

# Chapter 17

## Principles of Analysis

Gary R. Johnson and Tassos C. Kyriakides

### Introduction

The analysis plans for data from clinical trials involving operative procedures are driven by the special design features and definitions of study outcomes that arise for these types of studies. In this section, we will discuss some of the nuances and precautions for planning and conducting the analysis of clinical trials comparing operative procedures or devices and for clinical trials comparing operative procedures to other nonoperative procedures.

### Developing the Analysis Plan

An analysis plan should be prespecified in the protocol and finalized prior to conducting any scheduled interim analyses or final analysis. The analysis plan must include measures to avoid bias in the comparison of the trial interventions. Two major concerns are to ensure that: (a) the periods of risk are handled equally and (b) the study population to be included in the primary analysis is defined by the randomization assignment. This is usually handled by the following recommended practices:

---

G.R. Johnson (✉) · T.C. Kyriakides  
VA Cooperative Studies Program, Office of Research and Development,  
VA Connecticut Healthcare System, Cooperative Studies Program  
Coordinating Center (151A), 950 Campbell Avenue,  
West Haven, CT 06516, USA  
e-mail: gary.johnson4@va.gov

T.C. Kyriakides  
e-mail: tassos.kyriakides@va.gov

1. Define that the follow-up time for surveillance for study outcomes begins at the time of randomization, and define the parameters for the duration of that observation period such that the observation period is balanced for each intervention and occurrences of unexpected events are counted equally for all intervention groups.
2. Define what will be counted as an outcome for each intervention; this may be a single measure, for example, all-cause mortality, or may be a composite outcome, for example, graft failure, reoperation, and late sequelae.
3. Define the methodology to be used to compare outcomes (e.g., hazard ratios from life table regression, odds ratios from logistic regression, or the use of general linear models for comparison of means for a repeated measure of treatment effect).
4. Perform analyses using the intent-to-treat principle including all randomized participants. This will avoid bias that might occur by excluding participants who do not adhere to the assigned intervention. Problems with this type of non-adherence can also be reduced by designing the trial to minimize the time period between randomization and the completion of the assigned intervention or operative procedure.
5. Consider prespecified stratification factors to balance pre- and perioperative risk factors between interventions; these might be included as stratification factors in the randomization scheme or as planned covariates in the primary analysis.

Important elements to consider for inclusion in the statistical analysis plan beyond what is described in the initial study protocol are (Abridged from VA Cooperative Studies Program SAP guidance):

- Definitions of all primary and secondary endpoints
- Statements of hypotheses to be tested and the parameter estimation
- Levels of clinical and statistical significance (one-tailed or two-tailed)
- Description of the methods of analysis and presentation of results:
  - Rules for handling intervals in which study visits or assessments are scheduled to occur
  - Decision rules for the inclusion/exclusion of data in special cases
  - Definitions of compliance and adherence
  - Methods for handling multiple comparison methods
  - Use of baseline measurements in stratification or in adjustment of treatment effects
  - Specification of fixed or random effects models
  - Approaches for handling covariables or associated risk factors in the analyses
  - Rules for the calculation of derived variables
  - Rules for potentially early-stopping of the trial
  - Methods for handling missing data
  - Methods for handling outliers
  - Methods for handling withdrawals and protocol deviations

- Methods for point and interval estimation
- Description of any interim analyses, and specifications for any sample size reestimation
- Description of the content of what will be identified as the final statistical report (e.g., mock-up or templates of tables that summarize the planned interim monitoring and final analyses).

## **Analytical Approaches for Studies Evaluating Effectiveness of Operative Procedures**

*Intent-to-Treat Principle:* The primary analysis should proceed directly from the randomized assignment of individuals to the treatments being compared. This analytic strategy, known as intent to treat (ITT), requires that all randomized participants be included in the analysis according to their originally assigned treatment, regardless of what happens after the random assignment. Any analysis that either drops randomized participants from the analysis or, in the case they receive a different treatment than originally assigned, analyses them according to the treatment actually received and thus has the potential to introduce bias.

In some instances, a modified intent-to-treat analysis may be defensible for the primary analysis. A Modified Intent-to-Treat (MITT) plan may allow for dropping from analysis participants who never received the randomized treatment or received partial or very limited treatment. In these instances, the reason for not receiving the assigned treatment must be independent of the study intervention. Trials involving surgical procedures are usually not suitable for a modified intent-to-treat analyses. For example, a trial that compares two different surgical interventions where it is discovered after randomization that a patient is not a suitable candidate for the assigned intervention (but may have been able to receive the other intervention) would not be appropriate for a modified intent-to-treat analysis. Such a plan would introduce bias into the comparison of the interventions.

The primary analysis of a well-designed and well-conducted randomized trial may be very straightforward or very complex. A trial comparing 30-day complication rates for two surgical interventions could use a chi-square test for homogeneity to compare the proportion of participants who developed complications within 30 days in the two groups, or a more complex intervention trial might have the modeling of repeated measurements of intra- or perioperative markers of clinical risk as the primary analysis.

A common approach to evaluation of the success or failure of a procedure is to use survival or failure time analyses that compare the probability of survival or being event free during a follow-up period after randomization to a study intervention or procedure. Methodologies for analysis of survival data include test statistics for the comparison of survival distributions [1–3] and life table regression methods [4].

**Short-term outcomes:** When comparing two operative procedures or devices a time-to-event analysis may still be appropriate if the objective of the study is to assess the occurrence and time of peri- and postoperative events.

**Long-term outcome:** Survival analysis is more typically planned for comparing the long-term outcomes after a procedure where the outcome measure would include not only the short-term outcomes, but also late postoperative events and sequela and possibly recurring events.

**Repeated measures over time:** A trial may be designed to evaluate changes in repeated measurements over a specific time period. For example, repeated measurements of functional status, markers of clinical risk, severity of post-procedure pain or other symptoms, or measures of health-related quality of life may be used as primary or secondary outcomes. These types of data collected longitudinally and prospectively can be analyzed as dependent variables in mixed effects models. The analysis plan should identify the time points to be included in the analysis.

**Example: Carpal tunnel syndrome:** Participants randomized to endoscopic versus open procedures were assessed for postoperative pain (primary outcome) and other measures of functional status and quality of life at three weeks, six weeks, and three months, with additional assessments at 12 months [5]. While this study does not provide a good example of estimating an overall intervention effect over 12 months, it does demonstrate how repeated patient-reported outcomes can be analyzed to compare interventions.

## **Secondary and Supportive Analyses**

### ***Subgroups***

Complete the analysis of the comparison of intervention outcomes in prespecified subgroups by calculating the relative hazards ratios from models including the intervention and subgroup parameter and an interaction term. Presentation of these estimates of relative risk (e.g., relative risk estimate and 95% confidence intervals) in a table or a forest plot on the log scale will provide an easy way to assess the relative effects in different subgroups. The subgroups may be risk factors identified by previous research or observation. Prespecifying subgroups and specifying adjustments that will be made for multiple comparisons will help in the acceptance of study results, otherwise will be viewed as exploratory.

### ***Safety***

Adverse events for an operative procedure may overlap with events that have also been reported as study outcomes. For example, a complication of a surgical

procedure may be a component of a composite outcome [e.g., readmission for reoperation due to postoperative complication might be counted as a treatment-related hospitalization in a composite outcome].

### ***Supportive Analyses***

Supportive analyses should be included in the analysis plan such as analysis of cause-specific death in addition to all-cause mortality, and the analysis components of a composite outcome. The supportive analyses also include sensitivity analyses for alternatives to the primary intent-to-treat approach where the effect of operative interventions are assessed in a modified intent-to-treat population or other approaches that select a subset of the randomized population to be included in the analyses based in treatment assigned or received.

The analytic strategy known as a per-protocol analysis limits the analysis to participants who actually received or adhered to the randomly assigned intervention or strategy. The results of this type of analysis would not take precedence over the primary intent-to-treat analysis, but could provide additional supportive information that investigates the degree to which the intent-to-treat results may have been impacted by noncompliance with the randomized treatment assignment. Another approach is an as-treated analysis where the analysis groups are formed according to those who actually received the intervention, rather than according to the randomly assigned intervention. In this case participants who get the alternative intervention (crossovers) are grouped with those who adhered to the treatment per-protocol. Similar to per-protocol analysis, the as-treated analysis strategy is not based on the randomized intervention assignment and is inherently biased. These sensitivity analyses may not always produce results that are aligned with the primary analysis but if not done will leave a gap in the interpretation of the results. Of course, any potential biases in these supportive analyses should be recognized in any presentation.

### ***Example: REFLUX Trial [6]***

In this trial, participants were randomized to laparoscopic fundoplication surgical procedure or long-term medical treatment for gastroesophageal reflux disease. Main outcomes were disease-specific and general health-related quality of life measures and surgical complications. In the publication of the trial results, both the intent-to-treat and the per-protocol results were presented together with the justification of the large proportion of participants randomized to laparoscopic fundoplication (38%) who did not get the procedure. The adjusted treatment effect was greater in the per-protocol group (15.4, 95% confidence interval [CI]: 10.0, 20.9) than the intent-to-treat population comparison (11.2, 95% CI: 6.4, 16.0), and even

greater when the analysis was performed according to the treatment received (16.7, 95% CI: 9.7, 23.6) although with a wider confidence interval. In this example, the results were fortunately consistent for the three approaches and the supportive analyses were clearly identified and discussed as biased.

Thus, in a well-controlled randomized clinical trial, the intent-to-treat is considered the most conservative approach in the comparison of the study interventions and minimizes bias. Per-protocol and as-treated analyses provide a more direct comparison of the treatments actually received, but have more potential for bias because the randomized design is compromised.

Although intent-to-treat is theoretically an unbiased strategy, bias can still be introduced after the randomization. Outcome evaluation can be biased, especially if the outcome is subject to interpretation or subjective assessment with knowledge of the intervention assigned. Blinding of the treatment assignment protects against biased outcome evaluation, although this is often difficult to achieve in a study involving devices or operative procedures. Drop-outs or withdrawals of participants may occur during the follow-up and result in missing data, which can also introduce bias. This problem is not specific to trials of surgical procedures. Statistical methods to address the problem of bias due to missing data are the subject of another chapter in this book.

### *Cost-Effectiveness Analysis*

The protocol may plan for a cost-effectiveness analysis. Unless the study is completed in a setting where all costs can be identified (e.g., within one institution) these analyses usually are conducted on direct costs for the procedure and any complications or sequelae and do not include all indirect costs. The protocol may plan for only completing this exercise when a treatment effect has been found in the experimental arm, but such a comparison can be of value even when there is no significant difference. The active comparator group may be less costly than the control or standard of care group.

### *Interim Analyses*

Standard techniques can be applied for scheduling interim analyses. However, there should be special consideration for timing the analyses at a point when the necessary proportion of study events (information) defined a priori in the analysis plan has been observed, rather than at an enrollment or study duration milestone. For example, if the primary study outcome is postoperative status two years after randomization, enough two-year events need to be accrued before completing the analysis.

## Special Considerations and Cautions

### *Non-proportional Hazards*

The primary analysis can be conducted using a nonparametric approach such as the Kaplan–Meier method for calculating survival curves; this allows for a comparison of the interventions using the log-rank test without the assumption of proportional hazards. The events are weighted equally relative to the number of participants at risk no matter when the event occurs during follow-up. However, covariate adjusted analyses are usually planned and even with nonparametric approach, the results of the study can be misinterpreted. Therefore, an evaluation of possible time-dependent treatment effects and an assessment of hazard ratios over time should be undertaken. An extreme example of non-proportional hazards occurs when survival curves cross or hazard ratios change direction (relative to 1.0). In this case time-dependent effects need to be accounted for as well as the consideration of baseline hazard rates for risk factors [7]. In some cases, a piecewise analysis of segments of follow-up time might help explain the results.

### *Examples*

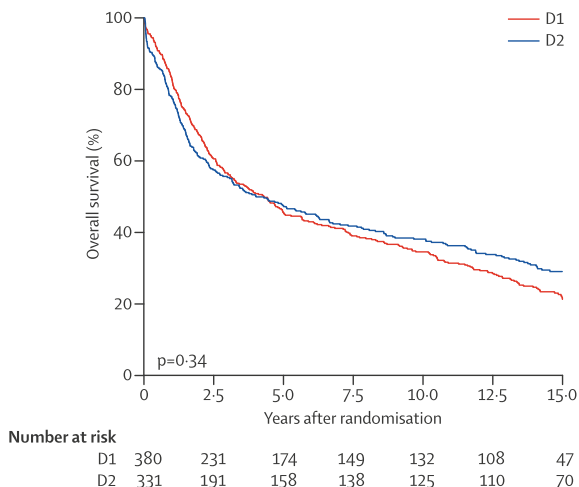
D1 versus D2 dissection for gastric cancer. Several trials on the treatment of gastric cancer demonstrate the problem of non-proportional hazards. The Dutch Gastric Cancer Trial [8] first presented the results of D1 versus D2 methods of dissection as having no difference in mortality with some early benefit of the D1 method, though acknowledging the non-proportional hazards. The survival curves crossed at approximately 4 years, with the subsequent hazard ratios in favor of D2. Long-term follow-up in this same population showed a long-term benefit of the D2 method [9] (see Fig. 17.1). Similarly, a concurrent study by the MRC comparing D1 to D2 showed non-proportional hazards. A methodology paper assessing the proportional hazards in the Dutch Gastric Cancer Trial demonstrated several methods for approaching this problem including time-dependent treatment and covariate effects and accounting for baseline hazards [7].

### *Learning Curve*

If the intervention trial is evaluating a new procedure or device, then the possibility of learning curve effects should be taken into account in the analysis, especially if the new procedure is being compared to a well-established procedure [10].

In addition, it is possible that during the course of a trial, multiple versions or modifications of the investigational device or procedure may be used in the same

**Fig. 17.1** Overall survival in patients treated with curative intent (N=711). D1=standard limited lymphadenectomy. D2=standard extended lymphadenectomy. Reprinted from [9] Example of non-proportional hazards. Reprinted From The Lancet Oncology with permission from Elsevier



study by intervention, by design, or by necessity to adapt trial interventions to changing technology. Some devices go through manufacturing revisions during the intervention phase of the trial, some might be withdrawn from the market, and each device might have a different period that it has been available for use. Thus, some consideration should be given to this in subsequent/sensitivity analyses. Major changes in device technology could introduce bias into a trial, especially in a watchful-waiting trial where an intervention group that gets the device or procedure early would not be a good comparator group if the group receiving the intervention later received a different version of the device or procedure.

Analyses when multiple sites treated within one subject: This needs to be considered when the randomization unit is one participant, but the procedure or intervention may be administered to many sites (e.g., multiple coronary grafts or stents or angioplasty to many vessels, or dental implants). The approach used by many trials is to rely on randomization or stratified randomization to balance the extent of disease in treatment groups, and to define the main outcomes as occurrence in any site (e.g., artery). Depending on the disease and possibility of varying outcomes depending on site, the analysis plan may need to take into account the measurement of outcomes for multiple interventions per randomized unit.

### ***Example: PREVENT IV Trial***

In this trial, two methods of preventing graft failure were compared in patients undergoing coronary bypass surgery (CABG) [11]. Vein grafts were treated ex vivo with either edifoligide or placebo in a pressure-mediated delivery system. The primary endpoint was all-cause mortality or 75% or greater stenosis of any graft.



Since patients were randomized and not arteries, the analysis may have been biased if there was an imbalance in number and which arteries were grafted. In secondary analyses both by patient and by graft, the generalized estimating equation (GEE) methods were used to adjust for the within-subject correlation among grafts.

## **Operative Versus Nonoperative Comparisons**

As discussed in other chapters, the comparison of an operative procedure to a nonoperative procedure needs to be carried out with precautions and special considerations in the analysis plan. Risks related to operative procedures are likely to be perioperative and early in the postoperative period, while the risks of not operating may be much later. Therefore, in the study protocol, the main hypothesis and objectives of the trial need to be explicitly stated over what time period the intervention comparisons will be made. To balance risk between operative and nonoperative interventions, a period after randomization equivalent to the operative-risk period can be defined for the nonoperative intervention for the surveillance of safety or effectiveness outcomes.

### ***Example: ADAM Study***

Time to death after abdominal aortic aneurysm (AAA) repair scheduled within 6 weeks after randomization compared to watchful waiting for symptoms of AAA growth or rupture before operating. [12]. In this trial, the secondary outcome of AAA-related death included deaths that occurred within 30 days after randomization for those randomized to surveillance, as well as deaths directly or indirectly caused by AAA rupture, AAA surgery, preoperative evaluation, late graft failure or complication, death related to recurrence of AAA after grafting, or any death occurring within 30 days after AAA surgery or any death.

Alternatively, another way to balance risk between operative and nonoperative interventions is to prespecify in the analysis plan that the initial procedure will not be counted as an outcome. In a trial randomizing one group to implantation of a device or a procedure, the hospitalization for the planned procedure might be excluded for the adverse experience analysis while the repeat of that operation is included.

### ***Example: COURAGE Study***

In this trial, time to death or nonfatal myocardial infarction after PCI procedure compared to intensive medical treatment only [13]. The secondary outcome of

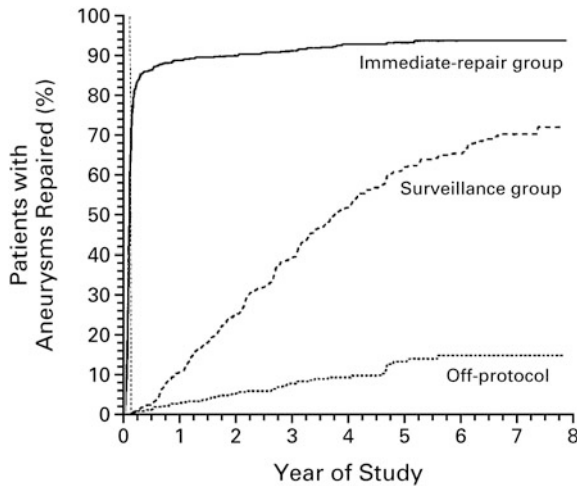
revascularization, did not count the initial percutaneous coronary intervention (PCI) in the PCI group as revascularization, and compared the number of patients requiring subsequent revascularization in the PCI group with all revascularization in the medical group.

Additional considerations for adherence to assigned intervention group:

In the case of watchful-waiting design studies and other intervention versus nonintervention analyses where measures must be taken to help reduce bias against penalizing the operative intervention over the nonintervention group, adherence to the assigned intervention should be considered for supportive analyses. If there is poor adherence to one or more interventions there may be poor separation of treatments. This might be due to a larger portion of randomized participants not receiving the intervention or a large proportion of the nonintervention participants crossing-over into the intervention arm. Adherence to per-protocol assigned intervention should be included in the presentation of results even though the primary analysis is by intention-to-treat.

### *Example*

In the ADAM Study of open repair of AAA (described above), randomization was to immediate AAA repair or to surveillance of changes in AAA size (growth to 5.4 cm) and symptoms of AAA rapid growth or symptoms of AAA rupture, and scheduling AAA repair when study criteria were met [14]. The cumulative proportion of immediate AAA repair participants operated on within 6 weeks was monitored for adherence to the study criteria, as well as AAA repairs that occurred later in follow-up. At 6 weeks, 72.7% of the immediate surgery group had AAA repair and by the end of the study 92.6% were repaired. In the surveillance group, 61.6% had AAA repair over the nearly 8 year of follow-up period (mean 4.9 years). Most repairs in the surveillance group were by protocol criteria, but the proportion of off-protocol repairs was also reported. These were where AAA repair was completed for AAA that did not meet the surveillance group criteria for repair; and occurred in 9.0% of the surveillance participants. In this example, there was good separation of intervention rates under protocol conditions, with a low crossover rates (low failure to have repair in the immediate surgery group, and low rate of AAA repair in the surveillance group prior to meeting criteria for repair). Even though this was a strategy trial of comparative effectiveness, the results of the trial would have been questioned had there not been a clear separation of interventions rates. (see Fig. 17.2).



**Fig. 17.2** Cumulative rate of repair of Abdominal Aortic Aneurysm, According to Treatment Group [14]. With permission from Massachusetts Medical Society

### ***Example VA CSP Study on Transurethral Resection of the Prostate (TURP) [15]***

Another example of separation of interventions in a watchful-waiting design is a trial comparing immediate surgery for benign prostatic hyperplasia with surveillance for symptoms before scheduling the procedure [15]. In this study, 89% of the surgery group underwent transurethral resection of the prostate within two weeks of randomization, while 24% had surgery over approximately 3 years.

## **References**

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53:457–81.
2. Peto R, Peto J. Asymptotically efficient rank invariant test procedures (with discussion). *J R Stat Soc Ser A.* 1972;135:195–206.
3. Schoenfeld DA, Tsiatis AA. A modified log rank test for highly stratified data. *Biometrika* 1987;74:167–75.
4. Cox DR. Regression models and lifetables. *J R Stat Soc.* 1972;34:187–220.
5. Atroshi I, Larsson GU, Ornstein E, Hofer M, Johnsson R, Ranstam J. Outcomes of endoscopic surgery compared with open surgery for carpal tunnel syndrome among employed patients: randomised controlled trial. *BMJ.* 2006;332(7556):1473.
6. Grant AM, Cotton SC, Boachie C, Ramsay CR, Krukowski ZH, Heading RC, Campbell K. REFLUX trial group. Minimal access surgery compared with medical management for gastro-oesophageal reflux disease: five year follow-up of a randomized controlled trial (REFLUX). *BMJ.* 2013;346:f1908.

7. Putter H, Sasako M, Hartgrink HH, van de Velde CJ, van Houwelingen JC. Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial. *Stat Med.* 2005;24(18):2807–21.
8. Bonenkamp JJ, Hermans J, Sasako M, van de Velde CJ, Welvaart K, Songun I, Meyer S, Plukker JT, Van Elk P, Obertop H, Gouma DJ, van Lanschot JJ, Taat CW, de Graaf PW, von Meyenfeldt MF, Tilanus H; Dutch Gastric Cancer Group. Extended lymph-node dissection for gastric cancer. *N Engl J Med.* 1999;340(12):908–14.
9. Songun I, Putter H, Kranenbarg EM, Sasako M, van de Velde CJ. Surgical treatment of gastric cancer: 15-year follow-up results of the randomised nationwide Dutch D1D2 trial. *Lancet Oncol.* 2010;11(5):439–49.
10. Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials.* 2004;1:421–7.
11. PREVENT IV Investigators. Efficacy and safety of edifoligide, an E2F transcription factor decoy, for prevention of vein graft failure following coronary artery bypass graft surgery PREVENT IV: a randomized controlled trial. *JAMA.* 2005;294(19):2446–54.
12. Lederle FA, Wilson SE, Johnson GR, Littooy FN, Acher C, Messina LM, Reinke DB, Ballard DJ. Design of the abdominal aortic Aneurysm Detection and Management Study. ADAM VA Cooperative Study Group. *J Vasc Surg.* 1994;20(2):296–303.
13. Boden WE, O'Rourke RA, Teo KK, Hartigan PM, Maron DJ, Kostuk WJ, Knudtson M, Dada M, Casperson P, Harris CL, Chaitman BR, Shaw L, Gosselin G, Nawaz S, Title LM, Gau G, Blaustein AS, Booth DC, Bates ER, Spertus JA, Berman DS, Mancini GB, Weintraub WS, COURAGE Trial Research Group. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med.* 2007;356(15):1503–16.
14. Lederle FA, Wilson SE, Johnson GR, Reinke DB, Littooy FN, Acher CW, Ballard DJ, Messina LM, Gordon IL, Chute EP, Krupski WC, Busuttill SJ, Barone GW, Sparks S, Graham LM, Rapp JH, Makaroun MS, Moneta GL, Cambria RA, Makhoul RG, Eton D, Ansel HJ, Freischlag JA, Bandyk D. Aneurysm Detection and Management Veterans Affairs Cooperative Study Group. Immediate repair compared with surveillance of small abdominal aortic aneurysms. *N Engl J Med.* 2002;346(19):1437–44.
15. Wasson JH, Reda DJ, Bruskevitz RC, Elinson J, Keller AM, Henderson WG. A comparison of transurethral surgery with watchful waiting for moderate symptoms of benign prostatic hyperplasia. The Veterans Affairs Cooperative Study Group on Transurethral Resection of the Prostate. *N Engl J Med.* 1995;332(2):75–9.