Users’ guide to the surgical literature: how to assess an article about harm in surgery

S urgeons often walk a proverbial tightrope, balancing the benefits and harms of surgical care. At all times, they must strive to ensure that the benefits of surgery outweigh any potential harm. For example, does the benefit of repairing minimally symptomatic inguinal hernia outweigh the risks in a patient with severe asthma and coronary heart disease? Indeed, patient safety is the cornerstone of good surgical practice.

The need for an evidence-based approach to harm reduction is imperative. In addition to surgical skills, we must possess the skills and confidence to identify and appraise the available evidence.

The patient in our clinical scenario wants to know if there is potential harm associated with breast reconstruction using silicone breast implants, particularly an increased risk of ALCL developing in the breast(s). She read in a magazine recently that ALCL, an unusual form of breast cancer, has been occurring in patients who have breast implants. She is very concerned that she might be at risk and asks for your opinion as to whether she should proceed with the procedure or not.

**Clinical Scenario**

You are a new plastic surgeon in the community and you are referred a patient interested in breast reconstruction. The patient is a 35-year-old female school teacher who had a bilateral prophylactic mastectomy 2 years earlier, as she was a BRCA1 gene carrier. Since she is of a petite build with very little subcutaneous tissue or extra skin in the lower abdomen, you decide that she is not a suitable candidate for an abdomen-based autologous tissue reconstruction. You recommend the technique of tissue expansion and silicone gel implants. She is concerned, however, about the possibility of anaplastic large cell lymphoma (ALCL) developing in her breasts. She read in a magazine recently that ALCL, an unusual form of breast cancer, has been occurring in patients who have breast implants. She is very concerned that she might be at risk and asks for your opinion as to whether she should proceed with the procedure or not.

The present guide will provide the readers with the tools to appraise surgical articles that deal with the issue of harm. The focus of the present article was to provide a global guide to appraise surgical articles that deal with the issue of harm.

The present guide will provide the readers with the tools to appraise surgical articles that assess harm. We used a similar framework to those used in previous evidence-based surgery (EBS) articles (Box 1).1-6

In line with our tenets of EBS, we must consult the best evidence for resolving issues about harm. Ideally, one would hope to acquire preappraised literature on the topic, providing expert review on primary literature and offering opinions for surgical management.2 However, with an emerging...
To identify the best evidence and inform the patient in our clinical scenario, we performed a literature search according to the Users’ Guide for Surgical Literature: How to perform a high-quality literature search. Deconstructing our research question using the PICOT (population, intervention, comparison, outcome, time horizon) format allowed us to choose important keywords for our search. 

- Population: female mastectomy patients
- Intervention: silicone gel breast implants
- Comparison: no intervention
- Outcome: ALCL
- Time horizon: any period of time after breast implant

Searching PubMed Clinical Queries using the search terms “anaplastic large cell lymphoma” AND “breast” AND (implant OR prosthes*), we identified 13 articles, including 1 case–control study, 1 retrospective historical cohort study and 3 systematic reviews that were relevant. With the above search strategy, the multiple case reports associated with this topic were eliminated. No RCTs were identified that dealt with ALCL in patients with breast implants. A systematic review represents an ideal source to answer our question, and we identified 3 studies of this type. Systematic reviews are, in general, of a higher level of evidence than a single study owing to the greater power of their pooled results. However, one must also take

Table 1: Description of the primary study designs, adapted from Levine et al

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Randomized controlled trial</th>
<th>Prospective cohort study</th>
<th>Case–control study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting point</td>
<td>Intervention status</td>
<td>Intervention/exposure status</td>
<td>Event/outcome status</td>
</tr>
<tr>
<td>Group allocation</td>
<td>Randomization; groups are balanced for known and unknown confounding factors</td>
<td>Groups are selected to intervention or exposure; groups may not be balanced.</td>
<td>Groups are selected to intervention or exposure; groups may not be balanced.</td>
</tr>
<tr>
<td>Outcome measures</td>
<td>Incidence of disease</td>
<td>Incidence of disease</td>
<td>Prevalence of disease</td>
</tr>
<tr>
<td>Measure of risk</td>
<td>Relative risk; odds ratio; risk difference</td>
<td>Relative risk; odds ratio; risk difference</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>Temporal relationship between exposure and disease</td>
<td>Easier to establish</td>
<td>Easier to establish</td>
<td>Harder to establish</td>
</tr>
<tr>
<td>Strength</td>
<td>Bias controlled</td>
<td>Bias uncontrolled</td>
<td>Bias uncontrolled</td>
</tr>
<tr>
<td>Validity (if well-designed)</td>
<td>Level I evidence</td>
<td>Level II evidence</td>
<td>Level III evidence</td>
</tr>
</tbody>
</table>

Box 1: Framework for critical appraisal of an article that deals with harm

I – Are the results valid?
- cohort studies: apart from the exposure of interest did the exposed and control groups start and finish with the same risk for the outcomes?
- Were patients similar in terms of prognostic factors that are known to be associated with the outcome (or was statistical adjustment necessary)?
- Were the circumstances and methods for detecting the outcome similar for patients and controls?
- Was follow-up sufficient to exclude bias?
- Case–control studies: did the cases and control group have the same risk (or was statistical adjustment necessary)?
- Were the circumstances and methods for determining exposure similar for cases and controls?
- Were the circumstances and methods for determining exposure similar for cases and controls?
- Was the correct temporal relationship demonstrated?
- Was there a dose–response relationship?

II – What are the results?
- How strong is the association between exposure and outcome?
- How precise was the estimate of the risk?

III – How can I apply the result my patient or clinical practice?
- Were the patients in the appraised study similar to the patient in my practice?
- Was follow-up sufficiently long?
- Is the exposure similar to what might occur in my patient?
- What is the magnitude of the risk?
- Are there any benefits that are known to be associated with exposure?
into account that a systematic review is only as strong as the articles included in its results; thus, it is important to ensure that the review has reported sufficiently appraised studies and appraised the methodology used for each of them. None of the systematic reviews retrieved using our search strategy offered a comprehensive critical appraisal of the included studies. Moreover, systematic reviews may include case reports and case series that are inherent to bias.

The cohort study identified in our literature search was excluded because no ALCL was identified. Thus, we selected the article by de Jong and colleagues because it represented the best available evidence addressing our clinical scenario. It was a case–control study that included all cases of patients with lymphoma in the breast from the entire population of the Netherlands in a 16-year span. In the absence of any RCTs, we believe this study design to be the most appropriate for measuring causation in the case of a rare harmful outcome, such as ALCL. The key methodological characteristics of the study by de Jong and colleagues are summarized in Box 2.

### Critical appraisal of the article

The 2 most common designs dealing with harm are the cohort and case–control designs. Box 1 includes questions that need be answered for the appraisal of either a cohort or case–control study. As the article by de Jong and colleagues uses a case–control design, our appraisal questions will pertain to this study.

**Are the results valid?**

Did the patients and controls have the same risk (chance) for being exposed in the past? Valid results are essential to making a clinical or surgical decision. Without adequate confidence that the results represent what they are intending to represent, there is insufficient evidence from which to draw conclusions.

Were patients and controls similar with respect to the indication or circumstances that would lead to exposure?

To assess possible causation in a case–control study, patients and controls with similar baseline characteristics are essential to minimize selection bias. This criterion outlines 1 area where randomized groups may be optimal; however, it is important for the surgeon to anticipate an absence of randomization and pay careful attention to how the groups were balanced. The study by de Jong and colleagues identified 389 women from a Dutch national database with histological evidence for lymphoma for the period 1990–2006; 11 women in total received diagnoses of ALCL. A standardized questionnaire was sent to the treating physicians for acquisition of medical information of each patient and control, including previous malignancies, staging results, presence of a breast prosthesis and mammographic results. Balancing of comparison groups was achieved to an extent by matching each patient with 3–7 controls for age (within 5 years) and year of diagnosis (within 2 years), all of whom were nested in the same cohort of female patients with primary breast lymphoma. Baseline prognostic factors were presented in Tables 1 and 2 of their article and included age at diagnosis, year of diagnosis, stage of the lymphoma, breast involvement and lymph involvement, year of placement of the prosthesis, and removal and the type of prosthesis (not provided in all cases). Based on the presented information, the comparison groups had similar baseline prognostic factors except for breast implant(s).

When assessing if the patient and control groups are comparable at baseline, it is important to ensure that all documented risk factors are addressed. Of course we cannot be absolutely certain of all the risk factors. Is the size of the breast, for example, a risk factor for ALCL? We presume that women who had breast implants had smaller breasts than controls, but we cannot be absolutely certain. It would be important in the appraisal of the study to be confident that those risks on their own could not account for the high ACLC rate. Based on the methods used, we were satisfied that patients and controls in the study by de Jong and colleagues were comparable at baseline.

**Were the circumstances and methods for determining exposure similar for patients and controls?**

There are certain biases that should be considered in a case–control study. The methods of diagnosing the outcome and assessing the exposure are particularly important to avoid case-ascertainment and misclassification bias. Ascertainment bias refers to the error associated with selecting patients and controls based on their exposure

---

<table>
<thead>
<tr>
<th>Box 2: Key methodological features of the matched case–control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source of cases</strong></td>
<td></td>
</tr>
<tr>
<td>• Netherlands population-based database: Pathologisch Anatomisch Landelijk Geautomatiseerd Archief (PALGA)</td>
<td></td>
</tr>
<tr>
<td>• 429 (389 women and 40 men) histologically proven cases of lymphoma of the breast</td>
<td></td>
</tr>
<tr>
<td><strong>Source of controls</strong></td>
<td></td>
</tr>
<tr>
<td>• Controls with non-ALCL breast lymphoma from PALGA database</td>
<td></td>
</tr>
<tr>
<td><strong>No. of cases</strong></td>
<td></td>
</tr>
<tr>
<td>• n = 11</td>
<td></td>
</tr>
<tr>
<td>• All female patients with ALCL</td>
<td></td>
</tr>
<tr>
<td>• 2 of 11 patients recently diagnosed by the authors</td>
<td></td>
</tr>
<tr>
<td><strong>No. of controls</strong></td>
<td></td>
</tr>
<tr>
<td>• n = 35, matched for age and year of diagnosis with a ratio of 3–7 controls to 1 case</td>
<td></td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>• Individually matched cases</td>
<td></td>
</tr>
<tr>
<td>• Conditional logistic regression estimated the odds ratio of ALCL associated with breast prosthesis</td>
<td></td>
</tr>
</tbody>
</table>

ALCL = anaplastic large t-cell lymphoma.
status, such that they do not have an equal chance of inclusion in the study. Misclassification bias refers to the error associated with the misidentification of an exposure status or disease such that the participant is assigned to the incorrect group. To confirm diagnosis of ALCL (outcome) in the study by de Jong and colleagues, all the histological material and medical records were retrieved, and additional immunohistochemical analysis and molecular studies were performed. To determine presence of breast implant, standardized questionnaires were disseminated to the treating physician; these questionnaires included patient history, such as previous breast malignancies, staging results and presence of prosthesis. Information was similarly collected for both patients and controls.

Surveillance bias may also occur in case–control studies; patients may receive extra attention for ascertaining exposure. In the study by de Jong and colleagues, we are not convinced that any surveillance bias was present, as the exposed and unexposed groups alike were extracted from a database of patients with diagnoses of lymphoma, and the same questionnaire was used to ascertain exposure. We are satisfied that the outcomes and exposures were measured comparably in both patients and controls. The authors, however, did not provide information regarding the placement or removal and type of implants used in 5 of 11 patients with ALCL.

Was the correct temporal relationship demonstrated?
To determine true causation, it is necessary to confirm that the surgical management preceded the harmful outcome (introduction of breast prostheses preceded the development of the ALCL). Case–control studies begin by identifying the outcome first and working back toward the exposure, which is why the issue of temporal relationship is so critical to this study design. The study attempts to elucidate the temporal relationship by providing data on time of prosthesis insertion and time of diagnosis of ALCL, which is presented in Table 1 of the article by de Jong and colleagues. However, their information appears to be limited to only 5 of the 11 cases reported. In each of these patients, the implantation of the prosthesis preceded the diagnosis of ALCL. Thus, we cannot be certain of the temporal relationship between exposure and outcome with all patients, leading to uncertainty in the appraisal of the study’s validity.

Was there a dose–response relationship?
Identification of a dose–response phenomenon between exposure and a harmful outcome is yet another measure to justify true causation. A classic example of a dose–response gradient germane to the harm topic is demonstrated in a study by Doll and Hill, wherein cigarette smoking (measured in pack years) showed a dose–response relationship to lung cancer. In our clinical example, we would be more confident in attributing ALCL development to breast implant exposure if we could demonstrate that greater exposure (i.e., greater silicone volume and longer-duration implants) increases the likelihood of ALCL. However, dose–response data may not be a realistic expectation when dealing with surgical studies in general, since the exposure often cannot be titrated to specific doses, as in medication studies.

What are the results?
How strong is the association between exposure and the outcome?
Statistical analysis for measuring effect in a case–control study is typically done using an odds ratio (OR). An OR measures how strong of an association there is between an exposure (breast implant) and disease (ALCL). It is different from other measures of effect, such as relative risk (RR) used in RCT and cohort study designs, and can be more difficult to interpret. The RR cannot be used with a case–control study design since incidence of the disease is unknown; however, the RR and OR approach similar values in the case of rare disease. In a case–control study, the data are classically represented in a 2 × 2 table (Table 2). Patients and controls are classified as exposed and unexposed. Table 2 presents a simple 2 × 2 table for the calculation of an OR, defined as the odds of an event in the exposed group (A/B) divided by the odds of an event in the unexposed group (C/D). An OR greater than 1 indicates that the risk of disease is higher when exposed to the risk factor in question, whereas an OR equal to 1 indicates no risk/association. In the study by de Jong and colleagues, this representation does not hold true since patients and controls were matched; instead a matched analysis method was used to calculate OR. Matching is essential to statistically analyze the results of the study owing to loss of independence between the 2 groups. In this case, conditional logistic regression analysis using the software program EGRET is used to more appropriately estimate the OR of ALCL associated with breast prosthesis while adjusting for between-group differences with respect to other risk factors. The study reported an OR of 18.2 (95% confidence interval [CI] 2.1–156.8) for ALCL in patients who received a breast prosthesis placed for cosmetic reasons, which means that the odds of ALCL developing in those exposed to breast implants is 18.2 times greater than in those with ALCL who have not had breast implants. This can be interpreted as 18 times greater odds for the

<table>
<thead>
<tr>
<th>Harmful outcome</th>
<th>ALCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>No</td>
<td>6</td>
</tr>
</tbody>
</table>

OR = (a + b)/(c + d) OR = (5 + 1)/(6 + 34) = 28.3

ALCL = anaplastic large T-cell lymphoma; OR = odds ratio.
development of ALCL in those with implants than in those without implants; note that the simple calculation included in Table 2 yields an OR of 28.3. de Jong and colleagues\(^6\) did not perform any subsequent analysis on missing data, instead excluding 1 patient in whom ALCL could not be confirmed.

The sample size in the study by de Jong and colleagues\(^6\) is too small to draw a valid conclusion. The article appropriately suggests that the findings are preliminary and recommends further confirmation of the association between ALCL and breast prosthesis.

**How precise was the estimate of the risk?**

Although the authors report an OR of 18.2 (95% CI 2.1–156.8), one must not make any swift conclusions when interpreting this value and take into account the characteristics of the specific harmful outcome. In our scenario, ALCL is a rare form of lymphoma, and thus its absolute risk remains very low; absolute risk refers to the probability of a cause-specific event occurring in a specific interval of time in the population, regardless of risk factors.\(^7\) The estimated incidence at all sites reported by de Jong and colleagues\(^6\) was 0.1/100 000 per year, which implies an exceedingly low overall risk of the disease developing. The authors estimated that the magnitude of risk for ALCL developing in the breast would be between 0.1 and 0.3/100 000 in women with breast prosthesis per year, based on 11 cases being identified in the Netherlands, with a population of 8 million women, during this period. Therefore the absolute risk of breast cancer developing in a breast containing a prosthesis is much higher than the risk of ALCL as reported in their study.

**How can I apply the result in my patient or clinical practice?**

As discussed earlier, the risk of ALCL with breast prosthesis is small. However, the CI was inclusive to a large odds ratio. It is in fact more likely that the patient in our clinical scenario will have breast cancer during her lifetime regardless of breast implant use because she is a carrier of the BRCA gene (11.7% in Canada) than our calculated risk of ALCL attributable to breast implant.\(^8\) Thus, ALCL need not be a primary concern despite the speculations in the patient’s magazine.

**Were the patients in the appraised study similar to the patient in my practice?**

The age of patients in the study by de Jong and colleagues\(^6\) ranged from 24–68 years (though they state a median age of 40 years), and the population (the Netherlands) is comparable to that of North America; our patient is median age of 40), and the population (the Netherlands) is comparable to that of North America. All patients are presumed to have been followed within a tertiary care setting. Thus, it is appropriate to apply the study results to the patient in our clinical scenario.

**Was follow-up sufficiently long?**

Adequate follow-up and measurement of the outcomes are important issues to consider in prospective RCTs and cohort studies. We are assessing a case–control study, the event has taken place previously and follow-up for diagnosis of ALCL was not a factor in the study performed. The investigators reported the study time period, which was 1990–2006. This period, extending to 16 years, may be insufficient for the identification of ALCL in controls, as the time from surgery to development of ALCL has been reported to be up to 32 years; however, the mean is 11 years.\(^12\)

The importance of follow-up should not be understated when appraising a harm article that involves a prospective study design. Whereas some harmful effects may occur early on in a patient’s follow-up, many harmful outcomes can manifest years after surgery. A prospective study with an insufficient follow-up period can mask the association of harm with a surgical procedure. Such was the case with Poly Implant Prothèse breast implants that were later found to be composed of improper quality materials leading to complications, including high rates of rupture.\(^9\) We are satisfied with the length of follow-up in the study by de Jong and colleagues\(^6\).

**Is the exposure similar to what might occur in my patient?**

The age of patients in the study by de Jong and colleagues\(^6\) ranged from 24–68 years (though they state a median age of 40), and the population (the Netherlands) is comparable to that of North America.; our patient is 35 years old, falling within their reported age range. The risk of ALCL may be slightly different for our patient in 2015. Our patient is interested in breast reconstruction, and she would most likely have a newer-generation implant with a cohesive silicone gel. The implants made today are different from those used in Dutch women in 1990–2006. Presently there is no evidence that the new implants have a risk profile identical to those of patients in the study by de Jong and colleagues.\(^13\)

**What is the magnitude of the risk?**

The OR does not tell us how frequently ALCL occurs. It tells us only that this harmful outcome occurs more often in the exposed group than in the unexposed group. To determine the clinical importance of the results, it is advisable to calculate the number of patients who would need to be exposed to breast implants to result in 1 additional harmful event; this value is known as the number needed to harm (NNH).

The NNH is conceptually and mathematically simple in studies where there are distinct exposed and unexposed groups (RCTs or cohort studies); however, the NNH becomes more complex when we attempt to calculate it based on OR values, as is typically the case with a case–control

studied. With case–control results, we also need to know the expected event rate for ALCL in the unexposed population, known as the patient-expected event rate (PEER) or control event rate (CER). The NNH is calculated as follows: \( \text{NNH} = \frac{[\text{PEER} \times (\text{OR} - 1)] + 1}{\text{PEER} \times (\text{OR} - 1) \times (1 - \text{PEER})} \). For PEER we can use an appropriate value of incidence for ALCL from the literature. In our case, we can apply the estimated incidence of ALCL found in the discussion of the study by de Jong and colleagues, which is 0.1/100 000 (0.000001) per year. Note that this is an estimated incidence, and thus our final NNH will also be an estimate. Our calculation was as follows: \( \text{NNH} = \frac{[0.000001 \times (18.2 - 1)] + 1}{0.000001 \times (18.2 - 1) \times (1 - 0.000001)} \). The calculation produces an NNH of just over 58 140, meaning that more than 58 000 patients would need to be treated with breast implants per year to result in 1 case of ALCL. The NNH provides both you and the patient with an easy-to-understand representation of the risk of harm.

Are there any benefits known to be associated with exposure?

After assessing the evidence that an exposure is causing harm and the results are applicable to our patient, we are faced with the difficult task of determining what the adverse effects are of not exposing our patient to the potentially harmful breast implants. There is ample evidence that breast implants in postmastectomy reconstruction have a beneficial effect for the patient that is both clinically important and statistically significant. The beneficial effects that must be considered in lieu of ALCL include improved well-being and long-term health in breast cancer survivors. In our clinical scenario where the magnitude of the risk is so small in contrast to the well-studied benefits of implant-based reconstruction, the decision moving forward may be much easier to make. To add context for future appraisals we recommend that one should know the risks. Basically one should ask if the benefits are worth the risk of harm.

**Conclusion**

For a surgeon counselling a patient on surgical care, informed consent is integral. A well-informed decision should involve communicating all known risks based on evidence that is not only up to date, but also stands up to the rigours of a well-conducted critical appraisal by the surgeon or other expert in the field. If upon reviewing the literature, there is sufficient concern regarding harm associated with surgical management, the surgeon should discuss the findings with the patient. The discussion should take into account the patient’s desires for a given technique/procedure, availability of resources, the surgeon’s own comfort with the procedures and the potential for undesirable results.

**Resolution of the clinical scenario**

By applying the 3 steps in Box 1, we concluded that the study by de Jong and colleagues holds up to methodological standards of the case–control study, demonstrating no important bias between groups that could render the results unreliable. We observed a significantly positive association (OR 18.2, 95% CI 2.1–156.8) between breast implants and ALCL. However, taking into account that the risk of ALCL developing is very low overall, we concluded that the risk of ALCL in our patient is very small (in this case, the NNH is 58 140). Having learned this information, our patient decided to proceed with breast reconstruction using the technique of tissue expansion and silicone gel implants.

**Affiliations:** From the Division of Plastic Surgery, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Thoma, Levis); the Surgical Outcomes Research Center, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Thoma, Kaur, Levic, Lovrics, Goldsmith, Farrokhyar); the Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Thoma, Farrokhyar, Goldsmith); the School of Rehabilitation Sciences, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Kaur); the Faculty of Medicine, University of Ottawa, Ottawa, Ont. (Waltho); the Division of General Surgery, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Lovrics); the Faculty of Health Sciences, Simon Fraser University, Vancouver, BC (Goldsmith).

**Funding:** M. Kaur is funded by the Canadian Institute of Health Research Doctoral Award (2015-18).

**Competing interests:** None declared.

**Contributors:** All authors contributed substantially to the conception, writing and revision of this article and approved the final version for publication.

**References**


