Arthroscopy Journal Original Article Template

• GENERAL
  o Word count: 4,000 words (excluding title page, blind title page, references, figure legends)
    ▪ Abstract word count: 300
  o Figure count: 7 (15 total figure parts)
  o Table count: 4

• TITLE
  o Concise, precise
  o Should be attention-getting, controversial
  o Watch abbreviations – spell out, ensure clarity
  o No industry names

• ABSTRACT
  o Should be able to “stand-alone” – any author/reviewer/editor should be able to read the abstract and obtain ALL the relevant information about study design, conduct, findings, conclusions
  o Abstract plays three roles (Ibrahim AM, Dimick JB; Writing for Impact: How to prepare a journal article; 2017):
    ▪ When writing
      • Improves your research question, manuscript writing
    ▪ Once submitted
      • Convinces editors it’s worthy of 1) peer review, 2) publishing
      • Half of manuscripts at high-impact journals are rejected based on abstract alone
    ▪ After publication
      • Getting the rest of the article read
      • Readers start at the abstract and decide if rest of article worth reading
  o Abstract PEARLS:
    ▪ Far too many authors choose a title and abstract just before submission, with little thought put into either. Unfortunately, these are the first (and frequently, the only) parts of the paper that will be read
  o PURPOSE
    ▪ Should verbatim match the purpose in the manuscript body’s Introduction.
    ▪ No background, no introduction, no hypothesis
    ▪ Should be specific, precise
  o METHODS
- If prospective study (prospective study design and conduct), then the number of subjects (and relevant demographics [e.g. age, sex]) reported in Methods.
  - If retrospective study (retrospective analysis of study participants – this includes retrospective analysis of prospectively-collected data), then number of subjects reported in Results
- If cadaver study, then number of specimens reported in Methods
- Include all main inclusion and exclusion criteria relative to study Purpose
- No commercial/industry/proprietary names unless exceptional reason(s)
- Brief description of pertinent statistical methods

  o **RESULTS**
  - Directly answer the primary purpose(s)
  - Results pertinent to primary (and relevant secondary or exploratory) outcome
    - Mean +/- standard deviation
    - Statistically significant findings should be stated with exact p-values (e.g. \(p=0.03\), \(p=0.43\), etc.) with 95% confidence intervals, not < or > (only exception \(p<0.001\))
    - If a result is statistically significant, then its clinically relevance, when applicable, should be stated
      - MCID (minimal clinically important difference), SCB (substantial clinical benefit), and PASS (patient acceptable symptom state) are a few examples of clinical relevance
    - If a result is not statistically significant, then a power analysis (a priori or post-hoc) will reduce the risk of beta error.
      - A post-hoc power analysis determines what sample size would be required for the non-significant results to be significant. Avoid calculating observed power. If the results fail to reach statistical significance, an observed power calculation will confirm what is already known- the study is underpowered.

  o **CONCLUSIONS**
  - State the most important finding(s) of your study based on your data, results, findings, outcomes without opinion, overstatement, speculation, suggestion, or editorialization
  - If the statement is not based on the results of the study, it does not belong in Conclusions section
  - Conclusions in abstract should verbatim match the Conclusions in manuscript body.
  - Only significant findings can be reported in Conclusions

  o **LEVEL OF EVIDENCE**
  - See Arthroscopy: The Journal of Arthroscopic and Related Research journal 2018 update on level of evidence – WILL UPDATE ONCE AVAILABLE
  - Step 1: determine type of evidence (based on Oxford Centre for Evidence-Based Medicine – OCEBM Levels of Evidence Working Group. The Oxford 2011 Levels of Evidence)
    - Therapeutic: investigation of treatments (intervention, surgery, medicine) with a pre- and post-assessment of participants
    - Diagnostic: investigation of diagnostic tests (imaging, fluid analysis, physical examination); can assess accuracy (versus a gold standard)
• Prognostic: investigation of prognostic entities (risk factors that can affect a
group of subjects, population, sample, group); what happens if no
intervention performed (control arm of randomized trial)
• Economic: investigation of outcome per unit of cost, to determine value of
an intervention
• Epidemiologic: cross-sectional investigation of a population; prevalence;
survey
  ▪ Step 2: determine level of evidence based on specific type
    • I‐IV: based on study design, conduct, and reporting
    • V: expert opinion, editorial, letter to editor
  ▪ Step 3: determine if level of evidence should be retained or downgraded
    • Example: Not all randomized controlled trials are level I evidence. If less
      than 80% follow‐up, less than two years follow‐up, high rates of dropout or
crossover, underpowered study with small sample sizes, high degrees of
heterogeneity, etc., then a Level I study may be downgraded to a Level II
  ▪ Level of evidence is not a firm rule, just a suggested or recommended single quality
    measure.
  ▪ Level of evidence does not provide a recommendation (such as GRADE – Grading of
    Recommendations, Assessment, Development, and Evaluation).

  ○ CLINICAL RELEVANCE
    ▪ One to two sentences
    ▪ For basic science, in vitro investigations – how does the study translate to clinical
      practice?

• INTRODUCTION
  ○ Concise summary of literature with appropriate references
  ○ Should give the author/reviewer/editor context and a reason to care about the topic, why it
    is important
  ○ Should create a knowledge gap, making it clear to the author/reviewer/editor what is
    known and unknown
  ○ Should state how the current investigation will address the knowledge gap with a specific
    purpose(s) and hypothesis(es)
  ○ Purpose
    ▪ Second to last sentence.
    ▪ Should verbatim match the purpose in the abstract
    ▪ Should be as specific as possible and address the primary outcome measure
  ○ Hypothesis
    ▪ Last sentence
    ▪ Specific, matches the purpose (should either be supported or rejected in the
      conclusion)
    ▪ Must be tested by the Methods

• METHODS
  ○ Should be able to “stand alone” – telling the author/reviewer/editor exactly and
    transparently everything you did.
  ○ Clear statement of IRB (Institutional Review Board) approval, when applicable
    ▪ Animal studies should have appropriate board approval (e.g. IACUC – Institutional
      Animal Care and Use Committee)
  ○ Clear, specific inclusion and exclusion criteria
- Should follow PICOS (participants, interventions, comparators, outcomes, study design) in order to optimize inclusion and exclusion criteria
- Specific inclusion and exclusion criteria for concomitant procedures as well (heterogeneity)
- Reasons for exclusion can be illustrated numerically via flow chart (figure)
- Surgical indications are separate from inclusion/exclusion criteria
  - Should register all prospective clinical trials on any national or international registry (prospective publication of clinical research study authors, title, purpose, hypothesis, methods, statistical analysis, and confirmation of IRB approval) – http://clinicaltrials.gov is an example
- CONSORT – CONsolidated Standards of Reporting Trials for parallel group randomized controlled trials – 25 item checklist to assist in improving quality of randomized trials. Provides minimum set of recommendation for design, conduct, and reporting. Checklist should be submitted for manuscripts of randomized trials.
- Should describe if efficacy or effectiveness design:
  - Efficacy: intervention(s) performed under ideal conditions
    - High homogeneity in intervention (high internal validity)
    - Exploratory trial that determines the pure effect of the intervention
  - Effectiveness: intervention(s) performed under everyday conditions
    - Higher degree of heterogeneity (high external validity, generalizability)
    - Pragmatic trial that better simulates daily practice
- Trial design – What are you trying to prove?
  - Superiority: when purpose is to prove intervention is better than control
    - Majority of trials in arthroscopic and related surgery
    - Caution: If no significant difference is detected between groups, this does not prove that there is truly no difference (beta error)
  - Non-inferiority: when purpose is to prove intervention is not worse than control
    - Maximum amount of possible difference between intervention and control is used, in favor of the control, and sample size calculated
    - Used when intervention not expected to be better than control, but advantageous because of another reason (safety, cost, compliance)
    - Requires larger sample size than superiority
  - Equivalence: when purpose is to prove intervention is similar to control
    - Maximum amount of possible difference between intervention and control is used to establish equivalence and sample size
    - Requires larger sample size than non-inferiority
    - Sample size of equivalence trial is approximately four times that of corresponding superiority trial
- If prospective study (prospective study design and conduct), then the number of subjects (and relevant demographics [e.g. age, sex]) reported in Methods.
- Should describe follow-up completely
  - Minimum and maximum permissible length
    - Greater than two years minimum preferred
    - Greater than 80% of enrolled subjects preferred
  - Investigator (blinding) that performed follow-up assessments
- Should describe intervention(s) completely
  - Number of surgeons, number of facilities
  - Surgical technique (references to commercial/industry/proprietary products listed here (including country, city, state of company)
  - Post-operative protocol (including physical therapy, durable medical equipment, return to activities, including sport)

- Should describe outcome measures completely
  - Objective clinician-measured
    - Physical examination (range of motion, strength, special testing)
    - Imaging (plain radiographs, magnetic resonance imaging, computed tomography)
    - Surgical findings
  - Subjective patient-reported
    - General health
    - Quality of life
    - Joint-specific
    - Limb-specific
    - Disease-specific
    - Activity level
  - Power analysis (sample size calculation) should evaluate the primary outcome measure
  - Inter- and intra-observer reliability measured, when applicable

- During study design (and preferably prior to study conduct), should utilize at least one appropriate methodological quality score or recommendation score to help identify the minimum relevant components of reporting once the study manuscript is ready to be written and published. Examples, based on study type, from the EQUATOR network listed here:
  - EQUATOR Network (Enhancing the QUAliity and Transparency Of health Research) – “umbrella” organization for guidelines, peer-reviewed articles, funding resources, and other entities collaborating to improve research quality
  - AGREE – Appraisal of Guidelines REsearch & Evaluation for clinical practice guidelines
  - ARRIVE – Animal Research: Reporting of In-Vivo Experiments
  - CARE – CAse REport guidelines for completeness, transparency, and data analysis in case reports
  - CHEERS – Consolidated Health Economic Evaluation Reporting Standards
  - CLEAR-NPT – CheckList to Evaluate A Report of a Non-Pharmacologic Trial
  - Cochrane Quality Assessment Tool – for randomized trials
  - Coleman/Modified Coleman – for randomized and non-randomized orthopedic trials
  - CONSORT – CONsolidated Standards of Reporting Trials for parallel group randomized controlled trials
  - COREQ – COnsolidated criteria for REporting Qualitative research – 32 item checklist for interviews and focus groups
  - Delphi - 8-item quality assessment tool for randomized and non-randomized trials
  - Detsky – 14-item quality assessment tool for randomized and non-randomized trials
  - Downs and Black – 27-item quality assessment tool for randomized and non-randomized trials
- GRADE – Grading of Recommendations, Assessment, Development, and Evaluation
- Jadad – simple three question (randomization, blinding, withdrawals) scale
- Level of evidence – I-V based on Center for Evidence-Based Medicine (CEBM)
- MECIR – Methodological Expectations of Cochrane Intervention Reviews
- MINORS – Methodological Index for NON-Randomized Studies
- Newcastle-Ottawa – quality assessment of non-randomized studies to be used in systematic reviews
- PRISMA – Preferred Reporting Items for Systematic reviews and Meta-Analyses
- QUADAS – QUality Assessment of Diagnostic Accuracy Studies
- Quality Appraisal Tool – percentile quality rating for non-randomized trials
- R-AMSTAR/AMSTAR – Revised Assessment of Multiple Systematic Reviews
- SORT – Strength Of Recommendation Taxonomy
- SPIRIT – Standard Protocol Items: Recommendations for Interventional Trials
- SQUIRE – Standards for QUality Improvement Reporting Excellence
- SRQR – Standards of Reporting Qualitative Research
- STARD – STAndards for Reporting Diagnostic accuracy studies
- STROBE - STrengthening the Reporting of OBServational studies in Epidemiology
  - Should assess bias qualitatively and attempt to mitigate
    - Selection, detection, performance, transfer, non-responder, publication, study design
  - Should address statistical analysis completely with all relevant details.
    - Statistical analysis should be final paragraph of the Methods
    - Should include comparison of statistically significant findings to that of clinical importance/relevance: MCID, PASS, SCB – it is critical to ensure that these concepts are used for “within-individual” change, not “group level”, “population level” changes observed with means of groups. See Harris JD, et al Arthroscopy 2017 Jun;33(6):1102-12 for a complete description of concept.
    - Power analysis (sample size calculation) – required a priori for most clinical studies and should include the size of the effect (anticipated difference in means or proportions, etc.) and the estimated amount of variability (anticipated standard deviation of the groups or effect).
    - 95% confidence intervals should be used with most, if not all, comparisons, including p-values
  - Variable types
    - Categorical (aka qualitative or discrete) variables
      - Ordinal
        - 2 or more categories, and have intrinsic order
        - Example: IKDC objective score, A, B, C, D, excellent, good, fair, poor
      - Nominal
        - 2 or more categories, but no intrinsic order
        - Example: 50 categories of USA states
      - Dichotomous
        - Type of nominal variable, but only 2 categories
        - Example: Male, Female; Yes, no.
    - Continuous (aka quantitative, numerical) variables
      - Interval
        - Measured along a continuum and have numerical value
Example: 20 deg Celsius is 10 less than 30 deg Celsius, which is same difference as 30 deg Celsius being 10 less than 40 deg Celsius

- Ratio
  - These are interval variables, but with the added condition that zero of the measurement means that there is none of that variable.
  - Thus, zero deg Celsius is just along the continuum, not that there is no temperature
  - However, Kelvin is ratio variable, b/c zero Kelvin means there is no temperature
  - Other examples: height, mass, distance

- Statistical comparisons (continuous); data distribution
  - Parametric
    - Makes assumption of normally distributed continuous data
    - Test for normality – Kolmogorov-Smirnov; Shapiro-Wilk
    - Independent variable (IV); Dependent variable (DV)
    - Two groups (IV):
      - Paired t-test: paired groups (IV); one continuous DV
        - Example: Pre vs. Post-Op Outcome scores in one group of patients.
      - Independent t-test: unpaired groups (IV); one continuous DV
        - Example: Post-Op outcome scores between two different groups, e.g. Single vs. Double Bundle ACL Reconstruction.
    - More than two groups (IV):
      - ANOVA: three or more independent groups (IV); one continuous DV
      - ANCOVA: three or more independent groups (IV); one continuous DV; comparison while controlling for covariates
      - MANOVA: ANOVA with multiple DV
      - MANCOVA: MANOVA comparison while controlling for covariates
  - Non-parametric
    - Not normally distributed data
    - Wilcoxon signed rank (paired) and Mann-Whitney U (independent) tests are non-parametric alternatives to the parametric t-test
    - Kruskal-Wallis is non-parametric alternative to ANOVA

- Statistical comparisons (categorical)
  - Chi-squared: comparison of distribution of categorical variable in one group with distribution of categorical variable in another group (comparison of frequencies or proportions)
  - Fisher’s exact: although frequently used for small samples (frequency less than 5), it is valid for all sample sizes
  - McNemar Test: comparisons of proportions in paired data.

- Correlations
  - Pearson (parametric)
  - Spearman (non-parametric or ordinal data)
  - Point biserial (one continuous variable and one dichotomous variable)

- Remember, correlation does not equal causation
- Regression
- **Linear Regression**: used to examine the ability of one or more IVs to predict a continuous DV variable. When more than one IV is included, it’s considered Multiple Linear Regression.
  - Example: Is the use of a femoral nerve block (IV) in ACL reconstruction predictive of the amount of opioids consumed (DV)?
  - Output: regression coefficient or “slope” which indicates how much of change in opioid consumption (DV) would be expected when a femoral nerve block is used (IV).

- **Logistic Regression**: used to examine the ability of one or more IVs to predict a dichotomous DV. When more than one IV is included, it’s considered Multiple Logistic Regression.
  - Example: Is the number of preoperative instability episodes (IV) predictive of the presence of an off-track lesion (DV) in patients with traumatic anterior instability.
  - Output: Odds Ratio (OR). An OR of 1.5 would indicate that for every one-unit increase in the number of preoperative instability episodes (IV), we expect to see a 1.5 times increase or 50% increase in the odds of an off-track lesion (DV).

- **Advanced Predictive Modeling**: The use of multivariable linear or logistic regression to predict an outcome while controlling for various covariates. Variables that are related to the outcome are entered into a multiple regression model. Different multiple regression models are trialed with the goal of isolating the variables most predictive of the outcome. Consultation with a statistician is recommended.

  - **Survival Analysis**: analysis of time to an event data.
    - Estimates the probability of an event occurring within given time points.
      - Example: In patients who have undergone arthroscopic rotator cuff repair, what’s the probability of structural failure in first 12 months following surgery?
      - Output: Cumulative probability of surviving each time period. In this example, the probability of surviving (not having a failed repair) the first 3 months would be cumulative probability of surviving the first, second, third months. This data is represented in Survival or Kaplan-Meier curves, which allow the probability of survival from one time point to the next to be represented graphically.
    - Cox proportional hazards regression: used to examine the relationship of one or more IVs and a single time to event DV.
      - Example: Does smoking impact the time to rotator cuff repair failure?
      - Output: Hazard Ratio (HR), which is the ratio of incidence of rates. An HR of 1.5 would indicate that smokers have 1.5 times or a 50% higher incidence rate of failure compared to non-smokers.
      - Consultation with a statistician is recommended for advanced, multivariable modeling with proportional hazards regression.

- **Other Statistical Considerations**
  - **Corrections for multiple comparisons**: Used when there are more than 2 groups being compared
• Common approaches include Tukey, Newman-Keuls, Scheffee, Bonferroni, Dunnett

• RESULTS
  o Should be able to “stand alone” – telling the author/reviewer/editor exactly and transparently everything you found – in essence, match the Methods section.
  o Should directly answer the primary (and secondary/tertiary) outcome(s)
  o Data should either be in text in Results section or in Figures/Tables – do not duplicate (minimize/avoid redundancy)
  o Should completely describe all relevant demographics of the study, participants, intervention(s), and outcomes – leave the author/reviewer/editor without a question to ask
    ▪ If retrospective study (retrospective analysis of study participants – this includes retrospective analysis of prospectively-collected data), then number of subjects reported in Results
    ▪ Follow-up time should be specific (mean +/- standard deviation; range) – preferred minimum 24 months necessary, with few exceptions
    ▪ Statistically significant findings should be stated with exact p-values (e.g. p=0.03, p=0.43, etc.) with 95% confidence intervals, not < or > (only exception p<0.001)
  o No discussion, opinion, or speculation – only data, results, outcomes, findings

• DISCUSSION
  o First paragraph should briefly summarize the principal findings of the study and whether hypothesis(es) were confirmed or rejected.
  o Should not just re-state the results.
  o Should compare and contrast results of current investigation with that of other similar trials (both original research and synthetic reviews like systematic reviews or meta-analyses).
    ▪ Does not need to be a complete review of the entire literature
  o Should discuss statistical significance versus clinical relevance for statistically significant results
    ▪ MCID, MDC, SCB, PASS
  o The precision of the results should be discussed. Specifically, the width of confidence intervals and their potential impact of the interpretation of the results.
  o Should discuss possibility of beta error in comparisons that failed to detect statistically significant differences – power analyses/sample size calculations
  o May provide conjecture or speculation as to interpretation of your results
  o LIMITATIONS
    ▪ Should be final paragraph of Discussion
      • All types of bias should be addressed and discussed
      • Selection: allocation or susceptibility bias; “apples versus oranges”
        o Improved by randomization, blinding, strict inclusion/exclusion criteria
      • Detection: recording bias; method of outcome detection differs between groups
      • Performance: heterogeneity in intervention(s)
      • Transfer: exclusion bias; occurs with inadequate follow-up
      • Non-responder: participation bias; respondents differ in meaningful ways from non-respondents
      • Publication: studies with positive findings have greater chance of and faster time to publication than studies with negative findings
• Confirmation: when researcher/author forms hypothesis and uses/selects/finds data to support that hypothesis
  ▪ Types of error:
    • Type 2 (beta): claiming no difference between two groups, when one actually exists (false negative)
      o Not rejecting the null hypothesis when it should be rejected
      o Frequent in underpowered studies
    • Type 1 (alpha): claiming significant difference between two groups, when one does not exist (false positive)
      o Rejecting the null hypothesis when it should not be rejected

• CONCLUSIONS
  o State the most important finding(s) of your study based on your data, results, findings, outcomes without opinion, overstatement, speculation, suggestion, or editorialization
  o If the statement is not based on the results of the study, it does not belong in Conclusions section
    ▪ Only significant findings can be reported in Conclusions
  o
  o Conclusions in abstract should verbatim match the Conclusions in manuscript body.

• REFERENCES
  o Should be up-to-date, most within the past five years
  o Follow PubMed journal abbreviations
  o Should be numbered consecutively in the order in which they appear in the text

• FIGURES, TABLES
  o Excellent way to efficiently summarize results
  o Legends must define all abbreviations
  o Should be able to “stand-alone”, so that the author/reviewer/editor can understand everything relevant to the figure/table with a single take-home educational message
  o Should always mention patient position, viewing perspective (portal), side, imaging orientation – labels on the actual figure are always helpful
  o Data should either be in text in Results section or in Figures/Tables – do not duplicate (minimize/avoid redundancy)
  o Legends for figures are almost always incomplete and require careful review.