ARTHROSCOPY JOURNAL EDITORIAL REVIEW TEMPLATE

- Confidential Comments to Editor
  - List positives, negatives, and what is required to revise manuscript to make acceptable if possible. If not possible, outline fatal flaws.
  - Overall
    - Impact on clinical practice
    - Impact potential for future citations
    - Proper English with good grammar
  - Confirm or correct Level of Evidence with reasoning
  - **For systematic reviews (SR)** with or without meta-analysis (i.e. quantitative synthesis):
    - Confirm if this SR is needed and adds to existing literature (especially if SR performed recently).
    - Confirm there are enough included studies without population overlap to justify this SR.
      - For meta-analysis, a minimum of three studies is required for quantitative synthesis and creation of forest plots.
      - In general, low level of evidence (LOE) studies (below Level II) should not be quantitatively pooled.
    - Confirm that the SR adds meaningful information (can be positive or negative), but should not be "inconclusive" due to poor quality/heterogeneity of the included studies.

- Title
  - Concise and precise.
  - Attention getting or controversial titles are preferred.
    - Watch abbreviations (some are made up) – spell out and ensure clarity.
    - No industry names in title.
  - Check the Short Title for accuracy
- Abstract – should be stand alone
  - Purpose
    - Same as Purpose in Introduction. No introductory remarks.
  - Methods
    - If prospective or a cadaver study, the number of subjects is reported in Methods.
    - If retrospective, the number of subject is reported in Results.
    - Include main inclusion/exclusion criteria related to the purpose
    - No commercial/proprietary names in abstract unless exceptional reason this could be required.
    - **For Systematic Reviews:**
      - Need to declare if used PRISMA guidelines.
      - Report the databases used.
      - Need to declare method of study methodological quality assessment
        - Coleman, Modified Coleman, Cochrane, CLEAR-NPT, Delphi, Detsky, Jadad, CONSORT, STROBE, Newcastle, etc.
  - Results
    - If prospective or a cadaver study, the number of subjects is reported in Methods.
    - If retrospective, the number of subject is reported in Results.

        - Contain the final results/data that are being presented
        - Contains specific *P* values; e.g., *P* = .xxx, not *P* < .xx
            - Unless *P* < .001 then acceptable.
    - Conclusions
        - Identical to Conclusions of text: word for word.
        - Narrowed, specific, and supported by data, study design, and results; everything else should be moved to the Discussion.
    - Level of Evidence
        - See Level of Evidence (LOE) table in Arthroscopy Instructions for Authors. The authors should include descriptive terms, i.e. Level I, Prospective randomized study.
        - Needs to fit into one of the categories – If incorrect, provide reasons and recommend new LOE.
        - Not needed for non-clinical studies and studies not in LOE table (see below), which require Clinical Relevance
        - Survey results studies have neither Level of Evidence nor Clinical Relevance.
        - **For Systematic Reviews:**
            - LOE is the lowest (highest number) of the studies that are included in the review. I.e. if authors use Level III and IV studies to conduct the SR, then the LOE is Level IV.
            - Meta-analysis: same LOE rules as for SR.
    - Clinical Relevance
        - 1 to 2 sentences.
        - Used for non-clinical studies, i.e. animal and biomechanical studies
        - Used for clinical studies that do not fit Level of Evidence table for Therapeutic, Diagnostic, Prognostic or Economic Investigations.
- Introduction
    - Concise summary of the literature with appropriate references
    - Identify the controversy - what is known and unknown about the topic
    - Purpose
        - Second to last sentence. Same as in abstract.
        - Purpose should be as specific as possible and focus on primary outcome measure.
            - Secondary measures may be mentioned as indicated and should be identified as secondary.
    - Hypothesis
        - Last sentence
        - Specific and narrowed and ultimately matches that of conclusion.
        - Must be tested by the Methods.
- Methods
    - Methods should be stand alone and be reproducible, i.e., like a cookbook
    - If prospective or a cadaver study, the number of subjects is reported in Methods.
    - If retrospective, the number of subject is reported in Results.
    - List specific inclusion criteria then specific exclusion criteria:
        - Inclusion/Exclusion of concomitant procedures
        - Exclusion criteria and number of excluded patients by reason for exclusion should be accounted for in the text or Tables.
        - The inclusion/exclusion criteria should be separate and distinct from surgical indications
        - Surgical indications should be specifically listed after selection criteria for cases during collection period
        - **For Systematic Reviews:**

        - Exclusion criteria and number of excluded studies by reason for exclusion should be accounted for in a PRISMA flow chart.
- IRB approval
    - Required for all studies except retrospective and cadaver studies (unless the institution where the study was performed requires it).
    - Animal studies require specific agency approval.
- CONSORT guidelines
    - Recommended for all RCT
- Trial Registration
    - Suggested for RCT beginning patient enrollment after January 1, 2012.
- Minimum of 24-month follow-up on all patients is ideal, but for outcomes which occur shortly after treatment, shorter could be acceptable (e.g., post-operative infection).
    - Prefer 24-month follow-up on ≥80% as ideal but can allow some leeway as this is not easy to achieve
    - Shorter follow-up requires justification, i.e., that follow-up is adequate to answer the clinical question.
- **For systematic reviews** (with or without Meta-analysis):
    - PRISMA guidelines recommended
        - http://prisma-statement.org/PRISMAStatement/Checklist.aspx
    - Search terms, dates, databases and article inclusion/exclusion criteria well described and appropriate
    - Methods clarify the population studied, the intervention studied, and the outcomes considered
    - Assesses the risk of bias of each of the included studies
        - Note: some tools evaluate quality of the reporting
            - e.g. was randomization described, which is an assessment of reporting not the risk of bias.
    - Eligible studies assessed for eligibility by 2 or more examiners
    - Quantitative synthesis of low LOE studies is generally not appropriate and requires adjustment for study heterogeneity where groups or treatments are not similar enough to combine) i.e. verify that the data that is pooled is as "homogeneous" as possible minimizing selection bias
        - Generally performed only for Level I and II studies.
    - If a meta-analysis is performed, the rationale for pooling data, methods used (fixed versus random effects), and measures to quantify heterogeneity are described.
        - Heterogeneity is typically quantified with the I^2 statistic, which is the percent of variability in the summary effect that is due to heterogeneity rather than chance.
    - Heterogeneity, if deemed substantial, is addressed
        - Sources of heterogeneity (clinical characteristics or methodological differences among the studies) should be evaluated, e.g. subgroup analysis, discussion, etc.
    - If quantitatively synthesis, need verification by a statistician.
- Number of surgeons, number of facilities
- Surgical Technique
    - References to commercial products
        - Proprietary names listed once only in Text not in Abstract, use generic name otherwise
- Description of postoperative protocol
- Outcome Measures
    - Should identify appropriate primary outcome measure, as well as secondary measures.
    - Power analysis should evaluate primary outcome measure.

- Define what the Clinically Significant difference is (i.e. MCID or minimal clinically important difference.
- Radiologic studies defined
  - Inter- and intra-observer reliability
  - Existing information on reliability of the outcome measures should be cited. Reliability should be assessed for new outcome measurements
- Statistical Analysis
  - Final paragraphs of Methods
  - Power Analysis
    - *A priori* analysis required for most clinical studies to determine power and number of subjects required.
    - Comparative papers should include a prospective power calculation when available.
      - If there was not a prospective power calculation, then this must be pointed out as a limitation.
        - While controversial, *post hoc* power analysis could be provided in this case to better estimate validity of the conclusions but better to analyze confidence intervals (as below).
    - What was the primary outcome variable of interest – how was this powered and was it powered appropriately?
      - Did authors determine clinically significant difference (pre and post intervention)
        - MCID (minimal clinically important difference)
        - MDC (minimal detectable change)
        - PASS (patient acceptable symptomatic state)
      - What did they use to power this - what was the main variable and what was used as prior data (Standard Deviation from existing literature or pilot data)?
      - Bearing in mind that most reviewers and editors are not statisticians, are the statistical differences clinically meaningful and could change practice?
  - 95% confidence intervals (CIs) should be shown for the major comparisons of the paper. This is preferred over a *post hoc* power analysis, in which case authors should provide an interpretation of the CI width and direction in the Discussion section.
  - Standard deviations are acceptable if *a priori* power analysis done but otherwise prefer 95% CIs.
  - Overlap of CI or SD *suggest* that differences may be clinically insignificant despite adequate p-value.
  - Significance
    - In Methods, authors should state clearly what a clinically significant (or meaningful) difference is and also clearly state why this study was appropriately designed to be able to detect such a difference.
  - Specific tests for different outcomes
    - Discrete variables – if outcome measured is binary (success/failure) or measured by categories
      - Fisher Exact test: preferred test for binary outcomes – works for large or small samples sizes (< 25 samples) – This works as a "non-parametric approach as well"
      - Chi Square tests – can be used for binary outcomes with larger sample sizes (> 25 samples)
    - Continuous variables – if outcome is measured on a numeric (continuous) scale
      - Parametric— comparisons that can be made if the outcome variable is thought to have a normal distribution.

        - Paired *t* test: if individuals are matched – or looking at a pre-post change within one group
        - Student *t* test: comparison of two groups of individuals
        - ANOVA: Comparison of more than two groups of individuals (e.g., with 3 or more possible treatment choices)
        - ANCOVA – analysis of covariance – method to compare continuous outcome data on groups of individuals while adjusting for characteristics of the individuals (e.g., adjusting for patient's age and gender when comparing WOMAC scores)
    - Non-parametric – if outcome variables do not follow a normal distribution
        - Wilcoxon or Mann-Whitney: comparison of two groups of individuals on a continuous (non-normal) outcome. Wilcoxon is for paired sample groups. It's a non-parametric alternative to the paired Student *t* test.
        - Kruskal Wallis – comparison of more than two groups of individuals on a continuous (non-normal) outcome. Similar to ANOVA except for non-parametric data.
- Correlations – if investigators wish to assess the linear relationship between two continuous variables in a group of individuals
    - Pearson correlation (parametric) – if both measures are normally distributed
    - Spearman correlation (non-parametric) – if one (or both) of the measures are not normally distributed
- Correction for multiple comparisons (multiple statistical testing)
    - Bonferroni correction – *P* value divided by number of comparisons
    - Scheffé's method – another possible approach to adjusting for multiple comparisons

- Analyze for Biases
    - Selection (Allocation or Susceptibility Bias)
        - Treatment group has different prognosis, i.e., comparing apples to oranges.
        - Improved by blinding, randomization.
        - Improved by strict inclusion and exclusion criteria.
    - Performance
        - Who performed the surgery? One versus many surgeons – no right answer.
    - Transfer
        - Lost to follow-up, prefer >80% at 2 years
    - Reporting
        - Check that the outcome measures are correctly selected for the condition tested.
    - Recording
        - How was the data collected and by whom?
    - Order
        - Performing steps in a certain order; may lead to bias (for biomechanical studies)
- Types of Error
    - Type 1 Error (alpha) – Claiming that there is a difference between groups when there really is not
        - Rejecting the Null hypothesis incorrectly (False-positive)
    - Type 2 Error (beta) – Claiming that there is no difference between groups when there really is
        - Not rejecting the Null Hypothesis when it should be rejected (False-negative)
    - Statistical Power = 1-Type 2 Error (Beta)

- In "negative" studies – where "no difference" or "no association" is found one needs to be sure adequate attention is paid to whether Type 2 error has occurred as a result of studying too few patients resulting in lack of adequate power.
  - Inclusion of 95% confidence intervals for all main outcomes must be shown to allow review/reader to see range of plausible values
  - Some statement indicating what magnitude of difference could have been detected in this study should be made. This can be done.
    - If an *a priori* power calculation had been performed (preferred)
    - A *post hoc* power calculation is provided based on observed variability in the study

- Other
  - Sensitivity
    - TP / (TP+FN)
  - Specificity
    - TN / (TN+FP)
  - False negative rate
    - FNR = 1-Sens
  - False positive rate
    - FPR = 1-Spec
  - Positive Predictive Value (Precision)
    - PPV = TP / (TP+FP)
  - Negative Predictive Value
  - * NPV = TN / (TN + FN)
  - Accuracy
    - (TP+TN)/(TP+FP+FN+TN)
  - Abbreviations
    - T=true, F=false, P=positive, N=negative

- Results
  - Everything reported in Methods should be reported in Results
  - Should be reported in the same order as the Methods
  - If prospective or a cadaver study, the number of subjects is reported in Methods.
  - If retrospective, the number of subject is reported in Results
  - Follow-up time
    - Average and range (minimum – maximum)
    - Prefer minimum of 24 months in all patients (which is difficult to achieve and exceptions can be considered.
  - Specific *P* values need to be provided throughout manuscript; e.g., $P = .xxx$, not $P < .xx$
    - $P < .001$ acceptable if very small number.
  - Use Table and Figures for data if possible
  - **For systematic reviews:**
    - Describe how many studies were included or excluded
    - Account for potential duplicate publication (where data on the same patient is reported in more than one included study)
    - Results extracted and well tabulated
    - If the authors perform subjective (qualitative) synthesis, does the interpretation of results seem unbiased? Does the interpretation and summary of the results seem appropriate with regard to level of detail? Is the "bottom line" result clear? Can the results be generalized to other populations?

- Discussion
  - First sentence is a brief summary of the results: "The principal findings of this study show…."

- Discussion should avoid redundant repeating of numerical results
- Was clinically significant difference achieved?
- Was there a statistical difference (but not a clinical one) or vice versa?
- Did there appear to be a clinically meaningful observed difference between groups that did not reach statistical significance?
  - If yes, was the study underpowered (too few patients)?
- Does the study show a statistically significant difference but not a clinically significant difference?
  - This can happen in an “over-powered” study (where a huge number of subjects are included) where a very small difference can be detected with a statistical test – but the magnitude of the difference is not clinically relevant.
- Must compare and contrast results with other publications, and explain contrasts.
  - However, should avoid extensive details of numerical results of all other studies.
- The clinical relevance of the results should be considered.
- Limitations
  - Section just before the Conclusions
  - Various types of bias should be considered in detail, as well as potential lack of power analysis if relevant, and anything else authors (or Reviewers) can suggest.
- **For systematic reviews:**
  - Do the authors summarize the results in the context of existing knowledge and literature and clinical expertise?
  - Do the authors (inappropriately) discuss their own opinions and bias? Do the authors address gaps in clinical knowledge?
  - Do the authors identify consistencies, or inconsistencies and conflicts, in the included, primary data?
  - Are directions for additional research proposed, and are these recommendations supported by the reported data?
  - Do the authors well address, and account for or attempt to mitigate against, the limitations of the study (including typical limitations of included articles such as low level of evidence and heterogeneity)?
  - In considering limitations, do the authors address publication bias and /or strengths and weaknesses of the primary evidence?

- Conclusions
  - Identical to Abstract Conclusions.
  - Specific and narrow, based only on the results. NOT overstated, NOT opinions.
  - Addresses whether the data does or does not support the hypothesis.
- References
  - Up-to-date within the last 5 years
  - How many references are within the last 5 years?
- Tables and Figures
  - Data should either be in the Results section or in a Table to avoid duplication of information
  - Excellent way to sum up results and present data
  - Ideas for additional Tables include lists of Tips, Pearls, Pitfalls, Key Points, Comparative studies if there are a large number
  - Indications, contraindications, risks, treatment algorithms, etc.
  - Labels on figures are generally always helpful.
  - Tables should have only a short title and any necessary notes, including abbreviations.
  - Figure legends must "stand alone"
    - i.e., contain a complete, take-home, educational message, as if a reader viewed only that Figure without looking at any other Figure or without reading the text.

- Legends must mention patient position, side, and viewing portal or MRI orientation as appropriate.
- Legends are almost always incomplete and require careful review.