total number of ratings for each level of the categorical items within and across journals.

We performed linear regression with the reporting quality score as the dependent variable and year of publication, journal, and type of arthroplasty (THA versus TKA) as predictor variables. Significance was set at $P = 0.10$.

The reporting of this systematic review is in accordance with the PRISMA (preferred reporting items for systematic reviews and meta-analyses) statement [25].

Results

A total of 132 observational studies met the aforementioned inclusion criteria and were subsequently reviewed. Baseline descriptive statistics are found in Table 1. The average number of arthroplasties per study was 281 (median 101.5, SD = 728). Continents where included studies were conducted are found in Table 2.

Table 1
Baseline Descriptive Statistics for All Studies by Journal, Year, and Type of Arthroplasty.

		Total Number of Joints			% of Joints in Female Patients		% of Joints for Primary OA		Mean Age	
	Number of Studies	Mean	Median	SD	Mean	SD	Mean	SD	Mean	SD
Journal										
AOS	4	285	179	326	58	4	23	23	56.5	11
BMC	4	380	214.5	473	52	44	37	53	61.3	9.7
CORR	23	241	101	477	54	19	43	41	49.6	12.3
JBJS	18	588	101	1451	54	26	54	42	54	12.5
JBJS (Br)	29	347	129	909	59	20	56	37	55.4	14.2
JOA	49	149	98	183	62	21	55	41	58.2	10.4
KSSTA	5	187	181	161	82	10	83	13	68.2	3.8
All journals	132	281	101.5	728	59	22	53	39	55.9	12.3
Year										
2010	76	273	103	660	58	20	52	39	54.7	13.5
2011	56	292	100	817	61	25	54	41	57.6	10.3
Type of arthroplasty										
THA	97	249	100	597	55	22	42	38	52.6	11.9
TKA	35	370	133	1012	70	16	83	23	65.9	7.1

AOS, Acta Orthopaedica; BMC, BMC Musculoskeletal Disorders; CORR, Clinical Orthopaedics and Related Research; JBJS, Journal of Bone and Joint Surgery; JBJS Br, Journal of Bone and Joint Surgery—British Volume; JOA, Journal of Arthroplasty; KSSTA, Knee Surgery, Sports Traumatology, Arthroscopy.

Please cite this article as: Cowan JB, et al, Evaluation of Quality of Lower Limb Arthroplasty Observational Studies Using the AQUILA Checklist, J Arthroplasty (2015), http://dx.doi.org/10.1016/j.arth.2015.03.020

Table 2
Continents Where Studies Were Conducted by Journal, Year, and Type of Arthroplasty.

	Number of Studies	Continent of Study					
		North America	South America	Europe	Asia	Africa	Australia
Journal							
AOS	4	0	0	3	1	0	0
BMC	4	1	0	2	1	0	0
CORR	23	12	0	9	2	0	0
JBJS	18	7	0	9	2	0	0
JBJS (Br)	29	2	0	22	5	0	0
JOA	49	20	0	7	18	0	4
KSSTA	5	0	0	2	3	0	0
All journals	132	42	0	54	32	0	4
Year							
2010	76	23	0	36	14	0	3
2011	56	19	0	18	18	0	1
Type of arthroplasty							
THA	97	25	0	47	23	0	2
TKA	35	17	0	7	9	0	2

AOS, Acta Orthopaedica; BMC, BMC Musculoskeletal Disorders; CORR, Clinical Orthopaedics and Related Research; JBJS, Journal of Bone and Joint Surgery; JBJS Br, Journal of Bone and Joint Surgery—British Volume; JOA, Journal of Arthroplasty; KSSTA, Knee Surgery, Sports Traumatology, Arthroscopy.

The reproducibility for inter-observer agreement of the two raters was excellent for reporting quality, with an ICC of 0.87 (95% CI 0.82–0.91), and excellent for methodological quality, with an ICC of 0.95 (95% CI 0.93–0.97). The ceiling effect was 1.5% for the reporting score (2/132 studies reached the maximum score) and 0% for the methodological score (no papers reached the maximum score).

Table 3 includes the average reporting quality scores organized by journal, year, and type of arthroplasty. There was no significant difference in average reporting quality score among journals of publication ($P = 0.575$), although CORR articles had the highest average reporting quality score while BMC articles had the lowest average reporting quality score. There was no significant difference in average reporting quality score between years of publication within journals ($P = 0.539$) or between THA and TKA ($P = 0.301$).

Table 4 summarizes the number of studies that adequately reported each "reporting quality item" of the AQUILA checklist. Nearly 98% (123/126) of studies that addressed revision surgery adequately reported reasons or definitions for revisions. Ninety-five percent (125/132) of studies adequately reported prosthesis brand and fixation in enough detail to be replicated. However, only 3% (4/132) of studies included adequate information regarding the number of patients who did not give informed consent or were not willing to participate in the study.

A summary of the methodology ratings is found in Table 5. Most studies enrolled subjects consecutively (52%) and had predefined follow-up appointments (66%). Notably, only approximately 15% (20/132) of studies clearly stated a primary research question or hypothesis and only approximately 11% (15/132) of studies reported a worst-case analysis or competing risk analysis for their endpoints. Three studies (2.3%) had unknown cohort construction, unknown loss to follow-up, and unknown information about how follow-up was performed ("methodological quality items" 2–4). Lastly, 42% (55/132) of studies reported more than 5% of patients were lost to follow-up.

Discussion

The AQUILA checklist was created to assess the methodology, reporting, and generalizability of case series and cohort studies in the total hip and total knee arthroplasty literature. Using this checklist, we evaluated observational studies from 2010 and 2011 in seven orthopedic journals, finding no significant difference in reporting quality items among journals or between years of publication or type of arthroplasty. However, a number of salient findings were identified regarding reporting quality and methodology.

With respect to the quality of reporting, the majority of articles reviewed did not adequately report inclusion and exclusion criteria and very few articles reported the number of patients who did not

Table 3
AQUILA Checklist Mean Reporting Quality Scores.

	Number of Studies	Reporting Quality	
		Mean Score	Standard Deviation
Journal			
AOS	4	5.0	1.8
BMC	4	4.8	1.3
CORR	23	5.7	1.2
JBJS	18	5.6	1.0
JBJS (Br)	29	5.1	1.2
JOA	49	5.4	1.2
KSSTA	5	5.6	0.9
All journals	132	5.4	1.2
Year			
2010	76	5.4	1.2
2011	56	5.3	1.1
Type of arthroplasty			
THA	97	5.4	1.3
TKA	35	5.5	0.8

AOS, Acta Orthopaedica; BMC, BMC Musculoskeletal Disorders; CORR, Clinical Orthopaedics and Related Research; JBJS, Journal of Bone and Joint Surgery; JBJS Br, Journal of Bone and Joint Surgery—British Volume; JOA, Journal of Arthroplasty; KSSTA, Knee Surgery, Sports Traumatology, Arthroscopy.

Table 4
Number of Studies With Adequate Reporting Quality by Item.

Reporting Quality Items	Number of Studies with Adequate Reporting (%)
1. Are the inclusion and exclusion criteria clearly reported?	48 (36%)
2. Is information regarding the number of patients who did not give informed consent and who were not willing to participate adequately reported?	4 (3%)
3. Are the baseline characteristics of included patients reported?	117 (89%)
4. Is the surgical technique adequately reported?	94 (71%)
5. Are the prosthesis brand and fixation reported with enough detail?	125 (95%)
6. Are the reasons or definitions for revision adequately reported?	123 (98%)
7. Is the number of revisions and revision rates regarding aseptic loosening adequately reported?	91 (69%)
8. Is the number of deaths, lost-to-follow-up, amputations, and revisions other than the primary endpoint adequately reported?	110 (83%)

* This item was applicable to 126 of 132 studies.

Please cite this article as: Cowan JB, et al, Evaluation of Quality of Lower Limb Arthroplasty Observational Studies Using the AQUILA Checklist, J Arthroplasty (2015), http://dx.doi.org/10.1016/j.arth.2015.03.020

Table 5
Number of Studies With Adequate Methodological Quality by Items.

Methodological Quality Items	Number of Studies with Adequate Reporting (%)
1. Is there a clear primary research question/hypothesis?	20 (15%)
2. How were the cohorts constructed?	
A. Consecutively	A. 69 (52%)
B. Non-consecutively	B. 12 (9%)
C. Unknown	C. 51 (39%)
3. How adequate was the follow-up (FU)?	
A. Fully completed FU	A. 26 (20%)
B. ≤5% lost-to-FU or FU quotient is ≤1	B. 30 (23%)
C. >5% lost-to-FU or FU quotient is >1	C. 55 (42%)
D. Unknown	D. 21 (16%)
4. How was the follow-up performed?	
A. Predefined (e.g., yearly)	A. 87 (66%)
B. When patients had complaints or chart review (of non-predefined FU)	B. 15 (11%)
C. Unknown	C. 30 (23%)
5. How many arthroplasties are at risk at the FU of interest?	
A. ≥20	A. 124 (94%)
B. <20	B. 8 (6%)
C. Unknown	C. 0 (0%)
6. Has a worst-case analysis or competing risk analysis for competing endpoints been performed?	15 (11%)

give informed consent or were not willing to participate in the study. This prevents readers from determining the internal and external validity of these studies, as well as whether selection bias may have affected results. Furthermore, if such studies are subsequently included in future systematic reviews, the results from these poor quality studies may be perpetuated and assumed to be of higher quality than is correct. Regarding this issue of informed consent (AQUILA Reporting Quality Item #2), many of the observation studies we reviewed included patients who were enrolled or whose data were collected prior to The Health Insurance Portability and Accountability Act of 1996 (HIPPA). Prior to HIPPA, certain types of retrospective, observational, and medical records-based research did not require individual consent [26]. Therefore while only 3% of studies adequately addressed this checklist item, this may be at least partially due to differing standards for research prior to 1996.

Most studies adequately report patient baseline characteristics, surgical techniques, prosthesis brand and fixation, and the reasons or definitions for revision. The collection of baseline characteristics is crucial to avoid confounding. Reporting this information, along with details regarding techniques and implants, also helps readers assess study generalizability and applicability. Overall, out of 8 possible points, the mean reporting quality score for all articles reviewed was 5.4 (SD = 1.2). Subjective quality stratification (i.e., excellent, good, fair, poor) based on numerical score has not been established for the AQUILA checklist.

Regarding methodological quality, a surprisingly small number of reviewed studies stated a clear primary research question or hypothesis. As statistical power may differ with different hypotheses, without this information, the adequacy of a study's power or what the study is powered to detect cannot be fully assessed. In other words, one cannot critically appraise the internal validity or results of a study without first knowing what question that study is attempting to answer. Most studies also did not include a worst case analysis or competing risk analysis for competing endpoints. This is important as such an omission may affect the reliability and applicability of results. For example, ignoring competing risks may lead to biased estimations of the probability of revision surgery, as one study found the Kaplan–Meier estimator to overestimate the probability of revision surgery up to 60% in case of competing endpoints [27]. In general, poor methodology has been shown to incorrectly estimate treatment effect and leave studies vulnerable to biases that may have been otherwise avoidable [2,28,29].

The results of our study suggest that there are a number of shortcomings and great potential for improvement in the reporting quality and methodological quality of lower limb arthroplasty observational studies. Our findings are consistent with numerous studies in the orthopedic literature that have reported evidence of suboptimal study methodology and reporting [1,28,30–40]. Bhandari et al [34] found that among seventy-two randomized trials published in The Journal of Bone and Joint Surgery from 1988 through 2000, fewer than half demonstrated adequate randomization concealment, blinding of outcome assessors, or reporting of exclusion criteria. Another study showed that RCTs in the orthopedic trauma literature adhered to only 32% ± 29% of CONSORT checklist items and that 88% of studies did not include an adequate hypothesis or clinical objectives [33]. The CONSORT checklist was also used by Montane et al [39] to show that 75% of RCTs focusing on postoperative analgesia demonstrated poor quality reporting. Similarly, one study found that among RCTs in the top five orthopedic journals, only 49% of methodology criteria were fulfilled and 42% of criteria could not be assessed due to poor reporting [41]. Herman et al [35] found that only 35% of 274 randomized clinical trials employed the intention-to-treat principle, and that only approximately 16% adhered to the strict requirements of the principle. Karanicolas et al [36] found that less than 10% of 171 RCTs reported using blinded outcome assessors even though blinding may have been feasible by as many as 89–96% of non-patient-reported assessors. The authors also noted that a high number of studies inadequately reported allocation concealment, crossover of intervention groups, and issues regarding loss to follow-up. In even the most powerful types of studies–systematic reviews and meta-analyses–in the orthopedic journals with the greatest impact factors, both reporting quality and methodological quality were found to be poor [42].

Limitations of this study should be noted. The AQUILA checklist was created specifically to look at studies focusing on aseptic loosening and arthroplasty revision rates. In the present study, we use the checklist to assess arthroplasty observational studies looking at a variety of outcomes and endpoints, including revision for any reason (i.e., aseptic and septic). Other outcomes such as radiological failure and clinical failure (i.e., patient-reported outcome measures below certain value) are of equally great importance, but revision for aseptic loosening was chosen in the development of AQUILA because it highlights potential problems such as competing risks or worst case analysis that may not be as obvious for other outcomes [43]. Regardless of the outcome or mode of failure considered, the mechanisms of potential selection bias are universal [27,43,44]. Therefore it is paramount that the possibility of selection bias is appraised in various phases of a study: hypothesis generating and reporting, patient and cohort selection, follow-up, and statistical analysis. The AQUILA methodological quality items address the possibility of selection bias in each of these phases and add to previous studies that have already established how to measure and define radiological failure [45]. Regarding limitations due to the included studies themselves, the mean and median numbers of arthroplasties per study were 281 and 101.5, respectively. This difference between the mean and the median is due to the six largest studies (each of which had between 1055 and 6070 arthroplasties) skewing the median number of arthroplasties. We also assessed observational studies published in only two calendar years (2010 and 2011), thus our findings cannot be generalized across other years. Nevertheless, the results provide a benchmark for contemporary reporting and methodological quality in orthopedic cohorts of lower limb arthroplasty.

This study also has several notable strengths. A large number of studies across a variety of journals were included in this analysis. Our authors include individuals with advanced training in orthopedic surgery and statistics, and affiliation with an epidemiology department, the latter of which has been associated with higher quality studies [34]. This study also assessed both reporting quality and methodological quality, both of which are crucial for a thorough evaluation of any study. The reproducibility of the AQUILA scores was excellent and the ceiling

Please cite this article as: Cowan JB, et al, Evaluation of Quality of Lower Limb Arthroplasty Observational Studies Using the AQUILA Checklist, J Arthroplasty (2015), http://dx.doi.org/10.1016/j.arth.2015.03.020

effect was very low. For comparison the Harris hip score has a ceiling effect of more than 50% [46]. These values indicate that reporting and methodological quality in observational orthopedic studies can be substantially improved and underscore the validity of the AQUILA in measuring observational study quality for lower extremity arthroplasty.

Other than the AQUILA checklist, we are not aware of a checklist that addresses the methodological quality of observational studies in arthroplasty studies. Other reporting tools such as STROBE are useful for assessing the reporting quality of observational studies, but do not address methodological quality. A systematic review by Sanderson et al [47] identified numerous tools for assessing methodology in observational studies. This review found that most tools address selection methods (93%), measurement of variables (86%), sources of bias (86%), confounding (78%), and statistics (78%); however, unlike the AQUILA checklist, these tools were not specifically designed for studies related to THA and TKA.

Lastly, while neither a weakness nor strength, it is important to consider that different results or conclusions may have been obtained from evaluators of varying levels of expertise. For example, general practice orthopedic surgeons and fellowship-trained arthroplasty surgeons may have differing opinions about whether a given item (surgical technique, for example) is "adequately reported" according to the AQUILA checklist. In other words, a surgeon with greater arthroplasty experience or training may find a less-detailed description to be "adequate" as compared with a surgeon with relatively less experience. These hypotheses remain to be tested.

Our results suggest that there is significant room for improvement in the reporting and methodology of total hip and knee arthroplasty observational studies. Authors are encouraged to seek out and adhere to guidelines for reporting and methodological quality. Journal editors are encouraged to promote such guidelines, including articulation of a clear primary research question or hypothesis, in their instructions to authors. Journal editors and peer reviewers are ultimately responsible for publication decisions and should hold authors to the highest standards of reporting and methodological quality. As observational studies have a significant impact on the clinical decision making of orthopedic surgeons, it is critical that they contain clear and thorough descriptions such that their results and conclusions can be considered valid and applicable to decision making.

References

1. Balasubramanian SP, Wiener M, Alshameeri Z, et al. Standards of reporting of randomized controlled trials in general surgery: can we do better? Ann Surg 2006; 244(5):663.
2. Schultz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273(5):408.
3. Bhandari M, Morrow F, Kulkarni AV, et al. Meta-analyses in orthopaedic surgery: a systematic review of their methodologies. J Bone Joint Surg Am 2001;83-A(1):15.
4. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609.
5. Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? J Bone Joint Surg Am 2007; 89(3):550.
6. Sharma R, Vannabouathong C, Bains S, et al. Meta-analyses in joint arthroplasty: a review of quantity, quality, and impact. J Bone Joint Surg Am 2011;93(24):2304.
7. Dijkman BG, Abouali JA, Kooistra BW, et al. Twenty years of meta-analyses in orthopaedic surgery: has quality kept up with quantity? J Bone Joint Surg Am 2010;92(1):48.
8. Cowan J, Lozano-Calderon S, Ring D. Quality of prospective controlled randomized trials: analysis of trials of treatment for lateral epicondylitis as an example. J Bone Joint Surg Am 2007;89-A(8):1693.
9. Poolman RW, Struijs PA, Krips R, et al. Does a "Level I evidence" rating imply high quality of reporting in orthopaedic randomised controlled trials? BMC Med Res Methodol 2006;6:44.
10. Soucacos PN, Johnson EO, Babis G. Randomised controlled trials in orthopaedic surgery and traumatology: overview of parameters and pitfalls. Injury 2008;39(6):636.
11. Mundi R, Chaudhry H, Singh I, et al. Checklists to improve the quality of the orthopaedic literature. Indian J Orthop 2008;42(2):150.
12. Schemitsch EH, Bhandari M, McKee MD, et al. Orthopaedic surgeons: artists or scientists? J Bone Joint Surg Am 2009;91(5):1264.
13. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453.
14. da Costa BR, Cevallos M, Altman DG, et al. Uses and misuses of the STROBE statement: bibliographic study. BMJ Open 2011;1(1):e000048.
15. Pijls BG, Dekkers OM, Middeldorp S, et al. AQUILA: assessment of quality in lower limb arthroplasty. An expert Delphi consensus for total knee and total hip arthroplasty. BMC Musculoskelet Disord 2011;12:173.
16. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995;311(7001):376.
17. Keurentjes JC, Van Tol FR, Fiocco M, et al. Minimal clinically important differences in health-related quality of life after total hip or knee replacement: a systematic review. Bone Joint Res 2012;1(5):71.
18. Nouta KA, Verra WC, Pijls BG, et al. All-polyethylene tibial components are equal to metal-backed components: systematic review and meta-regression. Clin Orthop Relat Res 2012;470(12):3549.
19. Pijls BG, Nieuwenhuijse MJ, Fiocco M, et al. Early proximal migration of cups is associated with late revision in THA: a systematic review and meta-analysis of 26 RSA studies and 49 survival studies. Acta Orthop 2012;83(6):583.
20. Pijls BG, Valstar ER, Kaptein BL, et al. Differences in long-term fixation between mobile-bearing and fixed-bearing knee prostheses at ten to 12 years' follow-up: a single-blinded randomised controlled radiostereometric trial. J Bone Joint Surg (Br) 2012;94(10):1366.
21. Pijls BG, Valstar ER, Nouta KA, et al. Early migration of tibial components is associated with late revision: a systematic review and meta-analysis of 21,000 knee arthroplasties. Acta Orthop 2012;83(6):614.
22. van der Voort P, Pijls BG, Nouta KA, et al. A systematic review and meta-regression of mobile-bearing versus fixed-bearing total knee replacement in 41 studies. Bone Joint J 2013;95-B(9):1209.
23. 2012 Journal Citation Reports®. New York: Thomson Reuters; 2013.
24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4(4):293.
25. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7).
26. Wilson JF. Health Insurance Portability and Accountability Act Privacy rule causes ongoing concerns among clinicians and researchers. Ann Intern Med 2006;145(4):313.
27. Keurentjes JC, Fiocco M, Schreurs BW, et al. Revision surgery is overestimated in hip replacement. Bone Joint Res 2012;1(10):258.
28. Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. J Bone Joint Surg Am 2007;89(9):1970.
29. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054.
30. Devereaux PJ, Manns BJ, Ghali WA, et al. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the consolidated standards of reporting trials (CONSORT) checklist. Control Clin Trials 2002;23(4):380.
31. Goldhahn S, Sawaguchi T, Audige L, et al. Complication reporting in orthopaedic trials. A systematic review of randomized controlled trials. J Bone Joint Surg Am 2009; 91(8):1847.
32. Dulai SK, Slobogean BL, Beauchamp RD, et al. A quality assessment of randomized clinical trials in pediatric orthopaedics. J Pediatr Orthop 2007;27:573.
33. Bhandari M, Guyatt GH, Lochner H, et al. Application of the consolidated standards of reporting trials (CONSORT) in the fracture care literature. J Bone Joint Surg Am 2002; 84-A(3):485.
34. Bhandari M, Richards R, Sprague S, et al. The quality of reporting of randomized trials in The Journal of Bone and Joint Surgery from 1988 through 2000. J Bone Joint Surg Am 2002;84-A(3):388.
35. Herman A, Botser IB, Tenenbaum S, et al. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. J Bone Joint Surg Am 2009;91(9):2137.
36. Karanicolas PJ, Bhandari M, Taromi B, et al. Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. J Bone Joint Surg Am 2008;90(5):1026.
37. Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg [Am] 2004;29(4):727.
38. Chaudhry H, Mundi R, Singh I, et al. How good is the orthopaedic literature? Indian J Orthop 2008;42(2):144.
39. Montane E, Vallano A, Vidal X, et al. Reporting randomised clinical trials of analgesics after traumatic or orthopaedic surgery is inadequate: a systematic review. BMC Clin Pharmacol 2010;10:2.
40. Somerson JS, Bhandari M, Vaughan CT, et al. Lack of diversity in orthopaedic trials conducted in the United States. J Bone Joint Surg Am 2014;96(7):e56.
41. Chess LE, Gagnier J. Risk of bias of randomized controlled trials published in orthopaedic journals. BMC Med Res Methodol 2013;13:76.
42. Gagnier JJ, Kellam PJ. Reporting and methodological quality of systematic reviews in the orthopaedic literature. J Bone Joint Surg Am 2013;95(11):e771.
43. Nelissen RG, Brand R, Rozing PM. Survivorship analysis in total condylar knee arthroplasty. A statistical review. J Bone Joint Surg Am 1992;74(3):383.
44. Nouta KA, Pijls BG, Fiocco M, et al. How to deal with lost to follow-up in total knee arthroplasty: a new method based on the competing risks approach. Int Orthop 2014;38(5):953.
45. Ewald FC. The Knee Society total knee arthroplasty roentgenographic evaluation and scoring system. Clin Orthop Relat Res 1989;248:9.
46. Wamper KE, Sierevelt IN, Poolman RW, et al. The Harris hip score: do ceiling effects limit its usefulness in orthopedics? Acta Orthop 2010;81(6):703.
47. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007;36(3):666.

Please cite this article as: Cowan JB, et al, Evaluation of Quality of Lower Limb Arthroplasty Observational Studies Using the AQUILA Checklist, J Arthroplasty (2015), http://dx.doi.org/10.1016/j.arth.2015.03.020