

Cochrane's Shoulder-and-Elbow Systematic Reviews Issues: Mistrustful Impact of Evidence

Jose Carlos Garcia Jr¹, Mario H L Gonçalves¹, Eduardo F. Cordeiro¹, Luiz F S Fink¹,
Alvaro M Cardoso Jr¹, Mauricio P Raffaelli¹, Marcelo Boulos Dumans e Mello¹

Abstract

Introduction: Evidence-based guidelines are important for Medicine. However they have repeatedly failed in presenting conclusive evidence in surgical field. The reason is limited availability of well-designed scientific studies.

In order to investigate limitations and their impact on Shoulder and Elbow Surgery field, we assessed data of Systematic Reviews (SRs) from the Cochrane Library database for risk of bias.

Analysis of the 35 SRs assessed presented the following risk of bias: Comparable outcomes within SRs studies – 42.86%; Blinding – 40.80%; Sample size calculation – 7.43%; Missing values handling – 16.14%; Adequate statistical analysis – 11.74%; Monitoring – 0.37%; and Meta-analysis data heterogeneity – 40.54%, showing that these papers have low methodological quality.

This study may call attention of Shoulder and Elbow surgeons to aim in quality improvement, data comparability and reliability of primary trials and SRs.

Future conjoint effort in defining and diffusing feasible standards for surgical trials must be made by local and international societies of surgical specialties.

Key Words: Shoulder, Elbow, Systematic Review, Methodological Quality.

Introduction:

Evidence-based medicine is the source for the highest level of evidence and reliability, driving to better decision-making in medical practice [25].

Despite all advantages of this method, one must be aware of the difficulty in accessing nonpublished Clinical Trials or Clinical Trials published in languages other than English, which are common limitations related to all Systematic Reviews (SRs) [29]. Clinical Trials whose statistical significance is not reached are frequently considered 'second-class' papers, being disregarded by clinical researchers and not accepted for publication by reviewers [12]. The absence of inclusion of these data into the SRs and meta-analysis leads to biased interpretations and conclusions (40).

The way SRs are conducted and how results

are extracted from primary studies vary among different areas of medicine. Each different area has its particularities and limitations for research. It is very clear when we compare clinical and surgical trials.

Following are the main particularities that preclude surgical field research to be conducted in similar way to clinical ones: [1] Impossibility of double blinding – some trials compare techniques that do not allow true blinding. For example, scars inherent to open surgery technique cannot be hidden when comparing it to an endoscopic procedure. [2] Learning curve – unlike of ministering manufactured pills, surgical performance will depend on surgeon's experience and it surely affect outcomes [33,10,3] Technological evolution of implants and devices – over time history showed us improvement in outcomes of

endoscopic procedures over open ones due to progress of devices [27]. Also it is not adequate to compare procedures done with different generation of implants. [4] Technique evolution – modification of some steps within the same technique can also impact learning curve and results. These and other characteristics make surgical research much more complex than those within clinical trials outside the surgical field.

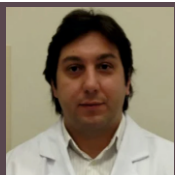
Recently Orthopaedic Surgery evidence-based medicine has promptly grown in importance after publication of some pivot Randomised Controlled Clinical Trials (RCTs) [24]. However, most SRs failed to present enough evidence to establish a reliable Evidence-Based Practice Guideline because strict selective criteria for inclusion in SRs limited the number of RCTs in the

meta-analysis. Methodological limitations, such as improper randomization (400, absence of Intention-to-Treat analysis [23], poor selection criteria [22] and lack of standardization of outcomes are pointed

¹NAEON-Santa Catarina Hospital

Address of Correspondence

Dr. Jose Carlos Garcia Jr., MD, MSc, PhD
NAEON-Hospital Santa Catarina-SP-
Brazil
Email: jose.cjunior@hsl.org.br



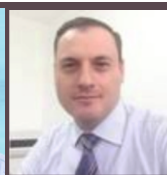
Dr. Jose Carlos Garcia Jr.



Dr. Mário Henrique
Lobão Gonçalves



Dr. Alvaro M
Cardoso Jr



Dr. Mauricio P
Raffaelli



Dr. Marcelo Boulos

© 2016 by Acta of Shoulder and Elbow Surgery | Available on www.asesjournal.com | doi:10.13107/ases.2454-5473.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1: Results of Cochrane's Methodological Quality Assessment tools for clinical SRs

Clinical Methodological Quality Assessment	By Review	By Paper
Randomization	Low Risk	Low Risk
	11.70/20 (58.50%)	111/175 (63.43%)
	Unclear	Unclear
	5.58/20 (27.90%)	24/175 (13.71%)
Allocation Concealment	Low Risk	Low Risk
	7.48/20 (37.40%)	51/175 (29.14%)
	Unclear	Unclear
	7.85/20 (39.25%)	54/175 (30.86%)
Balance at Baseline	Low Risk	Low Risk
	9.10/20 (45.50%)	90/175 (51.43%)
	Unclear	Unclear
	5.67/20 (28.35%)	27/175 (15.43%)
Provider, Physical Therapy Blinded	Low Risk	Low Risk
	3.93/20 (19.65%)	33/175 (18.86%)
	Unclear	Unclear
	0.11/20 (0.55%)	1/175 (0.57%)
Co-interventions Avoided	Low Risk	Low Risk
	14.14/20 (70.70%)	102/175 (58.28%)
	Unclear	Unclear
	0.64/20 (3.20%)	21/175 (12.01%)
Adherence Acceptable	Low Risk	Low Risk
	16.72/20 (83.60%)	132/175 (75.43%)
	Unclear	Unclear
	2.53/20 (12.65%)	31/175 (17.71%)
Patient Blinded	Low Risk	Low Risk
	7.75/20 (38.75%)	84/175 (48.00%)
	Unclear	Unclear
	1.40/20 (7.00%)	14/175 (8.00%)
Withdrawn Descript and Acceptable	Low Risk	Low Risk
	11.89/20 (59.30%)	109/175 (62.28%)
	Unclear	Unclear
	1.70/20 (8.50%)	19/175 (10.86%)
Outcome Assessor Blinded	Low Risk	Low Risk
	10.67/20 (53.35%)	102/175 (58.28%)
	Unclear	Unclear
	2.24/20 (11.20%)	20/175 (11.43%)
Timing of Assessment Comparable	Low Risk	Low Risk
	17.76/20 (88.80%)	129/175 (73.71%)
	Unclear	Unclear
	1.19/20 (5.95%)	14/175 (8.00%)
Intentio-to-Treat	Low Risk	Low Risk
	5.28/20 (26.40%)	48/175 (27.43%)
	Unclear	Unclear
	1.03/20 (5.15%)	7/175 (4.00%)

Table 2: Results of Cochrane's Methodological Quality Assessment tools for surgical SRs

Surgical Methodological Quality Assessment	By Review	By Paper
Randomization	Low Risk	Low Risk
	9.17/15 (47.93%)	44/93 (47.31%)
	Unclear	Unclear
	5.41/15 (36.07%)	40/93 (43.01%)
Allocation Concealment	Low Risk	Low Risk
	4.36/15 (29.07%)	27/93 (29.03%)
	Unclear	Unclear
	8.18/15 (54.53%)	51/93 (54.84%)
Blinding Outcome Assessor	Low Risk	Low Risk
	2.37/15 (15.80%)	13/93 (13.98%)
	Unclear	Unclear
	2.37/15 (15.80%)	19/93 (20.43%)
Incomplete Outcome Data	Low Risk	Low Risk
	8.76/15 (58.40%)	50/93 (53.76%)
	Unclear	Unclear
	2.56/15 (17.07%)	19/93 (20.43%)
Selective Reporting	Low Risk	Low Risk
	3.16/15 (16.07%)	14/93 (15.05%)
	Unclear	Unclear
	9.73/15 (64.87%)	63/93 (67.74%)
Balance at Baseline	Low Risk	Low Risk
	4.04/15 (26.93%)	23/93 (24.73%)
	Unclear	Unclear
	5.97/15 (39.80%)	43/93 (46.24%)
Free of Performance Bias	Low Risk	Low Risk
	5.05/15 (33.67%)	35/93 (37.63%)
	Unclear	Unclear
	6.53/15 (43.53%)	42/93 (45.16%)

out as main causes of failure of Orthopedics' and its subspecialties' Srs [30,41].

Better and deeper understanding of these limitations may be key to improve both primary research and consequently SRs in orthopaedic surgery. Thus, better decisions based on the best evidence will be able to be made.

Methods

Study Design

This Systematic Review assesses Cochrane's

Systematic Reviews of Shoulder and Elbow. The authors are orthopaedic surgeons familiar with this field of knowledge. So Shoulder-and-Elbow subspecialty will represent orthopaedic surgery addressed in this study, following Descartes' principle that "all complex matters need to be divided into many simpler parts in order to better focus on specific problems, thus creating greater possibilities of understanding and solving them".

As Cochrane Library is known as one of the

most complete, important and reliable sources of Srs [13], a review of all Cochrane SRs related to shoulder-and-elbow subspecialty was conducted to better understand limitations and characteristics of RCTs and SRs related to orthopedics.

Search Strategy

Two strategies were used: (1) Cochrane website's Search Tool – Medical Subject Headings (MeSH) terms searched: shoulder, elbow, scapula, humerus, cuff,

Table 3: Complementary Assessment analysed 9 topics looking for potential bias not considered on Cochrane's Methodological Quality Assessment

	Total by Review	Total by Paper	Clinical by Review	Clinical by Paper	Surgical by Review	Surgical by Paper
Heterogeneity less 40%	14.17/35 (40.48%) NA 5/35 (14.28%)	102/158 (64.56%)	6.92/20 (34.60%)	40/70 (57.14%)	7.25/15 (48.33%) NA 5/15 (33.33%)	62/88 (70.45%)
Outcomes comparable ?50%	15/35 (42.86%)	NA	7/20 (35.00%)	NA	8/15 (53.33%)	NA
Target population	34.16/35 (97.60%)	265/268 (98.88%)	19.75/20 (98.75%)	174/175 (99.43%)	14.41/15 (96.07%)	91/93 (97.85%)
Blinding	Low Risk 14.28/35 (40.80%) Unclear 3.33/35 (9.51%) NA 1/35 (2.86%)	Low Risk 125/268 (46.64%) Unclear 35/268 (13.06%) NA 4/268 (1.49%)	Low Risk 10.86/20 (54.30%) Unclear 0.96/20 (4.90%) NA 1/20 (5.00%)	Low Risk 107/175 (61.14%) Unclear 17/175 (9.71%) NA 4/175 (2.28%)	Low Risk 3.42/15 (22.80%) Unclear 2.37/15 (15.80%)	Low Risk 18/93 (19.35%) Unclear 18/93 (20.43%)
Sample size calculation	2.60/35 (7.42%)	14/268 (5.22%)	1.85/20 (9.25%)	11/175 (6.28%)	0.75/15 (5.00%)	3/93 (3.22%)
Adequate statistical analysis	4.11/35 (11.74%) Unclear/NR 28.46/35 (81.31%)	57/268 (21.27%) Unclear/NR 185/268 (69.30%)	4.11/20 (20.55%) Unclear/NR 13.93/20 (69.65%)	7/175 (32.57%) Unclear/NR 95/175 (54.23%)	0/15 (0.00%) Unclear/NR 14.53/15 (96.85%)	0/93 (0.00%) Unclear/NR 90/93 (96.77%)
Missing values handling, ITT	5.65/35 (16.14%) Unclear 1.53/20 (4.37%)	48/268 (17.91%) Unclear 10/268 (3.73%)	4.68/20 (23.40%) Unclear 0.83/20 (4.15%)	43/175 (24.57%) Unclear 6/175 (3.43%)	0.97/15 (6.47%) Unclear 0.7/15 (4.67%)	5/93 (5.38%) Unclear 4/93 (4.30%)
Data Entry	0/35 (0.00%) NR 35/35 (100%)	0/268 (0.00%) NR 268/268 (100%)	0/20 (0.00%) NR 20/20 (100%)	0/175 (0.00%) NR 175/175 (100%)	0/15 (0.00%) NR 15/15 (100%)	0/93 (0.00%) NR 93/93 (100%)
Monitoring	0.13/35 (0.37%)	2/268 (0.75%)	0.13/20 (0.65%)	2/175 (1.14%)	0/15 (0.00%)	0/93 (0.00%)

Table 4: Papers and SRs without meta-analysis were withdrawn and stratified by reason for exclusion

Withdrawn Cause	Total	Clinical	Surgical	Without Meta-Analysis
Randomization	183	82	74	27
Abandoned the Clinical Trial	14	4	10	0
Unacceptable Selection Bias	32	24	7	1
Incorrect Intervention	61	30	24	7
Incorrect/insufficient Enrollment	156	94	45	17
Ethics Issues	1	0	1	0
Did Not Begin/Ongoing	8	2	5	1
Retrospective/Observational/Case Report	154	109	15	30
Same Intervention Both Groups	5	3	1	1
Missing Data	3	2	1	0
Inadequate Follow up	15	8	7	0
No Specified Results	63	51	2	10
Review	23	18	5	0
Intervention Not Compared with Placebo	4	2	0	2
Language	2	1	1	0
TOTAL	724	430	198	96

ulna, epicondylitis and dislocation. The SRs that addressed the subspecialty of Shoulder-and-Elbow Surgery were chosen by assessing titles. (2) Direct search in Cochrane Bone, Joint and Muscle Trauma Group.

Thereafter three senior specialized in should-and-elbow orthopaedic surgeons reviewed all abstracts identified and consensually withdrew duplicated papers, the ones that are not related to shoulder-and-elbow, studies that were just protocols,

studies that were split into others and outdated ones.

As inclusion phase is from utmost importance, five senior surgeons analysed all papers and consensually withdrew studies not related to at least one of the following topics: physical exam, clinical or surgical therapeutics of shoulder-and-elbow subspecialty.

Data Extraction

All data were extracted as consensus by

three authors that assessed methodology of the selected SRs based on Cochrane's Methodological Quality Assessment Tool [8,14,5,20,19] and Complementary Assessment of RCTs [40,41,43,15,39,11].

Measurement Tools

The Cochrane Musculoskeletal Group grades SRs' evidence as: Platinum, Gold, Silver or Bronze levels, according to Cochrane's Methodological Quality Assessment. This has two different bias assessment tools:

Surgical and Clinical, each specific for the nature of the SRs analysed. They comprise a judgement and support for each entry in a 'Risk of bias' table, where each entry addresses a specific feature of the study as low risk, high risk or unclear risk. For trials of clinical treatments, an eleven-item risk of bias assessment was used: randomization, allocation concealment, balance in baseline, blinding provider, co-interventions avoided, adherence acceptable, patient blinded, acceptable withdrawn, outcome assessor blind, timing of assessment comparable and Intention-to-Treat analysis (8,14,5,20,19). For surgical treatment trials, a seven-item risk of bias was used assessing: randomization, allocation concealment, functional blinding, incomplete outcome (losses >20%), selective reporting, base-line balance and performance bias (8). The Complementary Assessment of RCTs assesses other important points not considered by Cochrane's tools. It was developed using important issues from CONSORT11, COMET32 and Cochrane Handbook44 related to quality standards8. For each issue assessed there were five possible answer categories: yes, no, unclear, not-applicable (NA) and not-reported (NR). There are three sections: (1) Issues related to SRs – Less than 40% heterogeneity between primary trials within the review, which means low to moderate heterogeneity8; (2) Choice of Outcomes – if study present at least one outcome as 'standard' comparable at more than 50% of primary papers included in the review; (3) Issues related to primary studies in SRs – Choice of Target Population, Blinding, Sample Size (reduce type 2 error), Adequate Statistical Analysis (assessment of curves characteristicsto understand if data is parametric or non-parametric, presence of standard deviations and standard errors, etc), Missing Values Handling (Intention-to-Treat or Per-Protocol), Data Entry Cautions (reading aloud, single data entry with cross-checks or double data entry) and Data Monitoring. Since clinical and surgical treatments present different patterns of assessment, the authors chose to stratify analysis through these two types of treatment to better understand differences between them. SRs that do not present any RCT were not

removed. Instead they were assessed separately to better understand causes of withdrawal of primary papers.

Data Analysis/ Procedures

Variance with standard error, standard deviation and confidence intervals were not logical analysis for this review, once the entire population of Cochrane's Shoulder-and-Elbow SRs was addressed. Statistical methods provide estimations with respect to a certain degree of uncertainty and results reflect all of the population of the SRs. Thus, this SR shows the ultimate results based on the whole data. Data for Cochrane Musculoskeletal Group's primary trials were presented only descriptively. Results for Cochrane's Methodological Quality Assessment and Complementary Assessment of RCTs were presented in frequency through proportions and percentagens of risk of bias for each item stratified into subgroups. Papers withdrew from SRs, including SRs without meta-analysis, were analysed for most common reasons for withdrawal and data presented in frequency by absolute numbers. Results were stratified into the following groups: all SRs, Clinical SRs, Surgical SRs and SRs without meta-analysis.

Ethics

As this study comprises a secondary data analysis, University of Liverpool Ethics Committee, Santa Catarina Hospital and the Brazilian Government granted approval based on the fact that there was no risk to human subjects.

Results

Thirty-eight SRs were included, 17 of them are surgical and 21 clinical. Of all surgical SRs, 2 did not present RCTs or quasi-randomised trials and therefore did not present meta-analysis. On the clinical group, one SR did not present meta-anaylisis. SRs were assessed in 3 groups: whole 38 SRs, Clirical (20 SRs) and Surgical (15 Srs). According to Cochrane Musculoskeletal Group grade for primary trial, including SRs without meta-analysis, of all 38 SRs selected only one paper reached platinum level of evidence. Three SRs reached silver level, one with high risk of bias due to

heterogeneous intervention and two low risk of bias. Three other SRs presented very limited evidence. Due to lack of any proper evidence others SRs could not be graded by Cochrane Musculoskeletal.

Results of Cochrane's Methodological Quality Assessment tools for clinical and surgical SRs are respectively shown on TABLES 1 and 2.

Complementary Assessment developed by the authors analysed 9 topics looking for potential bias not considered on Cochrane's Methodological Quality Assessment.

Results are shown on TABLE 3.

Target Population topic assesses if patients were suitable for the proposed intervention and almost all SRs presented adequate target population. But 3 primary papers were conducted among militaries and were considered as potential bias.

Heterogeneity <40% and Comparable Outcomes were found in round 40% of the Srs.

Blinding, as expected, showed a great difference between clinical (54,30%) and surgical (19,35%) papers.

Sample Size Calculation, Statistical Analysis and Missing Values were presented in less than 20% of all SRs, what made us very concerned about reliability of results.

Data Entry seems to be well difunded among SRs authors. All SRs analysed had reported it. However, none Primary Paper had it.

Monitoring was the worst aspect taken into consideration by researchers. Only one SR (0,37%) had adequate monitoring and it was a clinical SR.

Papers and SRs without meta-analysis were withdrawn and results are shown on Table 4, stratified by reason for exclusion.

Discussion

Cochrane's Systematic Reviews are known by high standard and reliability. Many physicians have been using them as guide for the best evidence possible in certain fields of medicine.

Despite they have proven to be effective to summarise results of trials in the Clinical field, they lack effectiveness in many Surgery trials. A commom complaint among surgeons is that SRs have not been able to provide reliable guidelines. Conclusions are

repetitively similar in most of SRs: “There is limited evidence available from RCTs on the relative effectiveness of [a certain surgical treatment]”.

Considering SRs analysed in this article were conducted in conformance with the high standards of Cochrane Collaboration, the absence of adequate evidence must be due to failure within primary papers and withdrawn ones. Of all reviewed SRs, we found a low rate of conclusiveness. Only one article reached the Platinum level of evidence according to Cochrane, showing that Shoulder-and-Elbow SRs have often failed to present enough evidence to establish reliable evidence-based practice guidelines [41].

Some authors stated that “absence of evidence is not evidence of absence” [1]. Therefore, the main cause of this lack of evidence may just be absence of data for establishing reliable results. The number of individuals necessary to suitably answer a research question may be more than those enrolled in the SR, making results not truly reliable. So, larger samples will lead trials to achieve narrower confidence intervals and high statistical power (31).

SRs may be guides to determine whether more studies are necessary or not to suitably answer a research question, as well as to determine what direction researchers need to follow in future trials, like sample sizing and research harms [9].

Deep analysis of the only Platinum-level SR found in this study showed: inclusion of a Silver-level primary study, only 50% of included papers had Intention-to-Treat, only 40% performed Sample Size Calculation and less than half of outcomes were comparable. So, despite having 90% of blinded primary papers and 70% of adequate statistical analysis, there is a lack of methodological quality that precludes adequate use of this data for meta-analysis. Less than 50% of SRs presenting at least one comparable outcome indicate an immediate necessity of discussion and standardization of outcomes among Shoulder and Elbow researchers. The Rowe scale³⁷ was the most ‘popular’ (50%) specific outcome for shoulder instability within Cochrane’s SRs probably because it is simple, fast and reproducible.

For rotator cuff interventions we found 3 major outcome assessment scores: UCLA

[21] (5 / 14) – the oldest and most used; Constant [35] (3 / 14) – mostly used in Europe; and ASES [28] (1 / 14) – mostly used in North America. Perhaps if all future primary papers could standardize the use of all three outcome-scores in a definite timeframe, papers would certainly be more comparable all around the world.

There is over emphasis in effectiveness within the SRs in expense of many times omitting safety information [16]. Indeed, medical decision follows a Paraconsistent logic pattern where positive and negative aspects are weighted to reach the best decision for each individual [3]. Thus, insufficient disclosure of safety aspects within a SR can overestimate effectiveness, driving to biased decisions regarding treatment.

Simple methodological issues like Blinding presented remarkable antagonism within clinical and surgical fields of the same subspecialty. Reliable blinding was found in only 19.35% surgical RCTs in contrast to 61.14% clinical RCTs. Surgical approach, post-surgical exams and ethics seems to be the main reasons why blinding is more feasible for Clinical Trials than Surgical ones. As Double-Blinding is extremely difficult to achieve in Surgical Trials, it is almost impossible for them to reach Cochrane’s Platinum level of evidence. In this way, Cochrane Musculoskeletal Group proper grading for primary trials does not seem suitable for surgical trials.

Sample Size Calculation is key factor to avoid Type II error but it is present in only 5.22% of primary papers analysed. This is an important reason why SRs lack statistical power and reliability [36] and it must be highlighted in congresses and discussions in order to avoid biased conclusions of primary papers.

Another important point to be considered in trials is data loss. There is tendency towards a higher drop out rate amongst patients who failed to respond to treatment² resulting in bias towards positive results⁶. Intention-to-Treat (ITT) is a good tool to avoid this type of bias. However in this study, the major part of the entire assessed primary papers that used ITT did not even mention the type of ITT technique used. Using Per-Protocol can estimate effectiveness to be overly optimistic, driving researchers to biased conclusions (2).

A reliable statistical analysis is also very important for achieving reliability. Different statistical approaches may be necessary in different conditions, depending on sample size, distribution of data and many other factors [4]. Many primary papers present results without mentioning data distribution verification and insufficient detail about statistical methods used. Only 18.69% of all SRs and 30.70% of primary papers presented analyses of statistical methodology. This must alert investigators about potentially biased results and misleading conclusions [38].

Current high standards for Clinical Trials require data entry to be scrutinized for reliability. Strategies include reading aloud [18], single data entry with cross check [42] and double data entry [7]. All SRs assessed showed strategies used, but no primary paper reported data entry strategies. In fact, the whole set of primary papers never mentioned it.

Good Clinical practices have standardised all legal and ethical steps to be followed in order to conduct a Clinical Trial. However, monitoring is main concern within this standardization [17]. Monitoring improves reliability by checking all steps within study. Only two primary clinical papers cared about this aspect in our study. A possible solution for improving monitoring could be subspecialty societies or international collaborations to assume this role. Heterogeneity is key factor within a meta-analysis. It shows how different results are among trials. So, the more heterogeneous results are, the more incompatible they will be, reflecting difficulty in assessing outcomes. In this review, 40.54% of SRs presented moderate to low rate of heterogeneity, suggesting better standardization of procedures, interventions, and/or assessments like the ones proposed by the IDEAL-Collaboration [34].

Randomization is ‘sine qua non’ for Primary Papers to provide data to meta-analysis within a Cochrane SR. But if not well done they can be under risk of bias. Quasi-randomization can be considered as high risk of bias. From all Shoulder-and-Elbow SRs analysed, randomization seemed not to be a concern at first look, as 63.43% of RCTs were considered low risk, but when stratified only 36.07% of surgical trials had

adequate randomization against 75.85% on clinical ones. It clearly shows another fragility of Surgical Papers.

Allocation Concealment is important to avoid selection bias and protect randomization. Only 29.14% of clinical RCTs and 29.03% of surgical trials presented low risk of bias allocation concealment. These rates are very low and require significant improvement.

Balance at Baseline ensures homogeneity of individuals among groups researched. If imbalance is found, all results might be compromised. The baseline of surgical trials was balanced (low risk of bias) in only 24.73%, while for clinical trials more than double (51.43%).

Only 58.28% of RCTs were free of co-interventions in Clinical Trials. This measure was not performed for Surgical Trials. This data exposes an alarming situation wherein confounding factors can be present in 41.72% of the papers. Better inclusion/exclusion criteria are necessary in order to foresee and avoid any possible co-interventions.

Adherence to treatment is crucial to adequate treatment and it is considered an important source of bias as noncompliant patients are removed from studies.

Withdrawal is considered acceptable when inferior to 20% of all enrolled individuals. We found 62.28% of RCTs to present acceptable withdrawn, which suggest biased conclusions once most dropouts are poor results [43,6]. Intention-to-Treat analysis with a growing curve, baseline data and admitting failure of treatment are helpful strategies to mitigate this issue.

Timing of assessment was just measured at the Clinical SRs group. Surgical trials did

not present this data. In 73.71% of the assessments there was a low risk of bias, which seems acceptable.

Selective Reporting was just measured at the Surgical SRs group. Low risk was found in just 15.05% of RCTs, which can drive investigators to biased results.

The main causes of withdrawal found are: deficiency in randomization (183 papers), retrospective studies [154], not specified results [63], incorrect or insufficient enrolment [62] and incorrect intervention [61]. Once again, this data point out deficit of high standard prospective papers in the Shouldered-and-Elbow field.

In summary, it is clear that Clinical and Surgical Primary Trials within the Shoulder-and-Elbow subspecialty need to improve their methodological quality in order to reach a suitable pattern of quality. Surgical trials tend to present less quality and more inconclusive results than clinical ones [41], because of following reasons: (1) learning curve of surgeons³³ – minimally invasive techniques and new procedures tend to be technically more demanding, thus having worse initial results; (2) surgical dexterity^{33,26} – natural variability among surgeons must be addressed at SRs as a potential cause of bias; (3) evolution of devices – rapid evolution of devices may add additional bias when compared to older models; (4) lower budgets for surgical trials; and (5) variability of lesions – surgical trials have to group variable similar lesions in packages in order to harmonize the baseline. Local and international societies of specialties and subspecialties will have to work hard together to get into an agreement about methodologies in their research. The IDEAL-Collaboration³⁴ has been a pioneer

on this field, establishing and improving research methodology for future Surgical Trials, but it still lacks recognition and acceptance by societies.

Limitations

This study is based only on SRs presented by Cochrane, which demands very high standards of methodology. Therefore, data presented here may underestimate methodological problem of primary papers. As data assessed in this paper were extracted from Cochrane's SRs, not directly from the primary trials, they might be susceptible to errors, misses or faults.

Conclusions

As most Systematic Reviews in Shoulder and Elbow Surgery contain primary trials and RCTs with a high risk of bias, their conclusions must be interpreted with caution.

Data from this study may be helpful for future generations of shoulder-and-elbow researchers helping them to focus on methodological quality improvements and thereby enhance worldwide discussion about surgical research methodology, standardized outcomes, timing of assessment and adequate monitoring. Also, it may help to establish new standards for primary surgical trials within shoulder and elbow surgery, like the IDEAL-Collaboration has done in many other surgical fields.

References

1. Alderson P. Absence of evidence is not evidence of absence. *British Medical Journal*. 2004; 328:476-7.
2. Amstutz HC, Sew HA, Clarke IC. Anatomic total shoulder Arthroplasty. *Clinical Orthopedic and Related Research*. 1981; 155: 7-20.
3. Bhandari BM, Morrow F, Kulkarni P, et al. Meta-analyses in Orthopedic Surgery. *Journal of Bone and Joint Surgery*. 2001; 83A (1), :15-24.
4. Blumenstein BA. Verifying keyed medical research data. *Statistics in Medicine*. 1993; 12(17) :1535-42.
5. Carvalho FR, Abe JM. (2011) Decision making by using the paraconsistent logic. 1st Ed, Blucher:São Paulo(Translated from the Portuguese).
6. Chalmers TC, Levin H, Sacks HS, et al. (1987) 'Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials.', *Statistics in Medicine*, 6(3), pp.315-328.
7. Chow SC, Chang M. (2008) 'Adaptive design methods in Clinical Trials: a review.' *Orphanet Journal of Rare Diseases*, 3, pp.11.
8. Cochrane Collaboration (nd) GMSG grading system, [Online] Available from: <http://musculoskeletal.cochrane.org/sites/musculoskeletal.cochrane.org/files/uploads/CMSG%20grading%20system.pdf> (Accessed 02 October 2014).
9. COMET (n.d.) Comet-Initiative [Online] Available from: <http://www.comet-initiative.org> Accessed 10 October 2014.
10. CONSORT (n.d.) The CONSORT statement [Online] Available from: <http://www.consort-statement.org> (Accessed 12 October 2015)
11. Constant CR, Murley AH. (1987) 'A clinical method of functional assessment of the shoulder.', *Clinical Orthopedic and Related Research*, 216, pp.160-164.

12. Day S, Fayers P, Harvey D. (1998) 'Double data entry: what value, what price?', *Controlled Clinical Trials*, 19 (1), February, pp. 15-24.
13. Handoll HH, Gillespie WJ, Gillespie LD, et al. (2008) 'The Cochrane collaboration: A leading role in producing reliable evidence to inform healthcare decisions in musculoskeletal trauma and disorders.' *Indian Journal of Orthopedics*, 42(3), pp: 247-251.
14. Harris JD, Quantman CE, Manring MM, et al. (2014) 'How to write a systematic review'. *The American Journal of Sports Medicine*, 42(11), pp.2761-2768.
15. Higgins JPT, Green S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0, The Cochrane Collaboration. [Online] Available from: www.cochrane-handbook.org (Accessed 8 August 2013).
16. Hopper AN, Jamilson MH, Lewis WG. (2007) 'Learning curve in surgical practice'. *Postgraduation Medical Journal*, 83(986), pp: 777-779.
17. ICH (1996) E6-GCP Guidelines [Online] Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1_Guideline.pdf (Accessed 27 July 2012).
18. ICH (1998) Statistical principles for Clinical Trials (E9) [Online] Available online from: www.ich.org (Accessed: 20 July 2013).
19. IDEAL (n.d.) Ideal-Collaboration [Online]. Available from: <http://www.ideal-collaboration.net> (Accessed: 20 October 2014).
20. Jadad AR, Moore RA, Carroll D, et al. (1996) 'Assessing the quality of reports of randomized Clinical Trials: is blinding necessary?' *Controlled Clinical Trials*, 17, pp. 1-12.
21. Kawado M, Hinotsu S, Matsuyama Y, et al. (2003) 'A comparison of error detection rates between the reading aloud method and the double data entry method.' *Controlled Clinical Trials*, 24 (5), pp. 560-569.
22. Khajuria A, Agha RA. (2013) 'Surgical Trials need for quantity and quality.' *Lancet*, 382(9908), pp: 1876.
23. Loke YK, Price D, Herxheimer A. (2007) 'Systematic reviews of adverse effects: framework for a structured approach.' *BMC Medical Research Methodology*, 5, pp.7-32.
24. McCulloch P, Altman DG, Campbell WB, et al. (2009) 'No surgical innovation without evaluation: IDEAL recommendations.' *Lancet*, 374, pp. 1105-1112.
25. Metzdorff MT. (2013) 'Evidence-based medicine: what it is, what it isn't, and are we practicing it?', *Journal of Trauma and Acute Care Surgery*, 75(6), pp.927-935.
26. Michener LA, McClure PW, Sennett BJ (2002) 'American Shoulder and Elbow Surgeons standardized assessment form, Patient self-report section, reliability, validity and responsiveness.' *Journal for Shoulder and Elbow Surgery*, 11(6), pp.587-594.
27. Newell DJ. (1992) 'Intention to treat analysis: Implications for quantitative and qualitative research.' *International Journal of Epidemiology*, 21: pp.837-841.
28. Peacock JL, Peacock PJ. (2011) *Oxford Handbook of Medical Statistics*. Oxford: Oxford University Press.
29. Prisma (2009) Welcome to PRISMA statement website [Online] Available from: <http://www.prisma-statement.org> (Accessed 3 April 2014).
30. Raboud JM, Montaner JS, Thorne A, et al. Impact of missing data due to dropouts on estimates of treatment effect in a randomized trial of antiretroviral therapy for HIV-infected individuals. Canadian HIV trials network A002 study group. *Journal of acquired immune deficiency syndromes and human retrovirology*. 1996; 12(1): 46-55.
31. Rowe CR, Patel D, Southmayd WW. The BAnkart procedure, a long term end-result study. *Journal of Bone and Joint Surgery*. 1978; 60(A):1-16.
32. Schulz KF, Chalmers I, Grimes DA, et al. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA*. 1994;272:125-8.
33. Shields PG. Publication Bias Is a Scientific Problem with Adverse Ethical Outcomes: The Case for a Section for Null Results. *Cancer Epidemiology, Biomarkers & Prevention*. 2007 Aug; 9: 771-2.
34. Shih WJ. (2002) 'Problems in dealing with missing data and informative censoring in clinical trials.' *Current Controlled Trials in Cardiovascular Medicine*, 3:4 [Online] Available from: <http://www.biomedcentral.com/1468-6708/3/4> (Accessed 20 October 2015)
35. Simes RJ. Publication bias: The case for an international registry of Clinical Trials. *Journal of Clinical Oncology*. 1986; 4:1529-41.
36. Somford MP, van Deurzen DF, Ostendorf M, et al. Quality of research and quality of reporting in elbow surgery trials. *Journal for shoulder and Elbow Surgery*. 2015; 4(10): 1619-26.
37. Stern JM, Simes RJ. Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal*. 1997; 315:640-5.
38. Streiner D, Geddes J. Some useful concepts and terms used in articles about diagnosis. *Evidence-Based Mental Health*. 1998; 1:6-7.
39. Struijs PA, Smidt N, Arola H, et al. (2002) Orthotic devices for the treatment of tennis elbow. *Cochrane Database of Systematic Reviews*, 1, Art. No.: CD001821. [Online] DOI: 10.1002/14651858.CD001821. (Accessed 1 March 2014).
40. Valania M. (2006) 'Quality Control and Assurance in Clinical Research - A system of checks and examinations that helps ensure the quality of Clinical Trials', *Applied Clinical Trials*, March [Online] Available from: <http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=a9h&AN=20194163&site=eds-live&scope=site> (Accessed 17 August 2012)
41. Velanovich V. Laparoscopic vs open surgery: a preliminary comparison of quality-of-life outcomes. *Surgical Endoscopy*. 2000; 14(1):16-21.
42. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized Clinical Trials conducting Systematic Reviews developed by Delphi consensus. *Journal of Clinical Epidemiology*. 1998; 51(12):1235-41.
43. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *Journal of Clinical Epidemiology*. 2003; 56:717-20.
44. Whitley E, Ball J. (2002) 'Statistics Review 4: Sample size calculations.' *Critical Care*, 6, pp.335-341

Conflict of Interest: NIL
Source of Support: NIL

How to Cite this Article

JC Garcia Jr, Gonçalves MHL, Cordeiro EF, Fink LFS, AM Cardoso Jr, Raffaelli MP, Fade MS. Cochrane's Shoulder-and-Elbow Systematic Reviews Issues: Mistrustful Impact of Evidence. *Acta of Shoulder and Elbow Surgery* Oct - Dec 2016;1(1):47-55