Critical Review of Articles: What I Look For?

Andrew Schmidt, MD
Faculty, Hennepin County Medical Center
Professor, Univ. of Minnesota

HennepinHealthcare

Why can’t I just read the conclusion?

Peer Review is not Perfect

- Reviewers have their own biases, and are subject to the same cognitive heuristics as everyone else
  - Confirmation bias
  - Framing
  - Counterfactual thinking

Do your own assessment…

Critical Literature Review

- What does this article *really* say?
- How does it apply to my practice / my patient / me ?

0022-5282/93/3404-0540$03.00/0
THE JOURNAL OF TRAUMA

Vol. 34, No. 4
Printed in U.S.A.

PRIMARY INTRAMEDULLARY FEMUR FIXATION IN MULTIPLE TRAUMA PATIENTS WITH ASSOCIATED LUNG CONTUSION—A CAUSE OF POSTTRAUMATIC ARDS?

Hans-Christoph Pape, MD, Michael Auf'm'Kolk, MD, Thomas Paffrath, MD, Gerd Regel, MD, Johannes A. Sturm, MD, and Harald Tscherne, MD

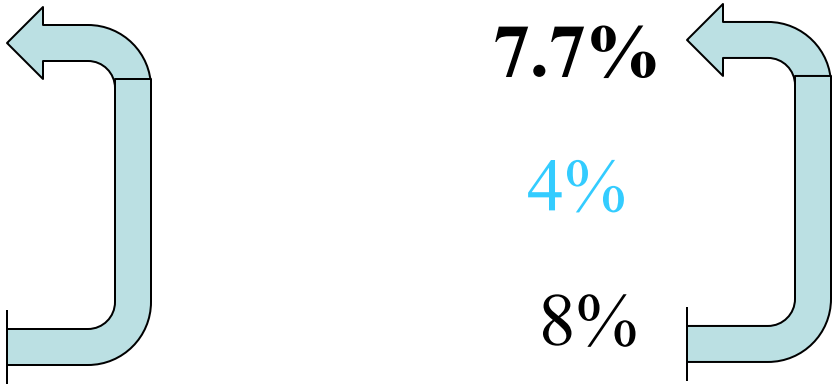
- **106 Patients**
  - ISS > 18
  - Nailed Femur Fx
  - Admitted within 8 hrs of injury
  - No death from head injury or shock
- **4 Groups**
  - With or without chest injury
  - Early or delayed femur stabilization (24 hrs)

With Chest Trauma

	Early Fixation	Late Fixation
Number	24	26
ISS	29.4	31.4
ARDS	33%	7.7%
Mortality	21%	4%

With Chest Trauma

	Early Fixation	Late Fixation
Number	24	26
ISS	29.4	31.4
ARDS	**33%**	**7.7%**
Mortality	21%	4%
Bilat Pulmonary Contusion	29%	8%



Is it the femur fracture?
Is it reaming?
Is it chest trauma?

Signs of a good paper

- There is a single hypothesis
- There is a primary outcome that can be measured reliably and is clearly related to the hypothesis
- If the outcome is a functional or QOL measure, the Minimal Clinically Important Difference (MCID) is defined.

Signs of a good paper

- The study is powered on the primary outcome
- Effect size large (mean diff/std dev).
- All secondary outcomes are pre-specified.
- Statistical comparisons are limited
- Confidence intervals are used
- Data is not presented for the first time in the conclusions

Signs of a good paper

- The conclusion addresses the hypothesis and is based on the primary outcome
  - Secondary outcomes remain secondary
  - Type II errors are identified
  - P values are not over-emphasized

Example
(excerpted from a real paper)

“Purpose: To review published literature assessing clinical outcomes and complication rates of XXX.”

“Methods: … searched PubMed and EMBASE for articles …

The inclusion criteria were studies of Level I to IV evidence in English, published in 1970 or later, involving human subjects, reporting clinical outcomes of XXX, including at least

1 patient.”

- “Results: Among the 3040 combined total results, 11 primary studies met our inclusion criteria. Adequate and safe XXX was achieved ... No significant difference was found in terms of complications such as infection and non-union.”

- “Conclusions: The comparative efficacy and safety of XXX is relatively equal to the YYY.”

???

Critical Review

- Strength of Evidence
- Prospective, randomized
  - Randomization concealed
  - Treatment blinded
  - Independent outcome assessment
- Case - control
  - Historical
  - Matched
- Case series – no control

Critical Review

- Does it apply to my practice?
  - Similar Institution?
  - Are the patients like mine?
  - Are the surgeons like me?

What factors are important?

- Rate of complications
- Re-operations
- Functional outcome
- Radiographic results
- Cost

Effect Size

- Absolute risk difference in treatments
  - 2 % vs 2.1 %
  - 2 % vs 4%
  - 2% vs 20%

ORIGINAL ARTICLE

Implementing 1-Dose Antibiotic Prophylaxis for Prevention of Surgical Site Infection

Silvia Nunes Szente Fonseca, MD, MPH; Sônia Regina Melon Kunzle, RN; Maria José Junqueira, RN; Renata Teodoro Nascimento, MD; José Ivan de Andrade, MD, PhD; Anna S. Levin, MD, PhD

Hypothesis: Replacing a 24-hour regimen with a 1-dose antibiotic prophylaxis for elective surgery would not increase rates of surgical site infection and would decrease costs.

Design and Setting: Before-after trial in a tertiary, private general hospital in Ribeirão Preto, São Paulo, Brazil.

Patients: Surgery was performed on 6140 consecutive patients from February 2002 through October 2002 (period 1) and 6159 consecutive patients from December 2002 through August 2003 (period 2). Studied surgeries included orthopedic, gastrointestinal, urology, vascular, lung, head and neck, heart, gynecologic, oncology, colon, neurologic, and pediatric surgeries. The study excluded patients with infection at the time of surgery.

Intervention: Decreasing the 24-hour prophylactic antibiotic regimen to 1-dose antibiotic prophylaxis.

Main Outcome Measures: Surgical site infections in both periods measured by in-hospital surveillance and postdischarge surveillance; compliance with 1-dose prophylaxis; and costs with cephazolin.

Results: We followed up 12 299 patients during their hospital stay; postdischarge surveillance increased significantly from 2717 patients (44%) to 3066 patients (50%, $P<.001$). One-dose prophylaxis was correctly followed in 6123 patients (99% compliance).The rate of surgical site infection did not change in either period (2% and 2.1% respectively, $P=.67$). The number of cephazolin vials purchased monthly decreased from 1259 to 467 with a corresponding monthly savings of $1980.

Conclusions: One-dose antibiotic prophylaxis did not lead to an increase in rates of surgical site infection and brought a monthly savings of $1980 considering cephazolin alone. High compliance to 1-dose prophylaxis was achieved through an educational intervention encouraged by the hospital director and administrative measures that reduced access to extra doses.

Arch Surg. 2006;141:1109-1113

Table 1. Surgical Procedures and Surgical Site Infections Among the Different Specialties

Type of Surgery	No. (%) Period 1 (n = 6140)	No. (%) Period 2 (n = 6159)	*P* Value
Orthopedic			
Surgeries	1615 (26)	1760 (28.5)	.42
SSIs	37 (2.3)	49 (2.8)	

Authors conclusion: "*One-dose antibiotic prophylaxis did not lead to an increase in rates of surgical site infection and brought a monthly savings of $1980 considering cephazolin alone.*"

Table 1. Surgical Procedures and Surgical Site Infections Among the Different Specialties

Type of Surgery	No. (%) Period 1 (n = 6140)	No. (%) Period 2 (n = 6159)	*P* Value
Orthopedic			
Surgeries	1615 (26)	1760 (28.5)	.42
SSIs	37 (2.3)	49 (2.8)	

What do you conclude?

- For infection in orthopedic surgeries, 2.3% v 2.8% = absolute risk difference of .5%, or .005.

- NNT = 1/.005 = 200

- 200 patients given 24 hours of Ancef rather than 1 dose prevents one FRI/PJI

Is that a clinically relevant difference !?!

"Statistical Difference" Does NOT Mean Clinically Important! An Evaluation of Orthopaedic Trauma Randomized Trials

Jinsil Sung; Judith Siegel; Paul Tornetta; Mohit Bhandari

- “A large percentage of RCTs that report "statistically different" results are likely NOT "clinically important."

OTA 2006

Lack of statistical difference Does NOT Mean there is no important difference!

- A large number of seemingly high-level papers there is no difference among outcomes that appear different and clinically relevant.

Type II statistical errors are rarely acknowledged

Distal femoral replacement for the treatment of periprosthetic distal femoral fractures around a total knee arthroplasty: a meta-analysis

Matthew J Wood[1], Talal Al-Jabri[2], Thomas Stelzhammer[3], Angela Brivio[4], James Donaldson[1], John A Skinner[1], David Barrett[4]

[1] Joint Reconstruction Unit, Royal National Orthopaedic Hospital, [2] Trauma and Orthopaedic Surgery, Imperial College London, [3] Trauma and Orthopaedic Surgery, Whittington Hospital, [4] Trauma and Orthopaedic Surgery, King Edward VII Hospital

Keywords: Periprosthetic fracture, locking compression plating, distal femoral replacement, knee replacement, arthroplasty, knee

https://doi.org/10.52965/001c.94574

Orthopedic Reviews

Vol. 16, 2024

Conclusion

This meta-analysis shows no statistically significant difference in the rates of re-operation, infection, mortality or return to pre-operative ambulatory status when comparing distal femoral replacement to locking compression plate.

Distal femoral replacement for the treatment of periprosthetic distal femoral fractures around a total knee arthroplasty: a meta-analysis

Matthew J Wood[1], Talal Al-Jabri[2], Thomas Stelzhammer[3], Angela Brivio[4], James Donaldson[1], John A Skinner[1], David Barrett[4]

[1] Joint Reconstruction Unit, Royal National Orthopaedic Hospital, [2] Trauma and Orthopaedic Surgery, Imperial College London, [3] Trauma and Orthopaedic Surgery, Whittington Hospital, [4] Trauma and Orthopaedic Surgery, King Edward VII Hospital

Keywords: Periprosthetic fracture, locking compression plating, distal femoral replacement, knee replacement, arthroplasty, knee

https://doi.org/10.52965/001c.94574

Orthopedic Reviews

Vol. 16, 2024

Conclusion

This meta-analysis shows no statistically significant difference in the rates of re-operation, infection, mortality or return to pre-operative ambulatory status when comparing distal femoral replacement to locking compression plate.

Results

Five articles met the inclusion criteria, reporting on 345 periprosthetic fractures. The overall rates of complications for distal femoral replacement and locking compression plate were: re-operation (6.1% vs 12.1%), infection (3.0% vs 5.3%), mortality (19.7% vs 19.3%), and return to pre-operative ambulatory status (60.9% vs 71.8%) (respectively).

Statistical Fragility

- A study's result is "fragile" if it would take only a few events occurring (or not occurring) to cause the statistical significance of that finding to change

	+ Complication	- Complication
IMN	1	24
Locking Plate	8	18
P Value		0.02

	+ Complication	- Complication
IMN	2	23
Locking Plate	8	18
P Value		0.08

FIGURE 2. Demonstration of the flip to reverse statistical significance with resultant FI = 1.

Parisien et al. / *Orthop Trauma* • Volume 33, Number 12, December 2019

Fragility Index

- Calculated by incrementally converting patients who did not have an event of interest into patients who had an event of interest until the p value changes from significant to nonsignificant.

- Fragility Quotient = FI / sample size

HIP International

The statistical fragility of arthroplasty versus fixation for femoral neck fractures: a systematic review of randomised controlled trials

HIP International
2025, Vol. 35(4) 410–417

DOI: 10.1177/11207000251343279
journals.sagepub.com/home/hpi
Sage

Auston R Locke[1], Avanish Yendluri[1], John J Corvi[1], John K Cordero[1], Michael N Megafu[2], Anya Wang[1], Steven J Yacovelli[1], David A Forsh[1], Calin S Moucha[1], Paul Tornetta 3rd[3] and Robert L Parisien[1]

- 985 RCTs were screened with 9 studies included in the analysis comprising 30 total outcomes

- The median FI across all outcomes was 5.
- In 11/30 of study outcomes, the number of patients lost to follow-up was greater than or equal to the median FI.

- Statistical findings in RCTs comparing arthroplasty to fixation for femoral neck fractures exhibit considerable fragility; minor changes in patient follow-up or outcome occurrences could significantly impact results.

J Orthop Trauma • Volume 33, Number 12, December 2019

ORIGINAL ARTICLE

Statistical Significance in Trauma Research: Too Unstable to Trust?

Robert L. Parisien, MD, Jesse Dashe, MD,* Patrick K. Cronin, MD,† Mohit Bhandari, MD, PhD,‡ and Paul Tornetta III, MD**

- All fx studies JBJS and JOT 1991-2013
- 198 comparative studies 775 total outcome events
  - 235 were initially reported as significant.
  - The median FI for the entire study was only 5 with an associated FQ of 0.046.
  - This represents just 3.8% of the total study population.

Final Thoughts

- Many papers state their goal is to study one thing, but they report on another
- Association is not causality
- Beware of conclusions based on p-values, especially small ones

JBJS REVIEWS 2020;8(1):e0079

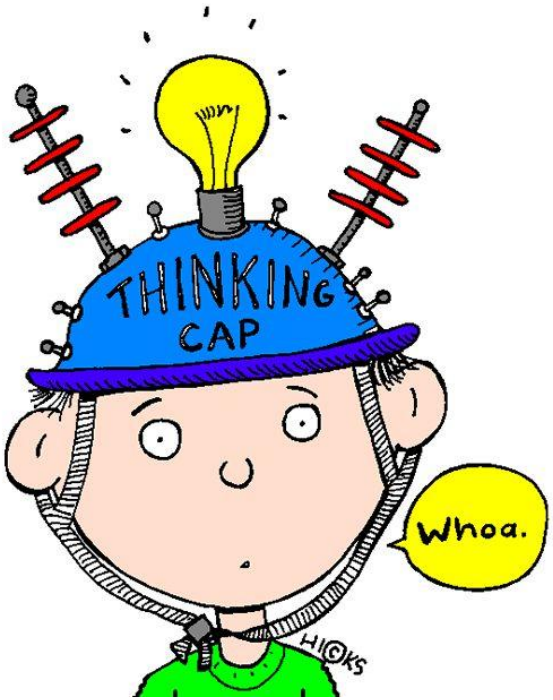
Revisiting the Sample Size and Statistical Power of Randomized Controlled Trials in Orthopaedics After 2 Decades

Aleksi Reito, MD, PhD
Lauri Raittio, BM
Olli Helminen, MD, PhD

- 233 RCTs from 7 major journals
- 136 reported a negative outcome; none had sufficient power to detect a small effect size, less than 1/3 powered to detect medium effect size
- < half were powered to detect the MCID

Conclusion

- Reading papers is not easy.
- Have a critical/adventurous mindset - there is something to “find” in every paper…



Thank You