Statistics for the Rest of Us

Ten traps for the unwary in surgical series: A case study in mesothelioma reports

Tom Treasure, MD,[a] and Martin Utley, PhD[b]

Case series are the commonest form of evidence in cancer surgery, an area of practice for which there are few randomized trials. There are pitfalls in the presentation and interpretation of this clinical material, and in this article we highlight a number of them. We have used as a case study the body of literature concerning resection of mesothelioma. Mesothelioma is a disease that concerns us greatly in Europe because we are facing an epidemic that will peak between 2010 and 2015,[1] and an even greater burden of disease will follow in parts of the developing world. It is a grim cancer, and the quality of evidence is poor.[2] However, the need to understand the limitations of case series spreads much wider than mesothelioma.

Authors may recognize their work in the examples we present, and so we may well cause offence. This is not our intention. Readers may be tempted to search our own work for examples of errors, and they will find them—more egregious than any of these. The alternative to citing real instances was for us to make them up or make them anonymous. However, then maybe some would say, "Surely no one would do that?" so they had to be real. We did not search systematically for examples; they come from articles we know through our research in mesothelioma. Finally, we have not shied away from using prominent examples, from leading authors, in leading journals. We needed to cite the best work or our concerns could be discounted as the result of a pedant's trawl of the world's literature. They are quoted with respect for the authors and editors, who we know have broad shoulders.

From the Department of Cardiothoracic Surgery, Guy's Hospital,[a] and the Clinical Operational Research Unit, University College London,[b] London, United Kingdom.

Received for publication Oct 31, 2006; revisions received Jan 31, 2007; accepted for publication Feb 12, 2007.

Address for reprints: Tom Treasure, Professor of Cardiothoracic Surgery, Guy's Hospital, London, United Kingdom (E-mail: tom.treasure@gmail.com).

J Thorac Cardiovasc Surg 2007;133:1414-8

0022-5223/$32.00

doi:10.1016/j.jtcvs.2007.02.014

Presumption of Efficacy

Radical surgery for mesothelioma was started in the 1970s.[3] There have been no randomized trials to show that radical surgery results in more survivors or longer survival than does no surgery.

The following statement (personal communication) from Eric Butchart, a pioneer of the pleuropneumonectomy procedure (also called extrapleural pneumonectomy or EPP), summarizes the situation: "We have recently analysed our experience with both pleuropneumonectomy and pleurectomy/decortication for mesothelioma. The very strong message from this analysis is that adjuvant therapy is essential in order to achieve any degree of long term survival with either surgical procedure."

The inference that no degree of long-term survival is achieved with surgery alone is embedded in the statement by Butchart. If there is a perceived improvement in survival, logically it should be attributed to the adjuvant chemotherapy, and yet authors of surgical case series tend to conclude that surgery confers a survival advantage.

This belief would seem to be widespread. In the Methods section of the first report of a randomized trial of pemetrexed (in which a difference in survival was found between two regimens of chemotherapy), there is the statement that "candi-

dates for curative surgery" were excluded.[4] The clear implication is that curative surgery exists for these patients and that somewhere surgical cure is being achieved, but no reference is cited.

When surgeons perform very radical surgery on many patients, there is deep felt belief, held by medics and the public alike, that the operation must be doing some good. This is a very human tenet, which we call the *presumption of efficacy*.

Lead-time Shift

There is a variable interval between the onset of malignant disease (which is hard to define) and the confirmation by histologic diagnosis (typically the point in time from which survival is measured in nonoperated cases). When patients have a disease surrounded by therapeutic nihilism (as was mesothelioma) and a clinical team declares interest, referrals are made sooner and the diagnosis is likely to be made earlier in the course of the disease. Historical series are drawn from the era before the active treatment was instigated, and these patients are likely to have received the diagnosis later in the course of the disease. For example, respiratory physicians in our own catchment area refer earlier now than they did in previous years because they know we are interested. It is known that we are interested in managing malignant effusion and in making the diagnosis of pleural thickening as early as possible because we wrote about it in the *British Medical Journal*.[5] It is highly likely that the interval between tissue diagnosis and death is longer now than it was previously for this reason alone. A case series collected in an earlier epoch would not be a reliable comparator against which to judge a present-day series.

The introduction to the most quoted surgical series affirms that the "natural history of malignant pleural mesothelioma includes a median survival of 4–12 months without intervention."[6] There are three citations in support of this statement. They were published in 1980, 1984, and 1989, 10 to 20 years before the citation in 1999. Two of these observational series specifically comment on the delay and lateness in making a diagnosis.

Statements such as "recent series show that overall survival is much longer than previously thought when [mesothelioma] is diagnosed at an early stage"[7] illustrate the potential for confusion between lengthening survival by therapy and by starting the clock sooner.

There are factors other than the vintage of the data. In 2001 we identified all patients in two London thoracic surgical centers in whom the diagnosis of mesothelioma was made. The two sets of 127 and 234 patients were similar in age, sex, and the mix of tumor types.[8] Neither group was treated aggressively; surgery at most was talc pleurodesis, and none had EPP. The survival of the two groups was very different. There are various possible explanations including chance, lead-time shift, and differing histologic criteria, but these uncertainties only add to the reasons for exercising caution when using historical series in comparisons. Either or both of these series of cases diagnosed in the 5 years up to 2001 would be a more fair comparison than the survival of patients diagnosed in the 1970s and 1980s, but they might still be misleading. Comparisons against historical controls are inherently of limited reliability.

Comparison Groups May Have a Different Case Mix

There is another source of error in the use of historical series in the EPP literature: We cannot be sure that the patients within such series match those whom we now select for radical surgery. In fact, we can be virtually certain that they do not. It is very misleading if the case mix of the series put up for comparison is not the same as the treated group. We would suggest that if any series is put up for comparison, due diligence must be exercised in ascertaining and reporting the extent to which it compares with the series being newly reported.

To make the point, let us look at one definable factor on which contemporary surgical series and historical survival series are different. The retrospective case series included all patients with a diagnosis of mesothelioma. More than 40% of patients would have had sarcomatoid histology, for which survival time is much reduced. We cannot know exactly how many, so we cannot correct for it. Other than retrieve all the pathologic material and have it re-examined with contemporary criteria, ideally "blind," we can never know. These sarcomatoid tumors are increasingly excluded from EPP in current practice.

To compare current surgical figures with an all-case historical series would be like reporting your present day results for surgically treated non–small cell cancer and comparing 5-year survival with the natural history of all lung cancer, including small cell, from the 1970s before computed tomography and before positron emission tomography.

If the results of current practice are to be held up for inspection against any comparator, historical or otherwise, surely just as much care should be exercised in the analysis of the comparison group as in that of the surgical series. These data may not be a finding of the research reported, but they are an intrinsic part of the evidence presented, and yet even the summary statistics may be quoted blandly without adequate definition or caveat, as in the example cited: "The natural history of malignant pleural mesothelioma includes a median survival of 4–12 months."[6] And yet it is done in the case of mesothelioma without qualification or citation in a recent "evidence based approach."[9]

There is a consequence worse than inadvertently misleading ourselves and our colleagues. Once such statements appear in reputable journals, it is reasonable for others to

STATS

cite them; they are regarded as fact. In 2006, by which time the historical series quoted were 20 to 30 years out of date, they appeared on a public web page in similar terms and in the present tense: "survival averages four to 12 months."[10]

In Multimodality Cancer Treatment We Cannot Laud a Single Component

In the seven reported series of multimodality therapy we[2] reviewed, radical surgery was combined with chemotherapy, radiotherapy, or both. Given this, it is not possible on the basis of clinical observation to determine whether the surgical component is conferring any advantage. For patients with mesothelioma, it would be as rational to attribute any perceived benefit to chemotherapy and radiotherapy, and to spare them the operation.

The series are reported as surgical series—the radical surgery is center stage. However, any survival gain cannot be attributed to radical surgery. If there is an advantage, it should be attributed to the full treatment package.

"Intention-to-Treat" Analysis Is Required

It is common for authors of case series to report an analysis of those patients who *completed* the full treatment schedule under review. They are reporting on the outcomes for a particular management strategy. This is the contrary of "intention-to-treat" analysis, which is the generally accepted standard of reporting for prospective studies and randomized clinical trials. The lapse into "per-protocol" analysis is perhaps natural inasmuch as these are the data retrieved by the surgical team—the patients who were treated in a particular way.

The most obvious breach of intention-to-treat analysis is to exclude from the survival curves the patients who died in the immediate postoperative period.[6]

It goes further than that. An extreme scenario in mesothelioma is as follows. Patients are selected for radical surgery, and in some (but not all) EPP is performed. Of those who survive surgery, some (but not all) are selected to have radiotherapy. Some of these (but not all) continue to survive, and some (but not all) of these are well enough to be given chemotherapy and complete the course. Patients who complete treatment have a median survival much longer than would have been expected in comparison with the natural history of the disease. This survival is claimed as evidence for the effectiveness of multimodality treatment.

The patient pool under consideration is reduced at each of multiple stages through both clinical selection and survival. The patients had to be worked up, undergo, survive, and recover from each of the first two components and then be worked up for and undergo the final component. All take time (maybe as much as 3 months each), each step may eliminate some patients, and each intervention takes a toll on patients. Those who arrive at the end are clinically and self-selectedly survivors.

There may be a place for per-protocol analysis in evaluation of the mechanics of some surgical techniques, but it can surely have no place in the evaluation of multimodality therapy as arduous and drawn out as that offered for mesothelioma. Intention-to-treat analysis is essential.

Face Validity

Usually we like to understand how an operation works. After hip replacement, the lame walk; after cataract extraction, the blind see; and relief of aortic stenosis results in symptom relief and much longer survival. These all make complete sense to surgeons, who are practical people; we prefer there to be face validity to what we do.

A fundamental principle of curing cancer by surgery has been eradication of all disease at operation. There may be an emerging role of reducing tumor bulk as an adjunct to other treatment, but for curative surgery the first principle is to clear all the cancer. It is for that reason that certain forms of cancers for which radical surgery used to be performed with intent to cure are no longer operated on. These include small-cell lung cancer, lymphoma, and many cases of breast cancer.

These cancers are characterized as already being systemic diseases at presentation, and the cancer is thus beyond the surgeon's knife. That distinction, between the surgically curable and incurable cancers, has been negotiated since the early reports of radical surgery for mesothelioma. On the face of it, the creeping nature of mesothelioma and the arbitrariness of the anatomic clearance of the endothoracic fascia make it an unlikely candidate for surgical cure. The onus to counter that argument is on those who believe that EPP can cure.

Analysis of an operated series revealed three factors associated with longer postoperative survival: histologic type, mediastinal lymph node involvement, and surgical clearance, in that order of effect size.[6] The smallest survival difference was associated with whether the resection margins were clear (R0) or not (R1).

That the difference between survival time for R1 and R0 (ie, the difference between the operation failing to eradicate all disease and the operation possibly eradicating all disease) is smaller than the differences attributed to nodal status and histology undermines the notion that this cancer lends itself to surgical cure.

Factors for "Case Selection" Should Be Tested on a New Data Set

Respected surgeons believe that there is evidence on which to select patients for EPP and cite the 1999 report by David Sugarbaker and associates.[6] It behooves us, therefore, to look carefully at the evidence on which they rely. The report concerns 176 patients who were the survivors from 183 cases selected for surgery from an unspecified but presumably large denominator, given the reputation of this group.

Three factors were found in multivariable analysis to be associated with survival. These were used to classify patients into groups with 0, 1, 2, or 3 of these characteristics: epithelioid histology, no extrapleural nodes, and clear resection margins. There were significant differences among the four groups thus created. However, inasmuch as the factors were derived from a multivariable analysis and then used to categorize the same data, this is completely unsurprising. As a numerical argument it is circular. To prove the validity of these factors as predictors, they should be applied to a new series, not to the one from which they were derived. To be fair, this point is made by the authors[6] but does not seem to have been heeded.

Importantly, even if validated as factors associated with survival after EPP, such analysis cannot be used to select those likely to "benefit most" from EPP because the notion of differential benefit is based entirely on the *presumption of efficacy*.

Criteria for Case Selection Must Be Applicable Before Surgery

There is a separate flaw in the promulgation of these factors as a means of selecting patients as those believed to benefit most from radical surgery: they were all defined or refined postoperatively. Resection margins can only be defined post hoc, lymph node status is revised as part of pTNM staging, and with all of the resected tumor in hand, more patients will be reclassified as having sarcomatoid or mixed histology.

If we are to have a set of criteria on which to select or counsel patients, they must be available before surgery and be amenable to validation, indeed have been validated, before they can be used.

Association Is Not Causation

Memorial Sloan-Kettering has analyzed 945 patients with mesothelioma over the past 15 years. Surgical resection was associated with improved survival.[11] Does this prove that surgery favorably influenced the outcome for these patients? The choice of words, "Multimodality therapy including surgery *yielded* a median survival of . . ." [our emphasis] suggests that the writer believed so.

The problem is that many factors are used by experienced and knowledgeable teams when they decide to operate on one patient but not another. These factors include all available clinical information but cannot all be accounted for: some are not consciously recognized, some are not made explicit, and some are unknown unknowns.

In Flores' Cox model,[11] perhaps it was not the operation that was the determinant but that a sage surgeon, supported by many investigations and many colleagues, deemed the patient a candidate for surgery. It would be indeed surprising if these choices did not produce a cohort of patients who did better than those not selected, that is, unless the authors are suggesting that the decision to operate is some sort of haphazard event. Hence *association* is the correct word.

Hope Springs Eternal in the Human Breast

Finally, it is argued that surgeons must offer hope—but surely not false hope by obscuring the truth from those who are entitled to know it. Active treatment for which there is no evidence of clinical benefit is sometimes preferred to inaction for "psychological reasons." EPP is too severe to be justified on these grounds, whether it is to comfort the patient or the surgeon!

Conclusions

Too often, retrospective reports of surgical series lack objectivity and start from a presumption of efficacy. At worst, they are a form of self-justification, not evidence. There is an inherent underlying assumption that the operation must be doing something to help rather than harm, or how could we face ourselves, let alone our patients? Patients, too, find it hard to accept that the operation they have been through may be unavailing or that they face a condition for which there is no recognized means of cure.

One solution to all of the problems raised might be a randomized trial. A randomized clinical trial provides a contemporary matched comparison group with a similar case mix, avoids the pitfalls of retrospective multivariable analysis, and is analyzed on an intention-to-treat basis. There are many forms of evidence for what we as surgeons do,[12,13] but once removed from rather obvious cause-and-effect, when benefits are small and treatments are complex, it is unlikely that a case series will give secure evidence.

We have to conclude that there is as yet no reliable evidence for the effectiveness of radical surgery for mesothelioma, and patients should be informed of that fact. Our belief is that in the context of mesothelioma, radical surgery should be offered only in a trial in which its effectiveness can be tested. Some worry that randomized trials have the effect of disadvantaging half the patients. Where they are wrong is in thinking that they know which half.

A total of 126 trials of cancer treatments have been reported to date in 36,567 children.[14] Who did better, those on the new treatments or those in the control groups? The answer is that the trial results are evenly spread either side of zero. For each group of patients denied benefit, another was spared a less effective novel treatment.

We owe it to today's patients with mesothelioma to get it right. The epidemic is still rising in Europe and there will be many patients in the next 10 to 20 years. We owe it to the rest of the world where asbestos was less well controlled or remains uncontrolled. Equally important is this: we owe it to our own scientific integrity.

References

1. Hodgson JT, McElvenny DM, Darnton AJ, Price MJ, Peto J. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *Br J Cancer.* 2005;92:587-93.
2. Treasure T, Sedrakyan A. Pleural mesothelioma: little evidence, still time to do trials. *Lancet.* 2004;364:1183-5.

3. Butchart EG, Ashcroft T, Barnsley WC, Holden MP. Pleuropneumonectomy in the management of diffuse malignant mesothelioma of the pleura. Experience with 29 patients. *Thorax.* 1976;31:15-24.
4. Vogelzang NJ, Rusthoven JJ, Symanowski J, Denham C, Kaukel E, Ruffie P, et al. Phase III study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma. *J Clin Oncol.* 2003;21:2636-44.
5. Treasure T, Waller D, Swift S, Peto J. Radical surgery for mesothelioma. *BMJ.* 2004;328:237-8.
6. Sugarbaker DJ, Flores RM, Jaklitsch MT, Richards WG, Strauss GM, Corson JM, et al. Resection margins, extrapleural nodal status, and cell type determine postoperative long-term survival in trimodality therapy of malignant pleural mesothelioma: results in 183 patients. *J Thorac Cardiovasc Surg.* 1999;117:54-63.
7. Rusch VW, Rosenzweig K, Venkatraman E, Leon L, Raben A, Harrison L, et al. A phase II trial of surgical resection and adjuvant high-dose hemithoracic radiation for malignant pleural mesothelioma. *J Thorac Cardiovasc Surg.* 2001;122:788-95.
8. Tan C, Swift S, Gilham C, Shaefi S, Fountain SW, Peto J, et al. Survival in surgically diagnosed patients with malignant mesothelioma in current practice [abstract]. *Thorax.* 2002;57iii:36.
9. Su S, Jaklitsch M, Sugarbaker D. Treatment of malignant pleural mesothelioma: is there a benefit to pleuropneumonectomy? In: Ferguson M, editor. Difficult decisions in thoracic surgery: an evidence based approach. London: Springer-Verlag; 2007. p. 451-60.
10. http:/www.brighamandwomens.org/excellence/COEsissy.aspx. One patient's story: Sissy. http://www.brighamandwomens.org/excellence/COEsissy.aspx [2006 Nov 27].
11. Flores R. Debate: surgery for malignant pleural mesothelioma: a guaranteed complete response. *Lung Cancer.* 2006;54(Suppl 1): S42.
12. Treasure T. The evidence on which to base practice: different tools for different times. *Eur J Cardiothorac Surg.* 2006;30:819-24.
13. Treasure T. From anecdote to EBM. *J R Soc Med.* 2006;99:267-70.
14. Kumar A, Soares H, Wells R, Clarke M, Hozo I, Bleyer A, et al. Are experimental treatments for cancer in children superior to established treatments? Observational study of randomised controlled trials by the Children's Oncology Group. *BMJ.* 2005;331:1295.