

# Evaluating Surgical Interventions

# 6

Aristithes G. Doumouras and Dennis Hong

Clinical decision-making in surgery has evolved dramatically in the past several years. Traditionally, surgeons coalesced anecdotal evidence, previous experience, expert opinion, and extrapolation of basic science into clinical decision-making. While this often served patients well, it is also possible that this decision-making process led to suboptimal treatment. To combat this, a plethora of surgical research has been ongoing for the past several decades, and the utilization and understanding of this evidence are vital to surgical practice in the twenty-first century [1]. The evidence itself can vary greatly in quality and therefore several grading systems have been developed to assess quality [2, 3]. Regardless of the grading system, the general hierarchy of evidence remains the same and is based on the propensity of a certain study type to introduce bias into its conclusions. Bias, specifically, refers to any systematic deviation from the truth, which could result in an underestimation or overestimation of the true effect of an intervention [4].

The lowest level of evidence after expert opinion consists of retrospective case series which present outcomes without comparison, and retrospective cohorts, which often are used to compare two or more therapies (see Chap. 5). While uncomplicated to design, the retrospective nature of these studies does not allow for the collection of all potential confounding factors. This limits the ability to infer causation from the conclusions. The next level of evidence consists of prospective, nonrandomized controlled trials. While the prospective nature of these studies allows for better data collection, the lack of randomization can still lead to an imbalance within groups for unmeasured variables that would not be accounted for fully with statistical analyses. Properly designed and conducted randomized trials provide a high level of internal validity to their conclusions but can be limited by their generalizability [3]. Accordingly, the highest level of evidence is when several randomized trials looking at the same topic can be combined into a meta-analysis [5, 6]. Meta-analyses combine the results of several trials in an effort to boost statistical power and minimize false negative results. However, one should be cautious of meta-analyses of observational data as *a meta-analysis of biased data is still biased* though these types of studies may be useful in summarizing the current observational evidence for a topic. Ideally, surgeons can evaluate the evidence they have for the clinical problem at hand and provide optimal care based on this assessment.

---

A. G. Doumouras  
Department of Surgery, McMaster University,  
Hamilton, ON, Canada  
e-mail: [aristithes.doumouras@medportal.ca](mailto:aristithes.doumouras@medportal.ca)

D. Hong (✉)  
Department of Surgery, St. Joseph's Healthcare,  
Hamilton, ON, Canada  
e-mail: [dennishong70@gmail.com](mailto:dennishong70@gmail.com)

## Clinical Scenario

You are a general surgeon on a surgical mission to a low-income country. In the course of an inguinal hernia repair requiring mesh, the junior resident who accompanied you on the trip asks you why you are using a sterilized mosquito net rather than polypropylene mesh, the mesh you use back home. You inform her that the typical mesh back home is very expensive and the mission hospital cannot afford it. After the case, you challenge her to review the literature and find out if there is a difference in effectiveness between these two approaches for the major outcomes of inguinal hernia repair such as complication rates and recurrence.

## Search Strategy

Your search is designed to cast a wide net and include all relevant articles while ultimately trying to focus on a few number of key articles. Accordingly, you search MEDLINE with a broad strategy that includes both the Medical Subject Headings (MeSH) “Herniorrhaphy,” “Hernia” and “Costs and Cost Analysis” and non-MeSH keywords “mesh” and “low-cost.” Using non-MeSH terms ensures the inclusion of all relevant abstracts. These search terms are combined with the “OR” Boolean function to create the initial database. From this, abstracts can be reviewed individually or, if still too numerous, the Boolean “AND” function can be used to narrow down the choices and create a more relevant set. For practical purposes, we can also limit it to the English language though this may introduce bias into systematic reviews. To further narrow down the list, we can also look directly for clinical trials (see Chap. 4). In doing this, we get four hits, and of these hits, one is a randomized trial with the desired clinical outcomes by Löfgren et al. [7]. The purpose of this chapter is to assist the reader to interpret the results/data presented in a surgical article. Therefore, we will appraise the Lofren et al. [7]

article and find if the authors’ conclusions are supported by the data they provide. The evaluation of surgical interventions follows the framework shown in Table 6.1.

When considering the evidence, there are three main aspects of a study that should be assessed: internal validity, the results, and external validity (Table 6.1). We will discuss each of these points individually and how to assess them. We will scrutinize the results of the Löfgren et al. [7], which, are presented in Table 6.2.

**Table 6.1** How to appraise an article evaluating surgical interventions [8]

Question	Appraisal
Are the results valid?	<ul style="list-style-type: none"> <li>– Was patient assignment randomized, and the randomization process “concealed”?</li> <li>– Were study personnel “blinded” to treatment and apart from the experimental intervention, were the groups treated equally?</li> <li>– Were all patients who entered the trial accounted for and was follow-up adequate?</li> <li>– Were patients analyzed according to the “intention to treat” principle?</li> <li>– Was the study large enough to detect a difference?</li> </ul>
What are the results?	<ul style="list-style-type: none"> <li>– What are the outcomes and how were they measured?</li> <li>– How large was the treatment effect?</li> <li>– How precise was the estimate of the treatment effect?</li> </ul>
Are the results applicable to my patients?	<ul style="list-style-type: none"> <li>– Were the study patients similar to my patients?</li> <li>– Were the measured outcomes clinically relevant?</li> <li>– Are my surgical skills similar to those of the study surgeons?</li> </ul>

**Table 6.2** Primary outcomes and mortality among study participants [7]

	Outcome	Low-cost mesh (n = 143)	Commercial mesh (n = 148)	Absolute difference		p-value
				Percentage points	95% CI	
Primary outcomes	<i>Hernia recurrence</i>	1 (0.7)	0	0.7	(−1.2 to 2.6)	1.0
	<i>Any postoperative complication</i>	44 (30.8)	44 (29.7)	1.0	(−9.5 to 11.6)	1.0
Distribution of Postoperative outcomes	<i>Hematoma or swelling in groin or scrotum</i>	35 (24.5)	35 (23.6)	0.8	(−9.0 to 10.7)	1.0
	<i>Superficial infection</i>	4 (2.8)	6 (4.1)	1.3	(−5.6 to 3.1)	0.38
	<i>Seroma</i>	1 (0.7)	0	0.7	(−1.2 to 2.6)	1.0
	<i>Impaired wound healing</i>	5 (3.5)	8 (5.4)	1.9	(−6.8 to 3.0)	0.29
	<i>Severe pain</i>	2 (1.4)	0	1.4	(−0.9 to 3.7)	1.0
	<i>Other complications</i>	2 (1.4)	4 (2.7)	1.3	(−4.8 to 2.2)	0.34
Death		2 (1.4)	3 (2.0)	0.6	(−3.9 to 2.6)	1.0

### Appraisal of an Article: Are the Results Valid?

Internal validity refers to the soundness of the methodology of the study and relates directly to our ability to infer causation from the results. In observational studies, the correlations presented in the results may not actually represent a true causal relationship and therefore using their conclusions without proper scrutiny may lead to unnecessary or potentially harmful patient care. Two types of errors can threaten internal validity [9]. Type I errors, or false positives, occur when calling a treatment useful when it is not. Type II errors, or false negatives, occur when concluding a treatment has no effect when it is actually useful (see Chap. 29). There are numerous examples of observational study conclusions that turned out to be equivocal or even harmful when examined in a randomized setting. Randomized trials provide the most sound experimental design and results that are most likely to be causal. This section will highlight some of the important points to look for when assessing the internal validity of a study. Importantly, some of this information may actually be in the study protocol rather than the paper itself.

### Was Patient Assignment Randomized, and was the Randomization Process “Concealed”?

Relationships between outcomes predictors of interest are subject to confounding. Confounding is the potential for a third variable to influence the relationship between the outcome and predictor thus limiting our ability to infer direct causation (see *Chapter 32: Confounding Factors and Interactions*). For observational studies, many of these confounders can be identified and adjusted for within multivariable statistical analyses. One of the advantages of prospective over retrospective studies is the ability to ensure data collection on important confounding factors. However, there are many unmeasurable or unknown confounding factors that cannot be collected and adjusted for. It is for these factors that randomization is crucial as it is the best way to balance these unknown or unmeasurable factors between groups. The method of randomization (usually a computer program) should be mentioned in the methods of any trial. Another vital aspect of randomization is allocation concealment. This refers to the concealment of the randomization process from the investigators so that it is unknown what group a patient will be

randomized into. If not done properly, investigators can intentionally or unintentionally direct patients towards the group they feel is most suitable thus introducing selection bias into the process and losing much of the benefit of randomization. In the study by Löfgren et al. [7] the operation list and order of patients was determined the day before but randomization was not performed until the patient was brought into the operative suite. In this way, randomization was concealed from the surgeons. In addition, randomization was done by a computer program in blocks of 4 and 6, rather than randomizing single individuals which allows for more balance of factors.

### **Were Study Personnel “Blinded” to Treatment and Apart from the Experimental Intervention, Were the Groups Treated Equally?**

Another major methodological concept is blinding. This is often confused with allocation concealment but they are distinct concepts. Blinding refers to ensuring that stakeholders in the trial do not know what treatment patients received as knowing this can influence the behavior of patients and investigators. While blinding of both patients and investigators is relatively straightforward in drug trials, it is often not possible in surgical trials as surgeons will often need to know the treatment the patient had. One way to minimize this issue in surgical trials is to have different clinicians provide the postoperative care to patients. In the study by Löfgren et al. [7], after randomization was done, the surgeons in the operative suite did know the type of mesh to be used within the procedure (blinded-patient). However, to minimize bias due to this, the two physicians performing the follow-up did not participate in the surgeries and were unaware of the study group assignments. Considering the study question, this was probably the strongest level of blinding possible.

### **Were All Patients Who Entered the Trial Accounted for and Was Follow-up Adequate?**

Another major issue with trials of all types is patient attrition rate. Readers should be concerned if not all patients are accounted for at a trial’s conclusion. If a large proportion of patients are unaccounted for at the end of a trial, the benefits of randomization may be lost. Moreover, bias can be introduced if the dropout is related to some aspect of the procedure itself. If the dropout was random, then the benefits of randomization should be maintained. Therefore, a full report of patient attrition is required. In addition, the follow-up should be rigorous, blinded and equal between groups to ensure that all adverse effects are accounted for. It should also be assessed similarly between groups and be long enough to ensure that the outcomes of interest can manifest themselves. In the Löfgren et al. [7] study, the follow-up was thoroughly conducted by two physicians who were blinded to the study group assignments. Overall, 4.4% of patients were lost to follow-up which was not different between groups. In addition, the time from surgery to follow-up was similar between groups. This is not unsurprising as there is little morbidity from a hernia repair and the follow-up period was relatively short.

### **Were Patients Analyzed According to the “Intention to Treat” Principle?**

The intention-to-treat principle is also fundamental to ensuring causal inference of results. This principle states that patients should be analyzed in the groups they were originally allocated to, regardless of the treatment actually received. This is vital in surgical trials as patients of poor operative status sometimes may not receive surgery, despite being randomized for surgery. If these patients were to be analyzed in the nonsurgical group, they can bias the results

by having healthier people in the surgical group. Sometimes, this principle can lead to issues of validity of conclusions if there are too many patients that did not receive the treatment but in most cases the strategy is sound. In addition, in surgical trials, if there are many patients not receiving the treatment it may provide a pragmatic answer as to the feasibility of the treatment.

In the Löfgren et al. [7] study, the intention to treat principle was followed for the final analysis. However, they make no mention of how the missing data was dealt with. It is likely, based on the small number of missing patients that the authors used a complete case analysis including data from only those patients who completed follow-up. This method of handling missing data is likely unbiased in this case due to the small numbers and did not likely substantially change the power for this study. Had patients dropped out due to measured or unmeasured confounders, the analysis will be biased. Furthermore, if too many patients were lost, even at random, the power of the study would be called into question.

### **Was the Study Large Enough to Detect a Difference?**

Ensuring an adequate sample size is essential to answering any clinical question. A randomized trial should clearly describe an a priori sample size and what factors they used to determine the sample size. Generally, calculating a sample size requires the rate of type I error (usually 5%), rate of type II error (usually 20%, also known as 80% power), the allocation ratio between groups (1:1, 2:1, etc.), the expected effect and the variance of that effect. While the first three are fairly standard in the sample size calculation, the last few variables can be controlled by investigators. A larger a priori expected effect means a smaller sample size but if that effect is unreasonable then it could lead to an underpowered trial. Conversely, the effect size chosen should also be large enough to be clinically relevant. Power is the complement of the type II error rate for a trial, and can be described as the chance for a trial to produce a

false negative result. Therefore, before a negative result can be truly established, the sample size and statistical power should be scrutinized and found to be adequate (see Chap. 29). In addition, the conclusions of other outcomes assessed, for which there was no sample size calculation, should not be assumed to be adequately powered. In the Löfgren et al. [7] study, the sample size was calculated at 150 patients within each group. This would give the study a power of 80% and a significance level of 5% to detect a five-percentage-point absolute difference in the rate of hernia recurrence at one year. Upon completion of the trial, this study failed to achieve the desired power level. The inability to reach power based on accrual issues should have been mentioned in the limitations so that it can be considered by the reader (see Chap. 29).

---

### **What Are the Results?**

The key findings of the Löfgren et al. [7] study are found in Table 6.2.

### **What Outcomes Were Used and How Were They Measured?**

Because of the wide variety of outcomes that can be assessed and the use of sophisticated statistical analyses, simply understanding what the results are can be a challenge for a surgeon. This section will provide a brief summary of how the most common outcomes are reported.

#### **Binary Versus Continuous Outcomes**

The vast majority of outcomes in the surgical literature are measured as either binary or continuous variables. Binary outcomes represent dichotomous occurrences where patients either have an outcome or they do not (e.g., death or anastomotic leak). Continuous outcomes represent data that are measured by real numbers such as weight or length of stay. The study by Löfgren et al. [7] had two primary outcomes, both binary, which were the hernia recurrence at 1 year and the overall complication rate at 2 weeks.

### Univariate Testing: Univariable and Multivariable Analyses

There is considerable confusion in the literature as to the nomenclature for testing let alone the actual test themselves. Univariate testing refers to statistical tests with a single response variable per observation. This represents the vast majority of tests in the surgical literature where a single patient will have a single outcome associated with them. Multivariate analyses refer to when a single patient has multiple outcomes associated with them and are rarely used (i.e., a patient has several weight changes over the course of the study) [10]. Univariate testing can take the form of univariable or multivariable tests. Univariable tests occur when the outcome is tested against only a single predictor. Examples include the chi-square test for dichotomous data and a t-test for continuous data. For observational data, these usually represent preliminary tests and are used to give some exposition to the data rather than as actual conclusions. Alternatively, for randomized trials, these may represent the final analysis because randomization precludes the need for multivariable analysis.

Multivariable analyses use multiple predictors to explain the outcome of interest. The simplest examples are linear regression for continuous data and logistic regression for binary data. These studies demonstrate the effect of a single predictor of interest while holding all other predictors steady. These analyses, therefore, account for the effect of many different potential confounders which is not done in univariable analyses. In the study by Löfgren et al. [7], because of the fact that it is a randomized clinical trial, multivariable regression was not required. This is because the groups are expected to be balanced due to randomization. Therefore, the main analysis was done using a chi-square test or the Fisher exact test, where appropriate.

### Multivariable Regression Results: Odds Ratios (OR) and Risk Ratios (RR)

Every surgeon has seen the results of a multivariable regression but correctly interpreting the results can be a bit more difficult. For linear regressions, the conclusions are represented by

the effect on the outcome of interest by a one unit change in the predictor. For example, if an outcome was weight and the predictor was age, the results would be represented as the amount the weight has changed for each year in age. For dichotomous data analyzed by logistic regression, results are represented by odds ratios. Odds ratios are related to risk ratios but are less intuitive. Their predominant use stems from the fact that they are much easier to calculate within statistical models. Risk ratios (also known as relative risks or the incidence rate ratios) represent the ratio between the actual event rate between groups. For example, if the event happened 60% of the time in group A and 40% in group B then the risk ratio would be 1.5. Similarly, odds ratios are the proportion of the odds of occurrence between the two groups. It is not vital to truly understand the difference between odds and risks but it is important to know that odds ratios approximate risk ratios when outcomes are rare but *overestimate* risk when the outcomes are common (>10%). Relative risks are common in randomized trials where multivariable logistic regression analyses are not required due to randomization. Although not utilized in the study by Löfgren et al. [7], one could easily calculate the risk ratio of postoperative complications (refer to Table 6.3 for calculation). In the commercial mesh group, the rate of complication was 29.7% whereas the rate of complication in the low-cost mesh group was 30.8%, therefore, the relative risk of complications in the low-cost mesh group was 1.04, or 4% higher, not a statistically significant difference. Furthermore, a relative risk of 4% is unlikely to be clinically important even if statistically different. One could use the 95% confidence interval and an a priori selected clinically important difference benchmark (non-inferiority margin) to determine whether the change is a large one to persuade us to accept the low-cost mesh. This is the basis of non-inferiority trials (see Chap. 13). Table 6.3 explains the measures that are used to explain the magnitude and precision of the treatment effect of a surgical intervention. The Löfgren et al. [7] article is explained through these measures in Column 3.

## How Large Was the Treatment Effect?

Importantly, the difference between absolute and relative measures needs to be understood by the surgeon contextualizing any results. Odds ratios and risk ratios represent a percentage change from the occurrence of an event. The absolute risk represents the change of the occurrence of an event on an absolute scale. From a clinical perspective, the latter is usually the much more important measure. For example, a study may report a predictor increases the event rate by 30% (risk ratio 1.30). This number seems large but if the event only occurs 1% of the time then the absolute risk only goes up to 1.3%. The same study could have reported a 0.3% increase in absolute risk—a much less provocative number. Conversely, a study could report a 5% relative increase (risk ratio 1.05) but if the event occurs

50% of the time then the absolute risk increase is 2.5%. Despite the clinical utility, absolute risks are usually only reported in randomized trials because they are much more difficult to model in regression analyses. Accordingly, surgeons must keep in mind the absolute risk of an event to contextualize the results of many trials. The study by Löfgren et al. [7] does report the absolute risk difference which is 1.1% and this was not seen as statistically significant. This, and calculations for the other results from the Löfgren et al. [7] study are shown in Table 6.2. The number needed to treat (NNT) of 91 (see column 3 Table 6.3) is an important measure which adds context to our interpretation of the data. It tells us that we have to treat 91 patients with the commercial mesh (instead of the low-cost mesh) to avoid one complication. Many surgeons may not see this as a huge benefit and may opt for the low-cost alternative.

**Table 6.3** Terms used to show the magnitude and precision of the treatment effect

Term	Description	Example from Löfgren et al. [7]
Risk	The probability that an event would occur calculated as the number of events of interest divided by the total number within the group. Also, known as the relative risk	Risk in treatment group: $44/143 = 0.308$ (30.8%) Risk in the control group: $44/148 = 0.297$ (29.7%)
Risk ratio	The ratio of the risk between 2 different groups	Risk in treatment 0.308 versus control 0.297 RR: $30.8/29.7 = 1.04$
Odds	A ratio of the probability that the event will happen to the probability that the event will not happen within the same group	Risk in the treatment group: $44/(143-44) = 0.44$ Risk in the control group: $44/(148-44) = 0.42$
Odds ratio	The ratio of odds between 2 different groups	Odds in treatment 0.44 versus in control 0.42 RR: $0.44/0.42 = 1.05$
Absolute risk reduction	Absolute reduction in events in one group compared with the other	Complication in treatment 30.8% versus in control 29.7%. ARR: $30.8 - 29.7 = 1.1\%$
Number needed to treat	The average number of patients who need to be treated to prevent one additional bad outcome. It is defined as the inverse of the absolute risk reduction	ARR from article: 1.1% NNT = $1/ARR = 1/0.011$ NNT = 91 patients
Relative risk reduction	Complement of relative risk, expressed as a percentage	Complication in treatment 30.8% versus in control 29.7%. RR (30.8-29.7)/30.8 = 3.6%
95% confidence interval	An interval of values that include the true value 95% of the time (calculated)	Various methods of calculation

## How Precise was the Estimate of the Treatment Effect?

The last thing a surgeon should examine are the measures of statistical significance and precision. The significance is represented by  $p$  values.  $p$  values represent the probability of obtaining the results if the null hypothesis (i.e. the assumption that the predictor has no effect) were true. Generally, when this probability is less than 5% ( $<0.05$ ) we consider the result statistically significant. It should be noted that this specific threshold is completely arbitrary and there is nothing magical between a  $p$ -value of 0.051 and 0.049. This threshold exists to limit the rate of Type I error (e.g. false positives) to 5%. Lastly, even if the  $p$ -value is below this threshold, statistical significance does not equate to clinical relevance. As a standard, most trials use a two-sided  $p$ -value of 0.05 as the threshold for statistical significance (see Chap. 27). In addition to  $p$  values, confidence intervals exist to better characterize the precision of the result. The 95% confidence interval is routinely used and can be interpreted as the interval in which the true effect lies 95% of the time if the same study with the same sample was repeated (see Chap. 28). Therefore, in studies with small sample sizes, the confidence interval can be quite large but with increasing sample size, the confidence interval becomes smaller. In the Löfgren et al. [7] study, the two-sided 95% confidence interval were calculated and for the main outcome of postoperative complication rate difference, this interval varied from  $-9.5$  to  $11.6\%$ . One way to interpret this interval is by saying that if this same study was repeated multiple times, the true rate difference would lie within this interval 95% of the time.

---

## Are the Results Applicable to My Patients?

External validity refers to the applicability of results to other groups of patients [11]. After the results are understood and the internal validity

assessed, this aspect of any trial should be clearly investigated by the reader and the questions from Table 6.1 should be asked.

## Were the Study Patients Similar to My Patients?

Comparing the patient population of a trial to the patient in front of you is essential to the application of evidence to surgical practice. Many trials have restricted criteria for enrollment and the results may not be directly applicable to the patient you are treating. In addition, they may only be enrolled from a specific patient group (e.g., veterans, men, uncomplicated surgical problems) which may not be the patient in front of you. Often, the patients we treat are older, have more comorbidities, and complex surgical problems than a trial would allow. These factors may mitigate the expected benefit of a treatment and thus the surgeon should know how the evidence relates to the patient in front of them before making any treatment decisions. The protocol for this study clearly outlines the inclusion and exclusion criteria. Specifically, it included patients 18 years of age or older with reducible, unilateral, primary inguinal hernias. It excluded females, recurrent hernias, femoral hernias, those on anticoagulation, those with current drug abuse and ASA 3 or above. In addition, these men were all from Uganda with a mean age of around 45 years old, mean BMI of 21 and ASA score of 1 for nearly 90% of the patients. These criteria give a clear picture of the type of patient these results can be applied to. Based on the mean BMI and ASA score of the patients within the study, this trial may not be as applicable in North America; however, it may be applicable to a low-income country. Considering the differences, it is important to carefully determine whether extrapolation to a different patient group is reasonable or whether this patient group has several things that may be too different to apply to your patients.



## Were the Measured Outcomes Clinically Relevant?

Another area that surgeons should question before accepting a “superior treatment” is the choice of main outcomes considered in the study. The choice of main outcomes should be relevant to both the surgeon and the patient, and should be clinically meaningful. Certain biochemical markers may be relevant to surgeons but of no relevance to patients while postoperative pain and the return of function may be less important to surgeons but of great substance to patients. The main outcomes of the study include complication rate at 2 weeks and recurrence rate at 1 year. While a 1-year follow-up is relatively long, many recurrences occur after this and therefore the long-term durability of this treatment cannot be determined based on the results of this trial.

## Are My Surgical Skills Similar to Those of the Study Surgeons?

Surgeons should also evaluate whether the treatment itself is feasible within their own practice. Trials on robotic surgery or expensive/difficult to obtain materials may have little relevance to a surgical practice, especially in low-resource settings. Moreover, certain surgical techniques may be beyond the technical proficiency of an individual surgeon and thus the treatment effect would largely be lost in the hands of that surgeon. If clearly published evidence does favor a certain treatment that is available but the surgeon does not perform it well then the surgeon is faced with three options: proceed with another operation that surgeon performs well while discussing all options for the patient, refer the patient to a colleague or seek additional training to master the new technique. Surgical proficiency creates a dilemma for surgical trials utilizing complex procedures and provides a major difference between surgical and medical trials. If the trial uses inexperienced surgeons it could bias the results away from the treatment, even if it does have a benefit in experienced hands. However, if the trial only utilized highly skilled surgeons the

result may not be applicable to the larger surgical community. The trial by Löfgren et al. [7] utilized a relatively simple surgical procedure which is outlined in the protocol. Specifically, these were day surgeries under local anesthesia which used a Lichtenstein tension-free method. This method could likely be replicated by most general surgeons.

---

## Resolution of the Clinical Scenario

A careful critique of this article demonstrates that the internal validity is quite high and therefore the results are likely valid to their goals. The resident who appraised this article likely felt that it was appropriate for her staff to use the low-cost mosquito net in this instance. It was comparable to the hernia mesh used in North America when used in these patients. However, it was clear that this conclusion may not be applicable to other groups at home in North America and the extrapolation of the results should be limited.

---

## References

1. Garas G, Ibrahim A, Ashrafian H, Ahmed K, Patel V, Okabayashi K, et al. Evidence-based surgery: barriers, solutions, and the role of evidence synthesis. *World J Surg.* 2012;36(8):1723–31.
2. Baker A, Young K, Potter J, Madan I. A review of grading systems for evidence-based guidelines produced by medical specialties. *Clin Med.* 2010;10(4):358–63.
3. Kavanagh BP. The GRADE system for rating clinical guidelines. *PLoS Med.* 2009;6(9):e1000094.
4. Higgins J, Green S, editors. 8.2: “Bias” and “risk of bias.” In: *Cochrane handbook for systematic reviews of interventions version 5.1.0* [Internet]. The Cochrane Collaboration; 2011 [cited 14 June 2018]. Available from: [www.handbook.cochrane.org](http://www.handbook.cochrane.org).
5. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ.* 1994; 309:1286–91.
6. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011;64(4):401–6.
7. Löfgren J, Nordin P, Ibingira C, Matovu A, Galiwango E, Wladis A. A randomized trial of low-cost mesh in groin hernia repair. *N Engl J Med.* 2016;374:146–53.

8. Urschel JD, Goldsmith CH, Tandan VR, Miller JD. For the evidence-based surgery working group. Users' guide to evidence-based surgery: how to use an article evaluating surgical interventions. *Can J Surg*. 2001; 44(2):95–100.
9. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-based medicine working group. *JAMA*. 1993;270:2598–601.
10. Hidalgo B, Goodman M. Multivariate or multivariable regression? *Am J Public Health*. 2013;103: 39–40.
11. Sedgwick P. External and internal validity in clinical trials. *BMJ*. 2012;344:e1004.