# Randomized Controlled Trial Comparing Surgical Interventions

# 11

Max Solow, Raman Mundi, Vickas Khanna
and Mohit Bhandari

## Introduction

Despite years of training and schooling, clinicians routinely encounter difficult clinical scenarios that go beyond the scope of their everyday practice. To answer such questions, physicians often rely on high quality, research evidence—or in other terms, evidence-based practice. With high-quality research being integral to the practice of evidence-based medicine, it is to no surprise that the National Institutes of Health, in the United States, invests nearly $37.7 billion annually toward funding medical research projects [1].

Historically, physicians made therapeutic decisions based on anecdotal reports and personal experience. More recently, guidelines have been endorsed to eliminate variability among patient care and to highlight the importance of evidence from clinical research [2, 3]. These guidelines typically emphasize randomized controlled trials (RCTs) as the optimal study design for eliminating bias and capturing the truest estimates of treatment effect. Yet, not all RCTs are equal in terms of validity. Simply stating that a study was randomized does not ensure it is a high-quality study. The purpose of this chapter is to introduce strategies for clinicians to use while evaluating the evidence, with a particular focus on RCTs comparing surgical interventions. Arguably, conducting an RCT in a surgical setting poses some unique challenges, such as with blinding and the surgical learning curve that if not properly accounted for can lead to potential confounders. We will illustrate how physicians may apply these strategies and identify potential confounders with the help of a clinical scenario.

## Scenario

A healthy 74-year-old woman comes into your clinic complaining of pain and a palpable mass on the lateral aspect of her right thigh. A year prior to her presentation, she had undergone open reduction and internal fixation with a sliding hip screw for a femoral neck fracture of her right femur due to a fall. On the anteroposterior radiograph, nonunion, deformed femoral neck, and implant migration are appreciated. You conclude that her implant has failed and she would likely benefit from a revision surgery. You discuss with her the nature of her problem, as well as the risks and benefits of further surgery. Your patient agrees to the procedure but asks you

M. Solow
St. George's University School of Medicine,
St. George's, West Indies, Grenada

R. Mundi · V. Khanna · M. Bhandari (✉)
Department of Surgery, Division of Orthopaedic
Surgery, McMaster University, Hamilton,
ON, Canada
e-mail: bhandam@mcmaster.ca

R. Mundi
e-mail: Raman.mundi@medportal.ca

"If I had my initial fracture fixed with a different implant, would it still have failed?" Unsure initially of how to answer this question, you decide to review the literature and assure her that at her next clinical appointment, you will provide her with the best possible answer.

## Literature Search

You begin your literature search by creating a well-defined research question that encompasses several aspects of the clinical scenario. Using the PICO(T) format [4], which incorporates information about the patient population (P), the intervention (I), comparative interventions (C), the outcome (O), and time period (T), you generate the following research question: In femoral neck fracture patients, does fixation with a sliding hip screw lead to higher revision rates compared with other methods of fixation? Using Medline, the National Library of Medicine's PubMed database [5], you enter "femoral neck fracture" AND "reoperation" in the search field. You limit your search to English articles, published in the last 3 years, clinical trials, and human subjects. The search yields five articles [6–10]. Three of the articles do not compare fixation methods [7, 8, 10] and one article compares hemiarthroplasty to internal fixation [6]. The article, "Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre randomized controlled trial" compares two methods of femoral neck fracture fixation and seems to address your clinical question.

## Summary of the Appraised Article

The FAITH trial [9] was conducted to compare the sliding hip screw to cancellous screw fixation for patients with low-energy femoral neck fractures. The study was an international, multicentre, and randomized controlled trial, which included 1108 patients across 81 clinical centers from 8 countries. Following randomization, 557 patients were assigned to receive a sliding

hip screw and 551 patients to receive cancellous screws. The primary outcome of this study was the need for reoperation within 24 months. Other important outcomes included mortality, fracture healing, complications, and health-related quality of life scores. The mean length of follow-up was 633 days (standard deviation [SD] 208 days).

Reoperations within 24 months, mortality, fracture healing, implant failures, nonunions, infections, medically related adverse advents, and health-related quality of life scores did not differ in either treatment arm. However, more cases of avascular necrosis were observed in the sliding hip screw group than in the cancellous screws group (50 patients [9%] vs. 28 patients [5%]; HR 1.91, 1.06–3.44; $p = 0.0319$). Likewise, prespecified subgroup analyses showed that sliding hip screws are favored in patients with displaced fractures, fractures at the base of the femoral neck, and in patients who currently smoke.

## Evaluating a Randomized Controlled Trial

When reading a research article, it is always important to critique whether or not the study was carried out in a manner that would produce reliable results. Notably, three questions should be asked: Are the results valid? What are the results? And are the results applicable to my practice (see Box 1) [11]?

---

**Box 1. Important points to consider when evaluating a surgical RCT**

*Are the results valid?*
- Was the learning curve taken into consideration?
- How was randomization and allocation concealment performed?
- Who was blinded?
- Were the patient groups similar to one another?
- How were patients' data analyzed?

- Were treatments standardized and all patients accounted for?

*What are the results?*

- What impact did the treatment have?
- How precise were the results?

*Are the results applicable to my practice?*

- Do the results apply to my patient population?
- How clinically relevant are the results?
- How do these results impact me?

This list was modified from [11].

## Are the Results Valid?

### Learning Curve and Expertise-Based Randomized Controlled Trials

Compared with drug trials, surgical RCTs require special considerations. To begin with, a study comparing a "novel" intervention may need to factor in the surgical learning curve. A learning curve refers to the fact that a surgeon's proficiency, efficiency, and expectant outcomes of a procedure will improve with experience [12, 13]. Simply speaking, the more cases a surgeon has completed, the better they will become with the technical aspects of the procedure. Likewise, as experience increases, surgeons will have a better understanding of the necessary adjunctive medications, the appropriate patient selection, and the necessary pre- and postoperative care regimens to optimize outcomes. If an RCT does not consider (or mention) a learning curve, the results may be biased in favor of the traditional, or more common, intervention. One solution to this challenge is the "expertise based RCT", in which study patients are randomized directly to a surgeon, rather than the intervention. The surgeon, in turn, only delivers the intervention in which they are an expert [14]. Yet, these too are not without fault as determining expertise, achieving adequate recruitment, and the context-specific nature of such studies prove challenging [15, 16].

In the FAITH trial [9], surgeons had done at least 25 hip fracture fixation procedures during their career, with at least 5 fracture fixation procedures having been completed in the year prior to participation. With this in mind, we can assume that participating surgeons had sufficient expertise in performing either intervention.

### Patient Randomization and Allocation Concealment

Randomization is a technique that assigns patients to either the treatment or control arm of the study in a manner that is entirely by chance and without taking into consideration patient or researcher preference [17]. The randomization schedule can often be created by computer-generated sequences. The purpose of randomization is to produce groups that are similar to one another in terms of both known and unknown characteristics that may influence the study outcome. Without randomization, researcher, patient, and physician bias may influence experimental outcomes and alter study results.

Similarly, allocation concealment aims to limit selection bias by concealing which treatment arm each prospective study patient will be assigned to [17]. In essence, it is a method to protect the integrity of the randomization sequence. An example of allocation concealment is the use of a central call-in center or computer program, which will reveal the treatment arm a patient has been randomized to only after enrolment. As such, physicians cannot predict which treatment the patient will receive prior to enrolment and randomization. Despite proper randomization, failure to adequately control allocation concealment may introduce bias and influence study results. For example, if investigators opened unsealed assignment envelopes and channeled participants with a better prognosis to the experimental group, this would lead to larger treatment effects [18].

In the FAITH trial [9], randomization and allocation concealment were performed with a centralised computer system, which provides a methodologically robust approach for randomization and allocation concealment.

## Were Patients, Surgeons, and Researchers Aware of Group Assignment (Blinding)?

Within RCTs, blinding refers to the precautions taken to prevent the patient, surgeon, or researchers from knowing a participant's group assignment [19]. The importance of blinding is that it minimizes bias in intervention implementation, outcomes assessment, results analysis, and patient dropout. Take for example the placebo effect, a scenario where a patient's belief about their treatment influences their outcomes [20]. In a meta-analysis comparing osteoarthritis patients receiving a placebo to an untreated control group, patients in the placebo group experienced significantly more pain and stiffness relief than the control group [21]. If patients are aware of their group allocation they may alter their answers on quality assessments and subconsciously put forth more of an effort in their rehabilitation, exaggerating the experimental treatment effect. Similarly, surgeons may favor one intervention over another and may unintentionally be more precise during its implementation, ultimately overestimating study results [22, 23]. Unfortunately, the very nature of surgery makes it nearly impossible for surgeons to remain blinded during intervention implementation. Finally, the research personnel in charge of assessing the outcomes, if not blinded, may alter study results. If, for example, the researchers rounded outcomes from the treatment arm of the study upwards compared to the control group, or if preferential treatment were given to either group during rehabilitation or assessment, study results could be distorted [24, 25].

In the FAITH trial [9], both the surgeons and the patients were not blinded, while the data analysts were. Because both treatment strategies were similar from a patient experience perspective, unblinded patients likely would not have biased the study results. Furthermore, the primary outcome of reoperation was unlikely to be substantially altered by the lack of blinding of patients. Unfortunately, as there is no real solution to surgeon blinding, it is difficult to comment on whether or not differential care was provided to either care group. However, as this study was large and incorporated multiple surgeons from various countries, it may be fair to assume that any differences during procedure implementation would have been balanced between the two groups.

## Were the Patients in Each Group Similar?

At the onset of the trial, it is imperative that the experimental and control groups be relatively homogenous. In other words, the more alike the two groups are to one another prior to trial commencement the less likely other factors can influence study results and the easier it will be to detect the true effects of the therapeutic intervention. Most commonly, studies present patient baseline demographic information and known prognostic variables—often presented in a table. In the FAITH trial [9], patients were included if they were 50 years or older and sustained a low-energy femoral neck fracture that required operative fixation and baseline patient characteristics were presented (Table 11.1).

Prior to randomization, it is important to consider whether or not any other variables exist that could influence treatment responsiveness. Stratification, another measure to ensure group balance, divides study participants into homogenous subgroups from which they are then randomized into the different arms of the trial [26]. Consider a study comparing two different treatment modalities for first-time traumatic anterior shoulder dislocations. Since patient age correlates very strongly with the rate of repeated dislocation, it would be critical to stratify patients in each trial group based on age [27]. Otherwise, the group with a larger proportion of younger patients may underestimate the treatment effect. Likewise, it is important to consider the variability among individual surgeons and clinical sites as these can influence outcomes [28]. In the FAITH trial [9], patients were stratified by clinical site, however, the authors did not report whether this had an impact on outcomes.

**Table 11.1** Patient baseline characteristics [9]

|  |  | SHS | CS |
|---|---|---|---|
| Age | (Years) | 72.2 (12.0) | 72.0 (12.3) |
| Sex | Male | 212/535 (40%) | 210/535 (39%) |
|  | Female | 323/535 (60%) | 325/535 (61%) |
| Ethnic origin | Native | 1/533 (<1%) | 3/535 (1%) |
|  | South Asian | 65/533 (12%) | 65/535 (12%) |
|  | East Asian | 6/533 (1%) | 4/535 (1%) |
|  | Black | 22/533 (4%) | 18/535 (3%) |
|  | Hispanic | 3/533 (1%) | 1/535 (<1%) |
|  | White | 436/533 (82%) | 444/535 (83%) |
| Smoking history | Never smoked | 268/533 (50%) | 276/532 (52%) |
|  | Current smoker | 101/533 (19%) | 100/532 (19%) |
|  | Former smoker | 164/533 (31%) | 156/532 (29%) |
| Current drugs | None | 170/535 (32%) | 179/534 (34%) |
|  | NSAIDS | 86/535 (16%) | 64/534 (12%) |
|  | General cardiac | 167/535 (31%) | 167/534 (31%) |
|  | Opioid analgesics | 43/535 (8%) | 56/534 (10%) |
|  | Pulmonary drugs | 58/535 (11%) | 69/534 (13%) |
|  | Anti-hypertension drugs | 244/535 (46%) | 252/534 (47%) |
|  | Osteoporosis drugs | 67/535 (13%) | 73/534 (14%) |
| BMI | Underweight (BMI 18.5) | 37/530 (7%) | 33/528 (6%) |
|  | Normal weight (18.5–24.9) | 276/530 (52%) | 300/528 (57%) |
|  | Overweight (25–29.9) | 159/530 (30%) | 148/528 (28%) |
|  | Obese (30–39.9) | 58/530 (11%) | 47/528 (9%) |
| Fractured hip | Left | 280/535 (52%) | 281/535 (53%) |
|  | Right | 255/535 (48%) | 254/535 (47%) |
| Mechanism of injury | Fall | 515/533 (97%) | 521/534 (98%) |
|  | Spontaneous | 13/533 (2%) | 6/534 (1%) |
|  | Other low-energy trauma | 5/533 (1%) | 7/534 (1%) |
| History of surgery to affected hip | Yes | 3/535 (1%) | 0/535 (0%) |
|  | No | 532/535 (99%) | 535/535 (100%) |
| Additional injuries | Yes | 67/535 (13%) | 72/535 (13%) |
|  | No | 468/535 (87%) | 463/535 (87%) |

SHS = sliding hip screw; CS = cancellous screws; NSAIDS = non-steroidal anti-inflammatory drugs; BMI = body-mass index

## Intention-to-Treat Analysis

Once the trial has begun, it is important to consider how the analysis was carried out. Consider the hypothetical example of an RCT comparing the rates of surgical site infection in 200 patients who underwent total knee arthroplasty. In this study, 100 patients are assigned to receive intraoperative local vancomycin powder to their wound plus preoperative ancef (experimental group), while the remaining 100 only receive the preoperative ancef for prophylaxis (control group). The surgeon then decides that 10 patients
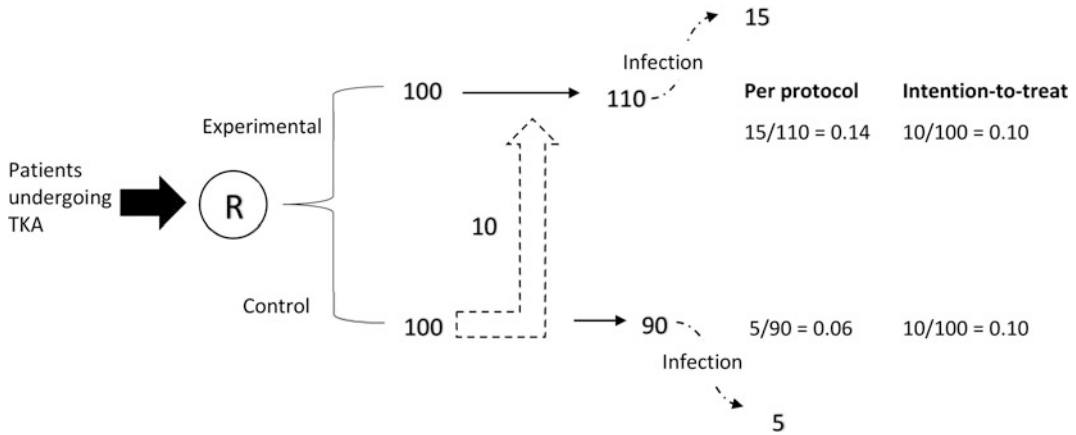
**Fig. 11.1** Bias introduced when using a per-protocol analysis versus an intention-to-treat analysis. Per-protocol analysis includes all those patients who received vancomycin, irrespective of their initial assignment. With this analysis, the event rate in the experimental group is more than double that of the control group. The intention-to-treat analysis analyzes patients in their original groups irrespective of the treatment they actually received. Under this model, we see that the event rates are equal. R = randomization; TKA = total knee arthroplasty

in the ancef-only group would benefit from the local vancomycin powder due to intraoperative complications. If 10 of the 100 patients in the initial experimental group develop an infection, plus 5 of the 10 patients that switched group's perioperatively go on to develop an infection, then the event rate would be 14%; however, the rate in the control group would be 6% (Fig. 11.1). These values represent a spurious reduction in infection rates for the control group. Intention-to-treat analysis eliminates this potential bias by analyzing patients in an RCT according to the treatment arm they were originally assigned to, irrespective of the treatment they actually received [29, 30]. In our example, if patients were analyzed with an intention-to-treat model the event rates would be 10/100 for both groups. Incorporating this method of analysis into a study eliminates any bias that may arise from participant attrition or crossover. In the FAITH trial [9], the authors state that an intention-to-treat analysis was employed.

### Treatment Standardization

Standardization of interventions is important in surgical trials to ensure treatment effects are not biased due to differential care between groups outside of the main surgical intervention (i.e., preoperative antibiotics, perioperative care, postoperative thromboprophylaxis, postoperative weight-bearing status, and rehabilitation). For example, differences in postoperative care could introduce bias if one group were given extra physiotherapy sessions. Standardization tries to address any bias that may be introduced in studies with multiple components by attempting to keep things as consistent as possible [31]. In the FAITH trial [9], patient positioning, fracture reduction, and surgical exposure were left to the surgeons' discretion. However, surgeons were given specific criteria for acceptability of post-fixation radiographic fracture alignment. The authors also provide supplemental materials in the appendix regarding procedural and rehabilitative standardization. Therefore, it is safe to assume that the necessary steps were taken to address any potential variability.

### Sample Size Calculation and Follow-Up

A predetermined sample size calculation is integral to the conduct of an RCT (see Chap. 29). A study with too few participants may fail to reach its objective by lacking capacity to answer the primary study question due to being statistically underpowered. Conversely, a study should not simply recruit an excessive number of patients as this would overpower the study and may result in findings of statistical significance that are not

actually clinically important [32]. As such, RCTs should have a sample size calculation based upon a determined minimally clinically important difference and desired study power [32]. Sample size calculations provide a study with adequate power to detect the minimally clinically important differences for outcomes used in the calculation. However, subgroup analysis is not included in this calculation.

Another caveat in any study is participant attrition. Patients that are lost to follow-up threaten the validity of the study since we do not know their outcomes, whether or not they died, or if demographic differences exist between them and the study group [33]. While there is not any hard and fast cutoff value at which attrition related bias becomes apparent, it is generally accepted that 5% loss to follow-up is of little worry, a loss of 20% should raise concern, and a loss between 5 and 20% may still produce bias but to a lesser degree [34]. To maintain the predetermined study power and avoid skewing study results, it is important to consider the anticipated loss to follow-up when calculating a trials minimum sample size in the planning phase of the study [35] (see Chap. 29 for further explanation). In the FAITH trial [9], the original sample size calculation found that enrolment of 1500 patients would give the trial a study power of 81.5%. However, upon reanalysis of completed follow-up data from the first 589 patients, it was found that a sample size of 1100 patients would provide 95.7% power to detect a relative risk reduction of 35%. The authors were able to recruit 1108 patients. The authors validate their study by providing details of sample size calculations and patient exclusion rationale in the appendix.

## What Are the Results?

### What Was the Impact from the Treatment?

Now that the study has been conducted, we want to know what the results mean and their clinical importance and statistical significance. Some common terminology used to report data are relative risk (RR), relative risk reduction (RRR), absolute risk reduction (ARR), and the number needed to treat (NNT) (Box 2). The RR describes the probability of an event occurring in one group of people versus another [36, 37] (see Chap. 6 for further explanation). In our study, the RR would tell us the probability of reoperation within 24 months in patients receiving a sliding hip screw (experimental) versus patients receiving cancellous screws (control). The RRR measures how much risk is reduced in the experimental group versus the control group [38]. Consider a hypothetical situation comparing stroke rates in hypertensive patients receiving an intensive treatment regimen (experimental) versus the standard of care (control). If 10% of the control group experienced a stroke, compared to 5% of patients in the treatment group, it can be said that the intensive treatment regimen resulted in a relative risk reduction of 50% (Table 11.2). The ARR describes the absolute difference in event rates between the control and treatment groups [38]. Using our hypothetical situation, the ARR for strokes would be 5%. The NNT refers to the number of patients that would need to be treated in order to prevent one additional adverse event from occurring and can be thought of as the inverse of the ARR [39, 40]. Referring to our hypothetical example once more, for every 20 patients treated with the intensive treatment regimen, 1 stroke would be prevented. In the FAITH trial [9], the authors report that reoperations of any kind within 24 months were roughly equal in the sliding hip screw group versus the cancellous screw group and there was no minimally clinically important difference or statistical significance (20% vs. 22%, $p = 0.18$); however, implant removal took place significantly less frequently in the sliding hip screw group than in the cancellous screws group although the findings were not clinically significant (5% vs. 9%, $p = 0.0009$). The authors also report that there were no statistically or clinically significant differences between the two groups in terms of mortality, fracture healing, complications, and health-related quality of life scores.

**Table 11.2** Sample calculations from the hypothetical hypertensive drug trial

|  | Experimental group | Control group |
|---|---|---|
| Total number of patients | 124 | 130 |
| Number of strokes | 6 | 13 |
| Event rates | $x = 6/124 = 0.05$ (5%) | $y = 13/130 = 0.10$ (10%) |
| Relative risk: $x/y = 0.05/0.10 = 0.50$ (50%) | | |
| Relative risk reduction: $(x - y)/y = (0.05 - 0.10)/0.10 = 0.50$ (50%) | | |
| Absolute risk reduction: $x - y = 0.05 - 0.10 = 0.05$ (5%) | | |
| Number needed to treat: $1/(x - y) = 1/(0.05 - 0.10) = 20$ | | |

Patients in the control group were 50% more likely to experience a stroke (RR). Similarly, the experimental intervention reduced the risk of stroke by 50% (RRR). Likewise, for every 100 patients treated with the experimental intervention, there would be 5 fewer strokes (ARR). Furthermore, in order to prevent one stroke, 20 patients would need to be treated with the experimental intervention (NNT).

---

**Box 2. Equations for common statistic terminology**

Relative Risk (RR)
$RR = \frac{x}{y}$

Relative Risk Reduction (RRR)
$RRR = \frac{(x-y)}{y}$

Absolute Risk Reduction (ARR)
$AAR = x - y$

Number Needed to Treat (NNT)
$NNT = \frac{1}{AAR}$

$x$—# of events in experimental group/total # of patients in experimental group.

$y$—# of events in control group/total # of patients in control group.

## How Precise Were the Results?

It is impossible to know the "true" reduction to reoperation rates within 24 months caused by the use of a sliding hip screw because of variables unbeknownst to researchers that may impact the study. The best we can do is come up with a close estimate, known as the point estimate, of the true value that would lie within that ballpark. To communicate the point estimate, researchers provide a variety of values, known as the confidence interval (CI), that specifies a probability within which one can be confident the true value lies [41, 42]. By convention, a 95% CI is generally used, meaning we can be 95% certain that the "true" value lies within the interval (see Chap. 28 for more information).

In the FAITH trial [9], results are reported using hazard ratios (HR), the chance of an event occurring in the treatment arm divided by the chance of the event occurring in the control arm [43], CIs, and $p$ values. They report an HR of 0.83 with a 95% CI of 0.63–1.09 and a $p$-value of 0.18. This means that the patients who received a sliding hip screw are less likely to require a reoperation and that the true value for this rate lies between 0.63 and 1.09. However, the difference in event rates between the two groups is not statistically or clinically significant.

## What Now?

### Generalizability

Before we decide to implement a particular intervention, we need to assess how relevant the information is to our patient population. In our clinical scenario, our patient is a 74-year-old woman who sustained a femoral neck fracture due to a fall. In the FAITH trial [9], they enrolled patients who were 50 years or older with a low-energy femoral neck fracture requiring operative fixation. Looking at the table of patient baseline characteristics, we also see that the mean age of patients was 72.1 years and that 61% of the study participants were female, therefore these findings could apply to our patient.

### Clinical Relevance

It is important to consider whether or not the outcomes that the researchers examined are clinically relevant. For example, a study comparing two treatment modalities for rotator cuff

tears that looked at "return to sport time" would be clinically useful for a 20-year-old pitcher, but less relevant for a 74-year-old patient. In the FAITH trial [9], the outcomes assessed included reoperation within 24 months, mortality, fracture healing, complications, and health-related quality of life scores. All of these outcome measures would be relevant to both physicians and patients alike, as they are clinically important with important implications on the ultimate outcome. For instance, revision surgeries tend to be more technically complex, impose added expenses to the health care system, and expose patients to further risks and potential perioperative morbidity [44].

## What Does This Mean for Me as a Healthcare Provider?

Now that we have assessed the evidence provided to us, how do we use it? For medical doctors, this may be straightforward because if a study shows that drug A is more efficacious than drug B, they can simply begin prescribing drug B. However, implementing new evidence poses potential challenges for surgeons. If a study shows that a particular intervention produces superior outcomes, how do they go about implementing it? Surgeons need to objectively critique their expertise and proficiency with a procedure and earnestly consider if they would be happy with the level of care provided. It would be unethical for a surgeon to perform a procedure they are not familiar with as patients would potentially experience a higher complication rate [45–47]. If a surgeon is not comfortable performing the procedure they can: (1) refer the patient to a colleague, (2) seek additional training, or (3) perform a different procedure after considering the evidence for it. Any of these options would be sufficient and the decision ultimately lies with each surgeon.

## Resolving Our Clinical Scenario

After thoughtfully examining the information provided to us from the FAITH trial [9], it is fair to conclude that either a sliding hip screw or cancellous screws would be acceptable treatments for low-energy femoral neck fractures in terms of the primary endpoints assessed (reoperation within 24 months, death, complications, and quality of life). Although our patient in the clinical scenario experienced an implant failure within a year, and will likely require reoperation, we can be confident in answering her that her failure likely was not due to the choice of implant.

## References

1. National Insititue of Health (NIH). What we do: budget [Internet]. 2018 Apr 11. [cited 2018 Aug 9]. Available from: https://www.nih.gov/about-nih/what-we-do/budget#note.
2. Graham R, Mancer M, Wolman Miller D, Greenfield S, Steinberg E, editors. Clinical practice guidelines we can trust. Washington, DC: The National Academies Press; 2011. p. 290.
3. Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, Hirsh J, et al. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268(17):2420–5.
4. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995;123(3):A12–3.
5. Motschall E, Falck-Ytter Y. Searching the MEDLINE literature database through PubMed: a short guide. Onkologie. 2005;28(10):517–22.
6. Lu Q, Tang G, Zhao X, Guo S, Cai B, Li Q. Hemiarthroplasty versus internal fixation in super-aged patients with undisplaced femoral neck fractures: a 5-year follow-up of randomized controlled trial. Arch Orthop Trauma Surg. 2017;137(1):27–35.
7. Bhandari M, Jin L, See K, Burge R, Gilchrist N, Witvrouw R, Krohn KD, Warner MR, Ahmad QI, Mitlak B. Does teriparatide improve femoral neck fracture healing: results from a randomized placebo-controlled trial. Clin Orthop Relat Res. 2016;474(5):1234–44.
8. Sprague S, Slobogean GP, Bogoch E, Petrisor B, Garibaldi A, O'Hara N, Bhandari M, FAITH Investigators. Vitamin D use and health outcomes after surgery for hip fracture. Orthopedics. 2017;40(5):e868–75.
9. Fixation using Alternative Implants for the Treatment of Hip fractures (FAITH) Investigators. Fracture fixation in the operative management of hip fractures (FAITH): an international, multicentre, randomised controlled trial. Lancet. 2017;389(10078):1519–27.
10. Watts CD, Houdek MT, Sems SA, Cross WW, Pagnano MW. Tranexamic acid safely reduced blood

loss in hemi- and total hip arthroplasty for acute femoral neck fracture: a randomized clinical trial. J Orthop Trauma. 2017;31(7):345–51.

11. Guyatt G, Rennie D, Meade MO, Cook DJ. User's guides to the medical literature: essentials of evidence-based clinical practice. 2nd ed. Hamilton, ON: McGraw-Hill Professional; 2008. p. 380.

12. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. Postgrad Med J. 2007;83(986):777–9.

13. Maruthappu M, Duclos A, Lipsitz SR, Orgill D, Carty MJ. Surgical learning curves and operative efficiency: a cross-specialty observational study. BMJ Open. 2015;5(3):e006679.

14. Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ. 2005;330(7482):88.

15. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. Trials. 2009;10:9.

16. Cook JA, Elders A, Boachie C, Bassinga T, Fraser C, Altman DG, et al. A systematic review of the use of an expertise-based randomised controlled trial design. Trials. 2015;16:241.

17. Kim J, Shin W. How to do random allocation (randomization). Clin Orthop Surg. 2014;6(1):103–9.

18. Schulz KF. Subverting randomization in controlled trials. JAMA. 1995;274(18):1456–8.

19. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. PLoS Med. 2006;3(10):e425.

20. Kaptchuk TJ. Powerful placebo: the dark side of the randomised controlled trial. Lancet. 1998;351(9117):1722–5.

21. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis. 2008;67(12):1716–23.

22. van der Linden W. Pitfalls in randomized surgical trials. Surgery. 1980;87(3):258–62.

23. Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, et al. Factors that limit the quality, number and progress of randomised controlled trials. Health Technol Assess. 1999;3(20):1–143.

24. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg. 2010;53(5):345–8.

25. Miller LE, Stewart ME. The blind leading the blind: use and misuse of blinding in randomized controlled trials. Contemp Clin Trials. 2011;32(2):240–3.

26. Randelli P, Arrigoni P, Lubowitz JH, Cabitza P, Denti M. Randomization procedures in orthopaedic trials. Arthroscopy. 2008;24(7):834–8.

27. Kirkley A, Werstine R, Ratjek A, Griffin S. Prospective randomized clinical trial comparing the effectiveness of immediate arthroscopic stabilization versus immobilization and rehabilitation in first

traumatic anterior dislocations of the shoulder: long-term evaluation. Arthroscopy. 2005;21(1):55–63.

28. Farrokhyar F, Bajammal S, Kahnamoui K, Bhandari M. Practical tips for surgical research. Ensuring balanced groups in surgical trials. Can J Surg. 2010;53(6):418–23.

29. Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. J Bone Joint Surg Am. 2009;91(9):2137–43.

30. Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ. 2001;165(10):1339–41.

31. Blencowe NS, Cook JA, Pinkney T, Rogers C, Reeves BC, Blazeby JM. Delivering successful randomized controlled trials in surgery: methods to optimize collaboration and study design. Clin Trials. 2017;14(2):211–8.

32. Zhong B. How to calculate sample size in randomized controlled trial? J Thorac Dis. 2009;1(1):51–4.

33. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. BMJ. 2006;332(7547):969–71.

34. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ. 2002;325(7365):652–4.

35. Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? Surgery. 2003;134(2):275–9.

36. Yelland LN, Salter AB, Ryan P. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. Int J Biostat. 2011;7(1):1–31.

37. Sistrom CL, Garvan CW. Proportions, odds, and risk. Radiology. 2004;230(1):12–9.

38. Irwig L, Irwig J, Trevena L, Sweet M. Smart health choices: making sense of health advice. London: Hammersmith Press; 2008. p. 1–242.

39. Haas M, Schneider M, Vavrek D. Illustrating risk difference and number needed to treat from a randomized controlled trial of spinal manipulation for cervicogenic headache. Chiropr Osteopat. 2010;18:9.

40. Bender R. Calculating confidence intervals for the number needed to treat. Control Clin Trials. 2001;22(2):102–10.

41. Flechner L, Tseng TY. Understanding results: P-values, confidence intervals, and number need to treat. Indian J Urol. 2011;27(4):532–5.

42. Sedgwick P. Randomised controlled trials: inferring significance of treatment effects based on confidence intervals. BMJ. 2014;349:g5196.

43. Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. J Clin Oncol. 2016;34(15):1813–9.

44. Bhandari M, Jon Smith, Larry E. Miller, Jon E. Block. Clinical and economic burden of revision

knee arthroplasty. Clin Med Insights Arthritis Musculoskelet Disord. 2012;(5):89–94.

45. Hammond JW, Queale WS, Kim TK, McFarland EG. Surgeon experience and clinical and economic outcomes for shoulder arthroplasty. J Bone Joint Surg Am. 2003;85–A(12):2318–24.

46. Jain N, Pietrobon R, Hocker S, Guller U, Shankar A, Higgins LD. The relationship between surgeon and hospital volume and outcomes for shoulder arthroplasty. J Bone Joint Surg Am. 2004;86–A(3):496–505.

47. Hervey SL, Purves HR, Guller U, Toth AP, Vail TP, Pietrobon R. Provider volume of total knee arthroplasties and patient outcomes in the HCUP-nationwide inpatient sample. J Bone Joint Surg Am. 2003;85–A(9):1775–83.