# **Vibration of effects from differing methodological choices when conducting a meta-analysis and how it may lead to conflicting results**

Nicolai Sandau MD1, Thomas Vedste Aagaard MSc 1,3, Asbjørn Hróbjartsson MD PhD2, Ian Harris MBBS MMed(Clin Epi) PhD4, Stig Brorson MD PhD DMSc1

1 Centre for Evidence-Based Orthopedics, Department of Orthopedic Surgery, Zealand University Hospital, Køge, Denmark

2 Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Denmark   
3 Department of Physiotherapy and Occupational Therapy, Holbaek Hospital, Denmark.  
4 Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, South Western Sydney Clinical School, University of New South Wales (UNSW Sydney), Liverpool, NSW, 2170, Australia

## Introduction

Despite the publication of numerous meta-analyses, no consensus currently exist concerning the optimal primary treatment for proximal humerus fractures (PHF) [1,3,4,7–11,14,19,20,22,26]. Some of these overlapping meta-analyses report conflicting results and conclusions thereby potentially leading us further away from reaching a consensus on the optimal treatment [2].

These conflicting results could be an effect of different methodological choices. Methodological choices regarding the definition of outcomes have been shown, due to the multiplicity of reported outcomes in trial reports, to influence the effect estimates of meta-analyses [23,24]. This, in combination with, other methodological choices regarding for example inclusion criteria, subgroup analyses, and methods of analysis available to authors, can potentially lead to a wide spectrum of effect estimates.

One way to investigate the cumulative effect of different methodological choices on effect estimates is by conducting a vibration of effects (VoE) analysis [6,18,21]. A VoE analysis is performed by simulating methodologically distinct modeling scenarios and analyzing the variability of the resulting effect estimates. By simulating a set of pre-defined methodological choices VoE allows for a simulation that more closely resembles realistic modeling scenarios compared to ‘all subsets’ and Monte Carlo simulations.

VoE has primarily been described in an epidemiological context with regards to the selection of covariates, but a recent study has investigated the VoE in meta-analyses used in the comparison of two drugs used in the treatment of alcohol use disorders [17]. In the study, the authors developed an algorithm that simulated a large number of unique combinations of methodological choices and for each combination conducted a meta-analysis. The authors reported that varying combinations of methodological choices could result in effect estimates favoring opposite interventions [17].

One way to reduce the potential VoE of a meta-analysis is with a pre-defined protocol. However, it has previously been reported that such protocols often lack the required specificity with regards to the definition of outcomes thereby potentially resulting in selective reporting bias [12,13,15,24]. In addition to defining outcomes of interest, the authors must also specify other methodological choices such as search strategy, language restrictions, subgroup analyses, and choice of meta-analysis model. This can be a challenging task as the authors do not know the nature of the available data beforehand. Therefore, we hypothesize that, despite a pre-defined protocol, a considerable VoE remains, potentially leading to selective reporting bias[16].

With this study, using the comparison of surgical and non-surgical interventions in the primary treatment of PHFs as a case study, we aim to:

1. Investigate the overall VoE for meta-analyses comparing surgical and non-surgical interventions in the primary treatment of PHFs, thereby allowing us to investigate the methodological scenarios leading to conflicting results.
2. Investigate the residual VoE of pre-defined protocols for published meta-analyses comparing surgical and non-surgical interventions in the primary treatment of PHFs.

## Methods

### Information sources

We will use published meta-analyses comparing surgical and non-surgical interventions in the primary treatment of PHFs to identify primary trials for data extraction.

### Eligibility criteria

All published meta-analyses, including network meta-analyses, comparing any surgical intervention with non-surgical intervention in the primary treatment of proximal humerus fractures will be considered eligible as an information source. No restrictions on language or publication date will be set.

### Search strategy

The bibliographic databases PubMed, EMBASE, The Cochrane Library, and Web of Science will be searched to identify any eligible meta-analyses. The search strategy will be developed for PubMed and will be adapted for the other bibliographic databases.

#### Selection process

The screening and selection process will be performed independently by the first and second authors. Any conflicts will be solved by consensus.

##### Meta-analyses

The abstracts of the search results will be screened for potentially eligible meta-analyses. The full texts of all potentially eligible meta-analyses will then be obtained and screened for final inclusion.

##### Trial reports

We will include all primary trials from which any of the identified meta-analyses have used outcome data in one of their analyses. Both randomized controlled trials (RCTs) and non-randomized studies of intervention (NRSI) will be included. NRSI will be defined using the definition from the Cochrane Handbook for Systematic Reviews of Interventions: “any quantitative study estimating the effectiveness of an intervention (harm or benefit) that does not use randomization to allocate units (individuals or clusters of individuals) to intervention groups.” [5]. We will only include comparative NRSI. Outcomes adjusted for confounding will be used in preference to unadjusted outcomes.

### Data items

Data will be extracted independently by the first and second authors using a piloted extraction spread-sheet. Conflicts will be resolved by consensus.

The following data items will be extracted from the included primary trial reports:

* Publication language
* Publication status (published, unpublished)
* Publication date
* The databases in which the report was retrieved
* Number of participants
* The average age of participants
* Neer fracture classifications
* Intervention
* Time of intervention
* Time to follow-up for all reported outcome measures
* Outcome results at any reported time point for:
  + Any validated patient-reported outcome measure (PROM) of upper limb function (e.g., Disabilities of the Arm, Shoulder, and Hand (DASH) and Oxford Shoulder Score (OSS))
  + Constant Shoulder score (CS)
  + Any validated health-related quality of life assessments (HRQoL)
  + Secondary surgery
  + Major complications (e.g., deep infection, nerve damage, avascular necrosis)
* Method of outcome analysis (per-protocol, intention-to-treat)

Despite multiple other objective outcomes related to functional outcome reported in trials comparing non-operative with operative treatment for PHFs (e.g., range of motion, strength, radiological outcomes), we have chosen to limit our outcome measures to the ones described above, due to these measures being the most widely reported in meta-analyses conducted on the subject [2].

Data from unpublished trial reports will be extracted directly from the meta-analyses in which the trial data is reported. In case of any discrepancies between two meta-analyses reporting on the same unpublished trial report, the data will be extracted from the most recently published meta-analysis.

If only a median value is reported for an outcome, the mean and standard deviation will be estimated using the method described by Wan et al. [25]. If a time point for an outcome is reported as a range without a median or mean, the minimum value in the range will be extracted as the timepoint for that given outcome. If the mean follow-up time is reported separately for each intervention group, the combined mean will be calculated using the method described in Cochrane Handbook for Systematic Reviews of Interventions [5].

Any missing standard deviations will be imputed using the methods described in Cochrane Handbook for Systematic Reviews of Interventions [5]. If it is not possible to impute a standard deviation for an outcome result, but a meta-analysis has imputed standard deviations using methods not described in the Cochrane Handbook for Systematic Reviews of Interventions, the standard deviation will be extracted from the corresponding meta-analysis.

DASH and quick-DASH will be reversed to conform with the scales of the other functional outcome measures such that an increasing score equals worsening disability. For the HRQoL outcomes, if physical and mental components are reported separately without a combined score, the physical component score will be extracted. If a trial report includes both adjusted and unadjusted CS, the unadjusted score will be extracted.

The RoB evaluations for each trial report will be extracted from the meta-analyses in which they were included. The Cochrane Risk of Bias tool (v2 preferred over v1) will be extracted for RCTs and ROBINS-I for NRSI. If a RoB evaluation for a given trial has been reported in multiple meta-analyses, data will be extracted from the meta-analysis with the latest publication date. If a trial report has not been evaluated using one of the aforementioned tools we will choose by consensus which RoB evaluation to extract.

Lastly, we will contact the authors of the meta-analyses used as information sources that did not have a registered protocol and request a copy of any pre-defined protocol or data analysis plan.

### Data synthesis

#### Simulation of methodological choices

Table 1 lists the planned methodological choices to be included in the algorithm. The choices regarding search strategy, inclusion criteria, method of analysis, and subgroup analyses are based on methodological choices we have observed in published meta-analyses regarding the treatment of PHFs.

The simulated choices regarding outcome definitions are based on the 5 items describing an outcome as described by Mayo-Wilson et al.: 1) domain (e.g. functional outcome), 2) measure ( e.g. Oxford Shoulder Score), 3) metric (e.g. change from baseline), 4) method of aggregation (e.g. continuous) and 5) timepoint of assessment [13]. We will conduct the VoE analyses for three outcome domains: functional outcome scores (continuous), HRQoL assessments (continuous), and harms (secondary surgery and major complications, both dichotomous). The definitions of the remaining outcome items are described in table 1.

If insufficient data is available to simulate a choice concerning an inclusion criterion, the trial will only be included for the option that includes all trials (e.g., ‘Inclusion of all trials regardless of sample size’ for the choice relating to trial sample size).

The simulation algorithm will be developed using the statistical software R.

#### Meta-analysis

For each distinct methodological scenario, we conduct a meta-analysis with a resulting effect estimate and p-value. The effect estimates for continuous outcomes will be calculated using the standardized mean difference (SMD) while dichotomous outcomes will be calculated as relative risk (RR). Non-operative treatment will be defined as the comparator for all three outcome domains. For each effect estimate, we also calculate P-values and 95% confidence intervals. Heterogeneity for each meta-analysis will be calculated using the Cochrane Q test and I2 index. Statistical significance will be defined as p-values less than 0.05.

The final number of unique combinations available for analysis depends on the number, and the heterogeneity, of the included trial reports. Combinations resulting in only one trial available for meta-analysis will be excluded.

#### Analysis of VoE

The results will be visualized with the effect estimate on the x-axis and the p-values on a log-transformed y-axis. This allows for visual analysis of the overall VoE. Furthermore, we will calculate the 1st, 50th, and 99th percentile allowing us to determine if a Janus effect is present. The Janus effect is defined as when the 1st and 99th percentile of the effect estimates are favoring opposite interventions [18]. If our results are normally distributed, we will calculate the standard deviation of the overall VoE. Lastly, we will present the specific methodological scenarios leading to the most conflicting results.

If we obtain any pre-defined data analysis plans, we will create separate algorithms that match each protocol as closely as possible. Using the methods described above we will then investigate the residual VoE for each of the obtained protocols.

Table 1: Methodological choices

|  |  |
| --- | --- |
| **Choices related to search strategy and inclusion criteria** | |
| Bibliographic databases | 1. Only inclusion of trial reports identified using a search strategy limited to Cochrane Library, PubMed, and Embase 2. Inclusion of all trials identified |
| Publication language | 1. Only inclusion of trial reports published in English 2. Inclusion of trial reports published in any language |
| Publication status | 1. Exclusion of unpublished trial reports 2. Inclusion of published and unpublished trial reports |
| Publication date | 1. Exclusion of studies published before 1990 2. Exclusion of studies published before 1995 3. Inclusion of all studies regardless of publication date |
| Type of trial | 1. Only inclusion of RCT’s 2. Inclusion of both RCT’s and prospective NRSI 3. Inclusion of both RCT’s and all types of NRSI |
| Sample size | 1. Exclusion of trials with a sample size less than 10 2. Exclusion of trials with a sample size less than 15 3. Inclusion of all trials regardless of sample size |
| The average age of participants | 1. Exclusion of trials with an average age below 16 2. Exclusion of trials with an average age below 18 3. Exclusion of trials with an average age below 50 years 4. Exclusion of trials with an average age below 55 years 5. Exclusion of trials with an average age below 60 years 6. Exclusion of trials with an average age below 65 years 7. Inclusion of all trials regardless of age distribution |
| Fracture classifications | 1. Only trials reporting on 3- or 4-part fractures 2. Inclusion of trials reporting on any type of fracture |
| Time of intervention | 1. Exclusion of trials including patients treated more than 48 hours post-injury 2. Exclusion of trials including patients treated more than 2 weeks post-injury 3. Exclusion of trials including patients treated more than 3 weeks post-injury 4. Inclusion of all trials regardless of time to treatment |
| Documentation of interventions | 1. Exclusion of trials without documentation for the specific operative technique and protocol for conservative treatment 2. Inclusion of all trials regardless of documentation of interventions |
| Time to follow up | 1. Exclusion of trials with less than 6 months of follow-up 2. Exclusion of trials with less than 12 months of follow-up 3. Inclusion of all trials regardless of time to follow-up |
| Loss to follow-up | 1. Exclusions of trials with a loss to follow-up greater than 15% 2. Inclusion of all trials regardless of loss to follow-up |
| **Choices related to outcome and analytical method** | |
| Outcome measures: functional outcome | 1. All PROMs analyzed in the same meta-analysis\* 2. PROMs and CS analyzed in the same meta-analysis\* 3. Each functional outcome measure analyzed in a separate meta-analysis |
| Outcome measures: HRQoL | 1. All types of HRQoL outcomes analyzed in the same meta-analysis 2. Each type of HRQoL assessment analyzed in a separate meta-analysis |
| Outcome metric | 1. Only outcomes reported as a change from baseline are included 2. Only outcomes reported as a difference at follow-up are included |
| Timepoint of assessment | 1. Data from the longest follow-up time for each trial is used for meta-analysis 2. Data from identical follow-up times are analyzed together 3. Data from differing follow-up times are grouped as short-, medium- and long-term with separate meta-analyses conducted for each group |
| Method of outcome analysis | 1. Only outcomes reported using intention to treat analyses are included 2. Only outcomes reported using per-protocol analyses are included 3. All outcomes are included regardless of the reported method of analysis† |
| Subgroup analyses: intervention | 1. Only trials using plate fixation 2. Only trials using tension band fixation 3. Only trials using either plate or tension band fixation 4. Only trials using an intramedullary nail 5. Only trials using hemiarthroplasty 6. Only trials using reverse hemiarthroplasty 7. Only trials using either hemiarthroplasty or reverse hemiarthroplasty 8. Inclusion of all trials regardless of intervention |
| Subgroup analyses: Fracture classifications | 1. Only trials with predominantly 2-part fractures 2. Only trials with predominantly 3-part fractures 3. Only trials with predominantly 4-part fractures 4. Only trials with predominantly 3- or 4-part fractures 5. All trials regardless of fracture classifications |
| Subgroup analyses: Risk of bias | 1. Exclusion of trials with a high risk of bias 2. Exclusion of trials with a high or moderate risk of bias 3. Inclusion of all trials regardless of the risk of bias |
| Missing outcome data | 1. Exclusion of trials with missing outcome data 2. Inclusion of trials with missing outcome data where imputation is possible |
| Meta-analysis model | 1. Fixed effect model 2. Random effect model |

\* A trial can only contribute with one outcome measure per meta-analysis. Therefore, if a trial reports multiple eligible outcome measures (e.g reports both CS and a PROM or two PROMs) the algorithm will simulate multiple methodological scenarios with each scenario only including one of the reported outcomes.

† If a trial reports outcomes using multiple methods of analysis, intention-to-treat analyses will be preferred over other methods.

#### Sensitivity analyses

We will conduct a sensitivity analysis by removing meta-analyses with high heterogeneity conducted using a fixed-effect model. High heterogeneity will be defined as I2 > 50% or Cochrane Q < 0.10. Furthermore, we will conduct sensitivity analyses by removing NRSI only reporting outcome measures not adjusted for confounding.

### Perspective

By investigating the overall VoE of meta-analyses comparing surgical and non-surgical interventions in the primary treatment of PHFs we will be able to identify specific methodological scenarios leading to conflicting results. Thereby, we will gain new insights into how methodological choices, and the combinations thereof, can result in meta-analyses reporting conflicting results, despite having access to the same trial reports.

Furthermore, we will be able to investigate the residual VoE of pre-defined protocols for published meta-analyses and identify the methodological choices responsible. This will provide new insights with regards to the needed level of specificity when developing a meta-analysis protocol to minimize any residual VoE.

## Conflicts of interest

None.

## References

[1] R.B. Beks, Y. Ochen, H. Frima, D.P.J. Smeeing, O. van der Meijden, T.K. Timmers, D. van der Velde, M. van Heijl, L.P.H. Leenen, R.H.H. Groenwold, R.M. Houwert, Operative versus nonoperative treatment of proximal humeral fractures: a systematic review, meta-analysis, and comparison of observational studies and randomized controlled trials, Journal of Shoulder and Elbow Surgery. 27 (2018) 1526–1534.

[2] B.S. Fu, H.L. Jia, D.S. Zhou, F.X. Liu, Surgical and Non-Surgical Treatment for 3-Part and 4-Part Fractures of the Proximal Humerus: A Systematic Review of Overlapping Meta-Analyses, Orthopaedic Surgery. 11 (2019) 356–365.

[3] T. Fu, C. Xia, Z. Li, H. Wu, Surgical versus conservative treatment for displaced proximal humeral fractures in elderly patients: A meta-analysis, International Journal of Clinical and Experimental Medicine. 7 (2014) 4607–4615.

[4] H.H. Handoll, S. Brorson, Interventions for treating proximal humeral fractures in adults, Cochrane Database of Systematic Reviews. (2015).

[5] J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, V. (editors) Welch, Cochrane Handbook for Systematic Reviews of Interventions, 2nd ed., John Wiley & Sons, Chichester (UK), 2019.

[6] J.P.A. Ioannidis, Why most discovered true associations are inflated, Epidemiology. 19 (2008) 640–648.

[7] Z. Jia, W. Li, Y. Qin, H. Li, D. Wang, C. Zhang, Q. He, D. Ruan, Operative Versus Nonoperative Treatment for Complex Proximal Humeral Fractures: A Meta-analysis of Randomized Controlled Trials, Orthopedics. 37 (2014) e543–e551.

[8] W. Li, G. Ding, J. Liu, J. Shi, C. Zhang, Q. Gao, [Operative versus non-operative treatment for three- or four-part proximal humeral fractures in elderly patients: a meta-analysis of randomized controlled trials]. [Chinese], 45 (2016) 28247610.

[9] Y. Li, L. Zhao, L. Zhu, J. Li, A. Chen, Internal Fixation Versus Nonoperative Treatment for Displaced 3-Part or 4-Part Proximal Humeral Fractures in Elderly Patients: A Meta-Analysis of Randomized Controlled Trials, PLoS ONE. 8 (2013).

[10] F. Mao, D.-H. Zhang, X.-C. Peng, Y. Liao, Comparison of Surgical versus Non-Surgical Treatment of Displaced 3- and 4-Part Fractures of the Proximal Humerus: A Meta-Analysis, Journal of Investigative Surgery. 28 (2015) 215–224.

[11] Z. Mao, L. Zhang, L. Zhang, X. Zeng, S. Chen, D. Liu, Z. Zhou, P. Tang, Operative Versus Nonoperative Treatment in Complex Proximal Humeral Fractures, Orthopedics. 37 (2014) e410–e419.

[12] E. Mayo-Wilson, N. Fusco, T. Li, H. Hong, J.K. Canner, K. Dickersin, Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis, Journal of Clinical Epidemiology. 86 (2017) 39–50.

[13] E. Mayo-Wilson, T. Li, N. Fusco, L. Bertizzolo, J.K. Canner, T. Cowley, P. Doshi, J. Ehmsen, G. Gresham, N. Guo, J.A. Haythornthwaite, J. Heyward, H. Hong, D. Pham, J.L. Payne, L. Rosman, E.A. Stuart, C. Suarez-Cuervo, E. Tolbert, C. Twose, S. Vedula, K. Dickersin, Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy, J Clin Epidemiol. 91 (2017) 95–110.

[14] C.M. Navarro, A. Brolund, C. Ekholm, E. Heintz, E.H. Ekström, P.O. Josefsson, L. Leander, P. Nordström, L. Zidén, K. Stenström, Treatment of humerus fractures in the elderly: A systematic review covering effectiveness, safety, economic aspects and evolution of practice, PLoS ONE. 13 (2018).

[15] M.J. Page, J.E. McKenzie, M. Chau, S.E. Green, A. Forbes, Methods to select results to include in meta-analyses deserve more consideration in systematic reviews, Journal of Clinical Epidemiology. 68 (2015) 1282–1291.

[16] M.J. Page, J.E. McKenzie, J. Kirkham, K. Dwan, S. Kramer, S. Green, A. Forbes, Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions, Cochrane Database of Systematic Reviews. (2014).

[17] C. Palpacuer, K. Hammas, R. Duprez, B. Laviolle, J.P.A. Ioannidis, F. Naudet, Vibration of effects from diverse inclusion/exclusion criteria and analytical choices: 9216 different ways to perform an indirect comparison meta-analysis, BMC Medicine. 17 (2019) 1–13.

[18] C.J. Patel, B. Burford, J.P.A. Ioannidis, Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations, Journal of Clinical Epidemiology. 68 (2015) 1046–1058.

[19] S. Rabi, Operative vs non-operative management of displaced proximal humeral fractures in the elderly: A systematic review and meta-analysis of randomized controlled trials , World Journal of Orthopedics. 6 (2015) 838.

[20] S. Sabharwal, N.K. Patel, D. Griffiths, T. Athanasiou, C.M. Gupte, P. Reilly, Trials based on specific fracture configuration and surgical procedures likely to be more relevant for decision making in the management of fractures of the proximal humerus, Bone & Joint Research. 5 (2016) 470–480.

[21] S. Serghiou, C.J. Patel, Y.Y. Tan, P. Koay, J.P.A. Ioannidis, Field-wide meta-analyses of observational associations can map selective availability of risk factors and the impact of model specifications, Journal of Clinical Epidemiology. 71 (2016) 58–67.

[22] J.Q. Song, X.F. Deng, Y.M. Wang, X.B. Wang, X. Li, B. Yu, Operative vs. nonoperative treatment for comminuted proximal humeral fractures in elderly patients: A current meta-analysis, Acta Orthopaedica et Traumatologica Turcica. 49 (2015) 345–353.

[23] B. Tendal, J.P.T. Higgins, P. Jüni, A. Hróbjartsson, S. Trelle, E. Nüesch, S. Wandel, A.W. Jørgensen, K. Gesser, S. Ilsøe-Kristensen, P.C. Gøtzsche, Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study, BMJ. 339 (2009).

[24] B. Tendal, E. Nüesch, J.P.T. Higgins, P. Jüni, P.C. Gøtzsche, Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study, BMJ. 343 (2011).

[25] X. Wan, W. Wang, J. Liu, T. Tong, Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range, BMC Medical Research Methodology. 14 (2014) 135.

[26] L. Xie, F. Ding, Z. Zhao, Y. Chen, D. Xing, Operative versus non-operative treatment in complex proximal humeral fractures: a meta-analysis of randomized controlled trials, SpringerPlus. 4 (2015) 1–9.