1

CHAPTER 1

Incidence of Morbidity Following Thyroid Surgery: Acceptable Morbidity Rates

Paolo Miccoli,[1] *Michele N. Minuto*[2] *and Mario Miccoli*[3]

[1] Department of Surgery, University of Pisa, Pisa, Italy
[2] Department of Surgical Sciences (DISC), University of Genoa, Genoa, Italy
[3] Department of Experimental Pathology, University of Pisa, Pisa, Italy

Introduction

The issue of complications in surgery is a very difficult topic to deal with. Few surgeons speak openly about their problems, many are tempted to under-rate their own incidence, and even debates in the most important international circles about complications may fail to fully encompass the scope of the problem.

Unfortunately, since the dawn of surgery, complications have been inescapable, although undesired, elements of the surgical discipline but they have also allowed surgery itself to constantly improve.

In the new century, surgeons should deal with patients undergoing surgery under their care in a completely different way. The road leading to the operation itself starts well before surgery, when the patient is informed about his operation, the way it will be performed and the possibility and incidence of relevant complications. The number of complications that a surgeon generally shares with the patient before surgery requires judgement; informed consent should be obtained after a thorough discussion of the common problems that might occur after surgery, starting from the possibility of a keloid scar (an event that is usually not related to the surgeon) to intraoperative or postoperative death, more often unrelated to surgery but due to other co-morbidities.

In between these two exceptional events, there is the real intraoperative complication that is directly or indirectly caused by the surgeon (iatrogenic) but that is not necessarily due to negligence.

Modern surgeons should be aware of how to deal with the complication and therefore instruct and start to treat the patient themselves or, at the very least, to correctly refer the patient to a relevant specialist.

In thyroid surgery, complications that may arise after surgery may vary from those that might be immediately life-threatening but resolve after proper treatment, often leaving no sequelae, to relatively minor problems that are immediately evident and can therefore cause significant impairment of the patient's quality of life. The management of those patients experiencing post-thyroidectomy sequelae can be difficult, and this book contains suggestions to help every surgeon properly manage their own patients both intra- and postoperatively, helping them to determine the possible options to deal with a selected complication.

Thyroid Surgery: Preventing and Managing Complications, First Edition. Edited by Paolo Miccoli, David J. Terris, Michele N. Minuto and Melanie W. Seybt.

Morbidity of thyroid surgery

Every experienced surgeon is aware that the incidence of intra- or postoperative complications in thyroid surgery is relatively common, starting from the 'frequent' postoperative hypoparathyroidism (transient in the vast majority of cases) that in some reports has a frequency as high as 53% [1, 2].

The relative rarity is also dependent upon the method of analysis: although the single morbidity (e.g. permanent recurrent nerve injury) may be uncommon, when looking at the total incidence of the complications as a whole, the incidence of morbidity rises sharply. The rarity of a complication is also strictly related to the overall activity of the surgical practice (and therefore to the experience of the surgeon); a surgeon performing 10 thyroidectomies every week may see an injury of the recurrent nerve more often than another good surgeon who performs 60 thyroidectomies per year, even if the first is unquestionably more experienced than the latter.

The literature contains many series with an almost 0% incidence of complications that cannot be considered straightforward. How can this happen? Every experienced thyroid surgeon is perfectly aware of the issues behind such a low incidence of complications, but an inexperienced one might be misled by the results, and legal operators and lawyers might use them to manipulate facts, twisting the relatively common events and turning them to an evidence of malpractice.

We would therefore like to address the complications issue in a different way than that of a single experience reported in literature, aiming to show every surgeon how to interpret the commonly reported results, and how a sound and thorough study of complications should be conceived, in our opinion.

- When dealing with a specific complication of thyroid surgery, it is necessary to contrast our own incidence of the single event with the general incidence as reported in literature; this comparison should be made with series that are similar in terms of numbers. Going deeper into the issue, a 0% incidence of a selected complication in a series of 100 patients is a good result indeed, but if the event in question has a very low incidence, this does not represent a significantly different result from that obtained by another surgeon who reports a single one.
- This leads to the issue of statistically significant numbers, which will be better developed later in this chapter. Due to the fact that a complication is a relatively uncommon event, when analysing the results reported by other authors, the series should have sufficient numbers to have statistical relevance. It is easy to understand that a 0% incidence of permanent recurrent nerve lesions, reported in a prospective series of 33 patients in a study designed to investigate the oncological thoroughness of minimally invasive video-assisted thyroidectomy versus conventional thyroidectomy, cannot be interpreted as a statement that the rate of recurrent nerve palsy in thyroid surgery for cancer should be 0, for example. Since the paper was not planned to investigate the incidence of complications, the numbers are clearly too limited for this. Nevertheless, it was necessary to report this result in the paper, since it has an important clinical (but no statistical) value.

Further in this chapter, we give the readers some information about how to interpret statistical data from the literature, and introduce some basic statistical notions on uncommon events such as surgical complications. These simple concepts should be the basis of any audit conducted within a surgical unit.

Acceptable rates of thyroid surgery complications

We will hereafter deal only with the two principal complications of this surgery: recurrent nerve injury (RNI) and hypoparathyroidism. All other issues will be thoroughly analysed in the relevant chapters. The data reported will be drawn from the most important experiences (strictly in terms of number of patients analysed) available from the literature.

Injury of the inferior laryngeal/recurrent nerve

This complication is generally considered the worst for its potential impact on the patient immediately after surgery and for its significant consequences on the patient's future quality of life. The event causes a major impairment in one of two situations: the voice (with the onset of typical dysphonia) or the ventilation, and the related symptoms are generally present in an inverse ratio. When analysing the incidence reported by various authors, the reader should be aware of the following parameters.

- The series should take into consideration a significant number of patients (see after in this chapter), and one should be aware that the incidence reported can be obtained from the total number of patients in the study or from the total number of nerves at risk (that may double the sample, if only patients undergoing a total thyroidectomy have been selected for the analysis).
- Is the series mixing cases of thyroidectomies for benign and malignant diseases and primary and reoperative surgery? The incidence of a RNI (as well as of hypoparathyroidism) is invariably higher when a thyroidectomy for cancer (possibly associated with a central neck dissection) is performed or when the operation comes after a previous surgery. The morbidity is also significantly increased when performing a thyroidectomy for a particularly aggressive cancer subtype; the more aggressive the tumor, the higher the possibility of RNI, as described by a multicentre study that includes almost 15,000 patients [3].
- Have the authors reported whether their results were calculated on the basis of routine postoperative laryngoscopy or only on the basis of the postoperative discomfort or voice alteration of the patient? It is well known that a RNI can exist also in the presence of a remarkably normal voice. Also, a preoperative laryngoscopy should be performed in every patient undergoing thyroidectomy, since evidence of preoperative paralysis of a vocal cord is present in as many as 1.8% of patients; although in the majority of them it relates to previous surgery, the rate of this unexpected finding is still significant (six out of 14 patients without any previous surgery in the series described by Echternach *et al.*) [4]. When either pre- or postoperative laryngoscopy is absent, the real incidence of RNI will be significantly affected, decreasing when a postoperative laryngoscopy is not routinely performed and, on the other hand, unjustly assigning complications to the surgeon when such a preoperative examination has not been done.
- Finally, when reporting the incidence of RNI, one should always check if the patients have been followed up for at least 6 (or 12) months, to have the possibility of dividing the transient lesions (that last for 12 months at the longest and then spontaneously resolve, leaving no sequelae) from the permanent ones.

An analysis of selected papers dealing with more than 500 cases [3–12] is summarized in Table 1.1. These represent the most reliable papers dealing with the incidence of complications following thyroid surgery. These published data allow one to show either a high or a low incidence of RNI following thyroid surgery; it is immediately evident that the results demonstrate wide variability in the incidence reported by experienced thyroid surgeons.

Table 1.1 Reported incidence of transient and permanent RNI in studies considering more than 500 patients.

Author	Patients/ nerves at risk	RNI (transient/ permanent)
Lo *et al.* [11]	500/787	5.2/0.9
Toniato *et al.* [7]†	504/1008	2.2
Chiang *et al.* [10]	521/704	5.1/0.9
Steurer *et al.* [12]	608/1080	3.4/0.3 (benign disease) 7.2/1.2 (malignant disease)
Lefevre *et al.* [9]§	685/n.a.	n.a./1.5
Efremidou *et al.* [6]*	932/1864	1.3/0.2
Echternach *et al.* [4]	1001/1365	6.6
Bergamaschi *et al.* [5]	1163/2010	2.9/0.3
Thomusch *et al.* [8]*	7266/13436	2.1/1.1
Rosato *et al.* [3]	14934/n.a.	3.4/1.4

*Only patients undergoing surgery for benign diseases.
†Only patients undergoing surgery for malignant diseases.
§Only patients undergoing surgery for recurrent thyroid disease.
n.a., not analysed.

Recurrent nerve injury has an incidence ranging from 0.3% described by Bergamaschi *et al.* [5] to 6.6% reported by Echternach *et al.* [4]. When we analyse their results more carefully, we can observe that Bergamaschi *et al.* report on a huge series (1192 operations and 2010 nerves at risk), dominated by benign disease (>90%) and reflecting a majority of patients who underwent less than total thyroidectomy (622), an operation that is less morbid than a total thyroidectomy. In contrast, the series reported by Echternach *et al.* reveals a significantly higher rate of RNI, but this result does not take into account the rate of transient and permanent lesions, since it does not have laryngoscopy follow-up 6 months after the operation, and therefore it cannot be used for a proper analysis of permanent RNI. In between these two extremes, the real and expected incidence of RNI exists.

When we consider the different series homogeneously, we can see how the reported incidence of RNI is similar for any experienced thyroid surgeon. In the studies reporting exclusively on benign diseases, the incidence appears very low (0.2% according to Efremidou *et al.* [6], who report their results on almost 2000 nerves at risk), whereas Toniato *et al.* [7], who describe surgeries for thyroid cancer only, report a 2.2% incidence of the same complication. To our knowledge, no study including a significant number of patients undergoing thyroid surgery for any indication (arguably consisting of more than 1000 nerves at risk, according to the authors of this chapter) reports a global incidence of RNI of less than 1%, whereas series electively dealing with surgery for benign non-recurrent thyroid diseases can obtain (but do not necessarily achieve) significantly better results. This result can be associated with a more or less aggressive surgery demanded by the nature of the disease itself. To support this speculation, we can observe that many large series still report a high incidence of less than total thyroidectomies performed when benign thyroid disease is preoperatively diagnosed, whereas when describing surgery for thyroid cancer not only are total thyroidectomies performed but they can be variously associated with central neck dissections. The results obtained by Rosato *et al.* [3] describe a significantly higher incidence of RNI when surgery is performed for an aggressive cancer (papillary and follicular < medullary < anaplastic), confirming the idea that the more aggressive the surgery is, the higher the possibility of an iatrogenic lesionis.

In conclusion, the incidence of RNI is indeed very low when a thyroidectomy is performed for a benign thyroid disease (generally less than 1%), but higher-risk groups exist that contribute to a significant rise in its incidence. These groups, as demonstrated by large experiences, include patients undergoing surgery for thyroid malignancy and those undergoing surgery for any recurrent thyroid disease. In these populations, the incidence of RNI is generally over 1% and can be as high as 2.2%. This is particularly true for postoperative hypoparathyroidism, which is well supported by results obtained from the literature.

Hypoparathyroidism

As for RNI, some general points should be raised before thoroughly analysing the incidence of this complication.

- Temporary hypoparathyroidism is not an uncommon event, especially in selected situations such as surgery for thyroid cancer, often associated with central lymph node dissection, or surgery for Graves' disease. Therefore, one should determine whether the experience reported is composed of patients selected for a certain diagnosis or if the different indications have been co-mingled, significantly affecting the true incidence of the event.
- Many papers dealing with complications fail to distinguish between different types of surgeries such as lobectomy and total thyroidectomy, alone or associated with various neck dissections: this is another important issue to verify since, as already stated, different operations have significantly different results.
- How do the authors define the term 'hypoparathyroidism'? Do they refer to a biochemical finding (this significantly increases the incidence of the problem) or to the symptoms triggered by the hypocalcaemia (a rarer circumstance)?

Table 1.2 Reported incidence of transient and permanent hypoparathyroidism in studies considering more than 500 patients.

Author	Patients	Hypoparathyroidism (transient/permanent)
Toniato *et al.* [7]†	504	6.3
Lefevre *et al.* [9]§	685	5/2.5
Efremidou *et al.* [6]*	932	7.3/0.3
Bergamaschi *et al.* [5]	1163	20/4
Thomusch *et al.* [8]*	7266	6.4/1.5
Rosato *et al.* [3]	14934	8.3/1.7

*Only patients undergoing surgery for benign diseases.
†Only patients undergoing surgery for malignant diseases.
§Only patients undergoing surgery for recurrent thyroid disease.

The results obtained from the most important papers published in the literature [3,5–9] are summarized in Table 1.2.

Hypoparathyroidism, including both its transient and permanent forms, is a more common issue following thyroid surgery than RNI, and can therefore be better analysed through series less important in strictly numerical terms. Its occurrence is reported to be between 0.3% and 6.3% (permanent hypoparathyroidism), and between 5% and 22% (transient hypoparathyroidism).

The lowest incidence of permanent hypoparathyroidism in recent literature has been described in the study by Efremidou *et al.* [6], that focuses exclusively on patients with benign thyroid disease, whereas the highest (6.6%), reported by Toniato *et al.* [7], considers only patients undergoing surgery for malignant disease. In between these extreme results lies the true incidence of this complication, that is generally present in more than 1% of cases and is described to be significantly higher in some specific groups (higher-risk groups), such as patients undergoing more extensive surgery than total thyroidectomy alone (e.g. when central neck dissection is performed) and in patients undergoing reoperations.

A thorough analysis of the literature can easily demonstrate many studies reporting an incidence of permanent hypoparathyroidism close to 0%. These studies generally aim at demonstrating the efficacy of the parathyroid autograft in preventing permanent hypocalcaemia (dealt with in Chapter 15), and include insufficient patients from which to draw conclusions on the true incidence of this morbidity. In older studies reporting a very low incidence of permanent hypoparathyroidism, this result may be affected by a high incidence of less than total thyroidectomies, that were performed with the purpose of obtaining a lower complication rate than that obtained with a thorough extracapsular total thyroidectomy.

In conclusion, when a comprehensive analysis of the results reported in the literature is performed, the evidence is that every experienced thyroid surgeon, treating every kind of thyroid pathology, cannot obtain a complication rate of less than 1% for either permanent RNI or hypoparathyroidism. The literature can also demonstrate that the incidence rate of such complications can be higher than 6%, in particular situations, even for the experienced thyroid surgeon.

After this review of the literature, aimed at ascertaining the average incidence of the most specific adverse events after thyroid surgery, we give below a quick explanation of the basis of a proper statistical analysis, and how it should be conducted, when dealing with an uncommon or rare event.

Statistical and epidemiological analysis to study the complications of thyroid surgery

Surgical complications are relatively uncommon and this should be kept in mind when a study is designed to analyse the outcome of an operation, but also when a comparison between surgical techniques is needed. Even the rarest events should be analysed through the inferential statistics and/or a thorough epidemiological analysis, that can be more or less complicated. For example, when two different techniques need to be compared, one should consider epidemiological data (gender, age of patients), temporal circumstances influencing surgery (different surgeons operating, different techniques or instruments), and other factors. A sporadic event should never be statistically analysed on the grounds of its rarity; on the contrary, a more careful and precise analysis is needed to obtain reliable results.

What is immediately evident to the expert's eye is the absence of a correct analysis of the statistical power in the vast majority of studies published in the common literature, that are therefore generally lacking any analysis on the numbers necessary to correctly draw statistically relevant conclusions on the results reported. In the same way, only a few studies report analysis of the correct mathematical functions needed to correctly investigate the issue being studied.

What exactly is the 'statistical power' of the study? To answer this question, it is necessary to introduce the 'type II error', the error of failing to reject a null hypothesis when the alternative hypothesis is true (in less technical but more friendly words, it is the possibility of obtaining a 'false-negative' result). The opposite of this situation, or 'the right conclusion on the correct statistical significance', is strictly related to the statistical power of the analysis, that defines when the right conclusions can be drawn ('true positive' or, more technically, when the null hypothesis can be correctly rejected).

In strictly mathematical systems, the type II error is labelled with the β symbol, and has a value between 0 and 1. The statistical power is its complementary, as expressed by the formula:

$$\text{Statistical power} = 1 - \beta$$

The statistical power is conventionally considered adequate when $1-\beta \geq 0.8$, and can be calculated in two different ways: *ex-ante* (Latin for 'before') or *ex-post* (after).

The analysis *ex-ante* allows determination of the number of subjects necessary to draw statistically relevant conclusions for a planned experiment or study before this has started. This analysis gives important information to the investigators about the feasibility of the research, and the time and resources needed for the study to be completed. On the other hand, the *ex-post* analysis is made after the enrollment of the subjects once the study has finished, and its rationale is to verify if the sample in analysis is sufficient to guarantee an appropriate statistical result.

The statistical power can be obtained using either nominal variables (e.g. the presence or absence of an anticipated complication) or continuous variables (e.g. operative time, incision length). The different statistical tests have their own formulas to determine the statistical power.

Examples of how to calculate the statistical power

We will assess the statistical power of an analysis performed to evaluate whether two different surgical techniques have significantly different complications.

A preliminary evaluation revealed that the expected incidence of complications for the two different techniques is 2% for the traditional operation and 1% for the new one. When dealing with such rare events, the number needed for a thorough statistical analysis will be extremely high. Different tests can be used to determine the statistical power for our study, and we will use in this example the free software 'R', version 2.12.1, available from the following internet address: www.r-project.org/.

The lowest power requested is 0.8, the lowest statistical threshold is generally 0.05, and the expected complications for the two different operations are 1 (p1) and 2 (p2)%, respectively.

On the 'R' software we will insert the following instructions:

```
power.prop.test (p1 = 0.01, p2 = 0.02,
   sig.level = 0.05, power = 0.8)
Two-sample comparison of proportions power
   calculation
n = 2318.165
p1 = 0.01
p2 = 0.02
sig.level = 0.05
power = 0.8
alternative = two.sided
NOTE: n is number in *each* group
```

The result obtained is n = 2318.165, which means that 2319 patients are needed *in each group* to draw reliable conclusions on the significant results that might be obtained by the statistical analysis performed.

Let's now assume that, during the study period, the *real* incidence of complications of the two techniques was revealed to be 27 out of 2319 when patients were operated on with the new technique, and 52 out of 2319 patients undergoing surgery with the traditional one. Through a simple chi-square analysis we obtain the following result:

```
prop.test (c(27,52),c(2319,2319))
2-sample test for equality of proportions with
   continuity correction
data: c(27, 52) out of c(2319, 2319)
X-squared = 7.4175, df = 1, p-value = 0.006459
alternative hypothesis: two.sided
95% confidence interval:
−0.018653104 −0.002907914
sample estimates:
prop 1 prop 2
0.01164295 0.02242346
```

The p-value obtained by this analysis is 0.006459, a significant result (<0.05) that allows one to draw conclusions about the incidence rate of complications, in favour of the most innovative technique over the traditional one. This result expresses that the possibility of error we can make when asserting that the two techniques are significantly different in terms of complication rate is low, since this result has been obtained through a statistically robust experience.

Let's now assume that, for example, the two populations studied had been lower and the complication rate had been 19 with the innovative technique and 35 with the traditional one. We would have obtained the following result:

```
prop.test (c(19,35),c(1500,1500))
2-sample test for equality of proportions with
   continuity correction
data: c(19, 35) out of c(1500, 1500)
X-squared = 4.243, df = 1, p-value = 0.03941
alternative hypothesis: two.sided
95% confidence interval:
−0.0208406889 −0.0004926444
sample estimates:
prop 1 prop 2
0.01266667 0.02333333
```

This result would also have indicated a statistically significant result (p <0.05): let's now verify the statistical power of the study with such results with a '*post hoc*' test:

```
power.prop.test (p1 = 0.0127, p2 = 0.0233,
   sig.level = 0.05, n = 1500)
Two-sample comparison of proportions power
   calculation
n = 1500
p1 = 0.0127
p2 = 0.0233
```

sig.level = 0.05
power = 0.588493
alternative = two.sided
NOTE: n is number in *each* group

The result of this test indicates that even though a statistically significant threshold has been reached with the previous test (p-value = 0.03941), the population enrolled in the analysis is not relevant enough to obtain a statistically reliable result, since a 42% possibility of error (when the power is 0.58) exists to commit a type II error when considering accurate this p-value.

From a statistical point of view the *ex-post* and *ex-ante* tests have the same validity.

The tests analysed can obviously be used also when the groups compared are more than two or composed of different numbers of subjects.

The previous examples show that when there is the need to perform a statistical analysis on rare events and on groups that can be similar, it is necessary to enroll a huge number of cases to demonstrate significant results. This is generally the case for studies dealing with surgical complications, that need an analysis with sufficient statistical power. On the other hand, when critically analysing a study about the complications issue, it is necessary to verify its statistical power to find out if the results are reliable.

When further considering the complications issue, it is necessary to introduce other statistical considerations, that can appear slightly more complicated in the beginning, but can be easily managed by every reader.

The studies on surgical complications tend to be performed through statistical tests based on nominal variables (a nominal variable is one that has two or more categories, without intrinsic ordering to the categories), such as the chi-square, the odds ratio or the logistic regression.

Various theorems of the central limit (e.g. the DeMoivre–LaPlace law) state that when the size of the sample tends to infinity, the sum of the random variables tends to lot as a normal casual one. These theories, although complicated, are particularly useful when considering rare events that need extremely large samples for a correct statistical analysis. Their final result is to allow the use of statistical tests that are used to study continuous Gaussian variables. This means that, in particular situations, a t-test can be used to evaluate the rare events in an analysis instead of a non-parametric test, or a multiple linear regression instead of a multiple logistic one. It is obviously not mandatory to use a test used for the evaluation of Gaussian variables in the presence of large samples; a statistician can decide to ignore the possibility given by the central limit theory and use instead a test for nominal variables.

It is necessary here to reiterate that the statistical power should also be calculated in these situations, since there are formulas available to evaluate it when using multivariate analysis.

When a project is set up to study a continuous variable (e.g. evaluating the severity of complications, the operative time, the length of an incision) and a sample of sufficient size to allow the use of the central limit theory cannot be obtained, it will be necessary to evaluate whether the variable in analysis shows a Gaussian distribution or not. This preliminary analysis can be done either graphically or by using a preliminary test, such as the Bartlett test, Fligner–Killeen test, Brown–Forsyth test, Hartley test, Cochran method or Levene test. When the desired variable does not follow a normal distribution, the power test will be a non-parametric test, such as the Mann–Whitney or Kruskal–Wallis.

It is not possible to show here every power test that can be used in different analyses, but it is worth noting that every statistical software program contains all the tests necessary for different situations.

Finally, it is important to underline the necessity of a preliminary statistical analysis when evaluating the desired aims of a study. During this preliminary analysis, it is essential to determine whether is necessary to demonstrate if a statistically significant difference is present or if an anticipated result is not different among the different samples. For example, if a researcher wants to demonstrate that the operative times of two different surgical operations are not statistically different, the aim of the study will be to demonstrate an equivalence and not a difference.

In such a project, it is not adequate to use a simple t-test aimed at demonstrating the absence of a significant difference ($p < 0.05$), since in this case the absence of a statistically significant difference only states that we do not have enough encounters to conclude that the two operations have different results; a situation identical to that of a suspect who is discharged for lack of evidence: the verdict does not necessarily mean that he is 100% innocent.

When a researcher wants to demonstrate the similarity of different treatments, a *test for therapeutic equivalence*

Table 1.3 Example of a meta-analysis (see text).

Studies	Complications (TS*)	Non-events** (TS)	Complications (IS***)	Non-events (IS)
Study 1	2	96	1	102
Study 2	2	51	0	55
Study 3	3	170	1	120
Study 4	2	60	2	95
Study 5	1	50	1	50
Study 6	4	215	1	117
Study 7	0	70	1	70
Study 8	1	42	0	57
Study 9	3	315	1	321
Study 10	1	73	0	76
Study 11	3	418	2	433
Study 12	1	83	1	96
Study 13	1	36	1	49
Study 14	1	162	1	187
Study 15	1	84	1	97
Study 16	2	126	1	157
Study 17	1	53	0	55
Study 18	1	89	1	117
Study 19	1	97	1	109
Study 20	2	213	1	217

*TS, traditional technique;
**Non-events, number of operations without complications;
***IS, innovative technique.

should be used; on the other hand, a *non-inferiority test* can be used when trying to demonstrate that one treatment is not less effective than another. Those tests are often used for pharmacological studies but can also be used in different fields of medical research. A test of equivalence does not refer to a confidence interval but to an equivalence interval and the rules are different from those used for the tests that have been previously discussed. The power tests that should be used are also different from those previously examined, although the rationale is exactly the same.

The MBESS package available for the most recent versions of the 'R' software (www.r-project.org/) contains the equivalence tests and allows expert statisticians to perform the relative power analyses.

It is necessary to point out that in the scientific literature, the tests for therapeutic equivalence are not commonly used to demonstrate an equivalence between two different surgical operations, and the tests that are generally, and erroneously, used are the more 'traditional statistical tests' (the t-test, Mann–Whitney test, etc.).

How to perform a meta-analysis

Proper evaluation of statistically rare events (demanding extremely rich samples) is aided by the use of a meta-analysis, which will include many different studies published in the literature, thus reaching a significant sample size. When none of the studies published in the literature reaches a significant sample by itself, the studies can be considered together, thus obtaining a proper number of cases. However, this target cannot be reached simply by adding the samples from all the different studies; the rules for creating a meta-analysis are given below.

Let's suppose, once again, that a surgeon needs to compare the outcomes of two distinct operations, a traditional one (TS) and an innovative counterpart (IS), in terms of morbidity. First, it is necessary to build a table that summarizes the number of complications (or 'events') of the surgeries, and the number of operations without morbidity (or non-events). The different studies considered should be relatively homogeneous in terms of number of cases analysed, and the final number should reach that of an adequate sample, according to the result obtained by an *ex-ante* power test.

Meta Analysis

Model	Study name	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value
	Study 1	2,125	0,190	23,816	0,611	0,541
	Study 2	5,388	0,253	114,915	1,079	0,281
	Study 3	2,118	0,218	20,605	0,646	0,518
	Study 4	1,583	0,217	11,542	0,453	0,650
	Study 5	1,000	0,061	16,435	0,000	1,000
	Study 6	2,177	0,241	19,701	0,692	0,489
	Study 7	0,333	0,013	8,323	-0,669	0,503
	Study 8	4,059	0,161	102,101	0,851	0,395
	Study 9	3,057	0,316	29,547	0,965	0,334
	Study 10	3,122	0,125	77,882	0,694	0,488
	Study 11	1,554	0,258	9,346	0,481	0,630
	Study 12	1,157	0,071	18,781	0,102	0,919
	Study 13	1,361	0,082	22,495	0,215	0,829
	Study 14	1,154	0,072	18,602	0,101	0,919
	Study 15	1,155	0,071	18,747	0,101	0,919
	Study 16	2,492	0,223	27,799	0,742	0,458
	Study 17	3,112	0,124	78,091	0,690	0,490
	Study 18	1,315	0,081	21,307	0,192	0,847
	Study 19	1,124	0,069	18,209	0,082	0,935
	Study 20	2,038	0,183	22,638	0,579	0,562
Fixed		1,774	1,002	3,141	1,968	0,049

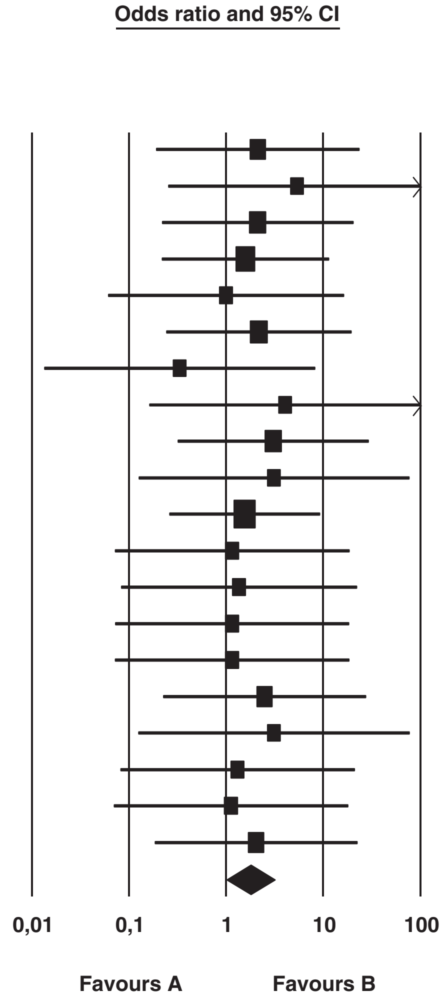

Figure 1.1 The results of the meta-analysis obtained by the 'R' software.

Table 1.3 summarizes an example of a meta-analysis. When all the patients in the 20 studies are considered, we obtain a significant population, which may demonstrate an adequate statistical power.

If we consider p1 and p2 values of, respectively, 0.01 and 0.02, the two samples are indeed 'strong' enough to be considered for a sound statistical analysis, since from the first example the sample needed was 2319, and the number of subjects here obtained is over 2500.

The statistical software www.meta-analysis.com will obtain the results summarized in Figure 1.1. It is easy to see that all the studies considered in the meta-analysis show p-values >0.05, and therefore are not statistically significant. The legend at the bottom of the figure represents the final result of the statistical analysis that takes into consideration all the 20 studies, demonstrating a p-value of 0.049 and an odds ratio of 1.774.

In conclusion, this meta-analysis works out the major issue of the size of the samples needed for a sound and powerful statistical analysis and, although contradicting the results of every single study, it represents their expression as a whole.

References

1 Pattou F, Combemale F, Fabre S, *et al*. Hypocalcemia following thyroid surgery: incidence and prediction of outcome. World J Surg 1998; 22(7): 718–24.

2 Olson JA, DeBenedetti MK, Baumann DS, Wells SA. Parathyroid autotransplantation during thyroidectomy: results of long-term follow-up. Ann Surg 1996; 223: 472.

3 Rosato L, Avenia N, Bernante P, *et al*. Complications of thyroid surgery: analysis of a multicentric study on 14,934 patients operated on in Italy over 5 years. World J Surg 2004; 28(3): 271–6.

4 Echternach M, Maurer CA, Mencke T, *et al*. Laryngeal complications after thyroidectomy: is it always the surgeon? Arch Surg 2009; 144(2): 149–53.

5 Bergamaschi R, Becouarn G, Ronceray J, Arnaud JP. Morbidity of thyroid surgery. Am J Surg 1998; 176(1): 71–5.

6 Efremidou EI, Papageorgiou MS, Liratzopoulos N, Manolas KJ. The efficacy and safety of total thyroidectomy in the management of benign thyroid disease: a review of 932 cases. Can J Surg 2009; 52(1): 39–44.

7 Toniato A, Boschin IM, Piotto A, *et al*. Complications in thyroid surgery for carcinoma: one institution's surgical experience. *World J Surg* 2008; 32(4): 572–5.

8 Thomusch O, Machens A, Sekulla C, *et al*. Multivariate analysis of risk factors for postoperative complications in benign goiter surgery: prospective multicenter study in Germany. World J Surg 2000; 24(11): 1335–41.

9 Lefevre JH, Tresallet C, Leenhardt L, *et al*. Reoperative surgery for thyroid disease. Langenbecks Arch Surg 2007; 392(6): 685–91.

10 Chiang FY, Wang LF, Huang YF, *et al*. Recurrent laryngeal nerve palsy after thyroidectomy with routine identification of the recurrent laryngeal nerve. Surgery 2005; 137(3): 342–7.

11 Lo CY, Kwok KF, Yuen PW. A prospective evaluation of recurrent laryngeal nerve paralysis during thyroidectomy. Arch Surg 2000; 135(2): 204–7.

12 Steurer M, Passler C, Denk DM, *et al*. Advantages of recurrent laryngeal nerve identification in thyroidectomy and parathyroidectomy and the importance of preoperative and postoperative laryngoscopic examination in more than 1000 nerves at risk. Laryngoscope 2002; 112(1): 124–33.