4 Nonrandomized Interventional Study Designs (Quasi-Experimental Designs)

David A. Axelrod, MD, MBA
and Rodney Hayward, MD

Contents

Nonrandomly Assigned Control (or Comparison) Group Studies
Time Series Analysis
Comprehensive Cohort Trials/Patient Preference Trials
Empirical Evidence Comparing Randomized and Nonrandomized Trials
Conclusions
References

In contrast to observational study designs, interventional studies manipulate clinical care to evaluate treatment effects on outcomes. Although surgeons have often relied on observational studies to establish the efficacy and effectiveness of operative and perioperative interventions, observational studies (also referred to as case series) are limited to demonstrating the correlation between the outcome of interest and the procedure. Prospective controlled interventional trials will provide a higher level of evidence for a true cause-and-effect relationship.

Interventional studies may be categorized into two large classifications: true experimental designs and quasi-experimental designs. The randomized, blinded clinical trial (RCT) is the prototypical example of a true experimental design. In an RCT, patients are allocated to treatment arms in a prospective, random fashion in an attempt designed to ensure comparability between groups. The intervention and outcome are then administered and recorded, often with blinding of the interventionalist, the evaluator and the subject to reduce bias. This study design is discussed further in Chapter 5.

Unfortunately, surgical interventions often do not readily lend themselves to randomized blinded trials *(1)*. Consent for randomization is often difficult to obtain for surgical interventions because patients may have a preconceived notion of what treatment they wish to receive, blinding is often impossible (e.g., the patient and surgeon both know whether a cholecystectomy was performed laparoscopically or through open surgery),

From: *Clinical Research for Surgeons*
Edited by: D. F. Penson and J. T. Wei © Humana Press Inc., Totowa, NJ

and ethical concerns usually render sham surgery controls unacceptable *(2)*. Furthermore, the technical nature of surgery can make randomization difficult. Surgeons are more likely to be skilled in certain operations, results are likely to improve over time as a result of learning curve effects, and surgery often involves small incremental improvement rather than dramatic changes. As a result, quasi-experimental techniques, which do not require random assignment, are more often used in the assessment of surgical interventions than in the assessment of other medical treatments.

The most basic experimental research design is a comparison of outcome before and after a planned intervention without the use of a control group (also known as the pre/post design). Essentially, this is a systematic case series in which a new intervention or treatment is introduced during the period of study *(3)*. Unfortunately, interpretation of simple pre/post intervention studies is difficult. Changes in the outcome of interest may be due to the intervention; however, it may also reflect disease natural history (as the condition improves over time or clinical therapy improves with experience), patient selection (patients before and after the intervention may have differed in clinically important attributes), or placebo effects (because neither patient nor provider is blinded). In addition, there is a natural tendency for processes to regress to the mean, which may occur without intervention. Therefore, in this chapter we will examine alternative study designs, which, although not randomized, can often provide more rigorous evidence of a treatment effect than a simple pre/post design.

In this chapter, three principal interventional study designs will be considered:

- nonrandomly assigned control (or comparison) group
- time-series design with pre- and posttest comparisons
- preference allocation (patient, physician)

For each research design, we will consider appropriate research questions, basic design elements, allocation of subjects, outcome measurement, analytic techniques, and overall assessment of the strength of the design.

1. NONRANDOMLY ASSIGNED CONTROL (OR COMPARISON) GROUP STUDIES

In nonrandomly assigned control group studies, at least two separate groups are evaluated—one of which receives the intervention of interest and another that serves as a control or comparison group (Figure 1). Thus the nonrandom control group is similar in design to a RCT, except that patients are assigned to treatment groups in a nonrandom fashion. Quasi-experimental designs differ from that of an observational trial in that the patients are allocated to treatment groups by research protocol, whereas in an observational study the natural history of treatments is studied (i.e., there is no allocation to any intervention).

1.1. Identify Appropriate Questions

There are two main instances in which a nonrandomized control group trial is a good choice. The first is when an RCT would be ideal but practical considerations (e.g., costs, unacceptability to patients or providers) make a high-quality RCT infeasible. The second is when you are trying to establish the effectiveness of large-scale dissemination and implementation. Still, just as in observational studies, predictive variables need to be identified and measured to ensure comparability between the study groups. As is also true of observa-

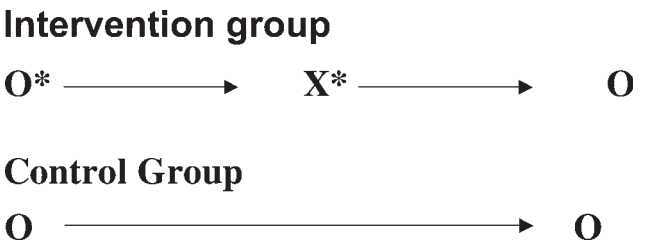

Figure 1: Nonrandomized control groups. *O represents observation; X* represents a study intervention.

tional studies, quasi-experimental studies are less desirable when studying outcomes that are multifactorial or are less well understood. The strength of these studies is in part determined by the investigator's ability to control for potential confounders using multivariable analysis, therefore, being able to identify and accurately measure these potential cofounders (such as patients' illness severity and comorbidities) is essential to minimizing the risk of bias.

1.2. Define Inclusion/Exclusion Criteria

Next, specific inclusion and exclusion criteria need to be established for the study population. Inclusion criteria must be identical for both the intervention and comparison groups. There is an inherent tension between using criteria that are broad enough to ensure that recruitment of an adequate sample and generalizability, but not so broad that meaningful comparison is not possible. For example, including all first-time uncomplicated hernias in men may be reasonable, whereas limiting the trial to asymptomatic hernias less than 2 cm may be unnecessarily restrictive. Exclusion criteria have two main purposes: (1) to exclude study subjects who present substantial risk to the scientific quality of the study (e.g., inability to follow up, plan to move out of state) and (2) to assure safe and ethical conduct of the study (e.g., inability to tolerate general anesthesia, contraindications to the one or both of the treatment arms, unable to give informed consent). Both clinical characteristics of potential study subjects and social/cognitive criteria should be considered when identifying exclusion criteria.

1.3. Estimate Sample Size

Finally, an estimate of the appropriate sample size needs to be determined. The size of the patient cohort is likely to be larger in a nonrandomized study to permit the application of multivariate regression techniques to adjust for differences in baseline characteristics. Although the mechanics of regression analysis are considered elsewhere in this text, it is important to remember that the number of independent variables that can be used in a regression model for a categorical outcome can be quite limited. About 10–20 outcomes (e.g., hernia recurrence) are required for each variable included in the regression analysis. Even if your study includes 1000 surgeries, if there are only 50 adverse outcomes you should include no more than 5 independent variables in the regression model *(4)*. For continuous outcomes variable (e.g., health-related quality of life or exercise tolerance) this restriction is considerably less (10–20 study subjects per independent variable). Clearly, the investigators need to ensure that there is adequate research support to recruit and follow a sufficient number of study subjects.

1.4. Allocate Subjects Between Groups

The selection of study sites and the allocation of subjects to treatment groups are among the most challenging issues in nonrandomly assigned control group studies. Subjects should be allocated to treatment groups in a manner that allows the groups to be generally comparable and to minimize the introduction of bias. For example, the investigator may choose to randomly select one of two comparable private hospitals as the intervention site and the other as the comparison site. Whether to randomize by hospital, ward or clinic, or physician will depend on feasibility, risk of contamination, and the nature of the intervention. The use of pseudo-randomization (e.g., every other patient) should be discouraged because it does not offer any great advantage over true randomization and is subject to manipulation by clinicians. The use of patient or physician preference to allocate patients to treatments can be used, but is less desirable. A discussion of preference trials is included later in this chapter.

1.5. Collect Baseline Data

In the nonrandomized controlled trial, it is crucial to collect a comprehensive dataset including all variables that can reasonably be expected to influence the outcome of the procedure. Baseline and follow-up (i.e., pre/post) measures should be collected at both the intervention site and the control sites. This data will allow the investigator to judge the comparability of the two groups and can also be used to statistically adjust for measured differences between the groups. In addition, pre/post data from the control sites can provide an external temporally synchronous control. Unfortunately, just as in observational studies, even with detailed data collection, it is still possible that unmeasured confounders will influence the study's results. This should always be reported as a limitation of this study design.

1.6. Measure the Outcome

The outcome of interest should be established before initiating the study and measured as accurately and reliably as possible. Because neither the patient nor the investigator is blinded to the nature of the procedure, the use of physician- and patient-reported outcomes can be quite problematic. The use of independent, blinded evaluators, imaging studies, or physiologic measurements (e.g., blood flow rate, degree, residual stenosis) may be less prone to bias than patient-reported outcomes. However, it is critically important that the outcome measure is clinically meaningful. For example, although measuring range of motion may be more objective and less prone to bias, it is also less clinically relevant than measures such as pain or ability to return to work. There may be some instances in which sham procedures or other placebo treatments are ethically acceptable and in such instance subjective outcome measured (such as pain, health-related quality of life) are likely to be much less subject to bias *(5)*.

1.7. Analysis of the Data

Although randomized controlled trials can be analyzed using relatively straight forward bivariate statistical analysis (e.g., *t*-tests, chi-squared statistics) when successful randomization is demonstrable, analysis of nonequivalent comparison groups generally requires multivariable modeling. First, the groups are analyzed to determine the degree of comparability using simple descriptive statistics. Bivariate statistics can be used, but it is critical to realize that both clinical and statistical significance of differences between

intervention and control subjects should be considered. For example, if there are clinically substantive differences in an important preintervention patient attribute(s), then those variables should be adjusted for in the analyses even if the bivariate difference was not statistically significant.

Next, multivariate regression techniques are used to "control" or "adjust" for any observed differences in baseline characteristics. Treatment assignment is entered as an independent variable controlling for these potential confounders and the effect of treatment is determined from the regression coefficient. Just as in a true experiment, this variables signifies intention to treat (i.e., was the subject in the assigned intervention or the control group), not whether the subject received the treatment. Standard multivariate analysis assumes measurement of all potential confounding variables (although if you know the degree of measurement error, adjustments for low or moderate precision can be performed).

There are several threats to the interpretation of data from a nonrandomized clinical trial, of which unmeasured confounders is particularly prominent. For example, a hernia may recur more frequently in one hospital because the patients are more likely to be poor and must return to work earlier. Because patients are usually not "blinded" to the study intervention, there may also be differential degrees of placebo effects that may account for the clinical differences, especially of outcomes based on patient self-report. Finally, the investigator must consider issues that are relevant to any trial, including RCTs, such as the need for complete follow-up, the ascertainment of an unbiased outcome assessment, and concerns regarding the generalizability of the findings into a nonexperimental setting. The inherent uncertainty of achieving complete case-mix adjustment has left some experts to question whether we can rely on these statistical methods to account for differences in the characteristics of the comparison groups *(6)*.

1.8. Advantages of the Nonrandomized Controlled Trial

When a true experiment is not feasible, there are several potential advantages of including a control group (even nonrandomized) instead of relying solely on simple pre- and postintervention comparisons. The control group principally helps to account for threats to internal validity from temporal trends, regression to the mean, and the learning curve. A temporal trend bias is the potential that other advances or changes in clinical care, the nature of the disease, or patient population may account for observed changes. As long as these changes are reflected in both the control and experimental groups, they are likely to be identified using this design. Similarly, the impact of the learning curve has been widely established for new surgical procedures. Thus outcomes may improve over time, which must be accounted for in any analysis. Finally, the outcomes at the extreme are likely to moderate naturally over time, leading to a phenomenon of regression to the mean. Without controlling for this trend, observed effects may reflect chance rather than true clinical changes.

The use of a nonrandomized control group may also reduce the threats to external validity that limit the value of RCTs results. First, RCTs tend to be done at a few, highly selected sites, and are rarely done in community settings. Quasi-experimental designs can often involve more providers and settings, making the results more generalizable. Second, the lack of randomization often facilitates recruitment of a larger proportion of eligible patients, thus further increasing generalizability.

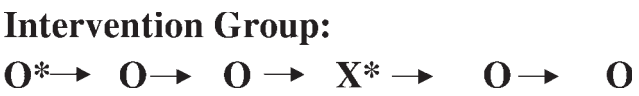

Figure 2: Time-series analysis. *O represents observation; X* represents a study intervention.

1.9. Disadvantages of the Nonrandomized Clinical Trial

The principal disadvantage of this design is the potential for bias from confounding. The direction of this bias is unpredictable from study to study. For example, clinicians may differentially include the sicker patients in the intervention trial to provide the "best chance" for the patient, thus biasing the trial against the intervention. Alternatively, the healthiest patients may be included to ensure that the intervention has the optimal opportunity to work. Therefore, the investigator should try to preempt "hand-picking" study subjects who receive the intervention. Even when optimally conducted, this design can never ensure that unmeasured or imprecisely measured social, economic, cultural, or clinical variables do not account for the apparent treatment effect. Thus the results of these trials must be evaluated in a larger context, and internal and external validity may be best assessed through the replication of results in a variety of clinical settings.

2. TIME SERIES ANALYSIS

Time series analysis can provide a more robust method for addressing the problem of secular trends in clinical care. Essentially, the investigator measures the outcome of interest several times before initiating the experiment to establish a baseline value and trend in the data (Figure 2). After the intervention, the investigator will again measure the outcome several times to establish the impact of the intervention. This design differs from a standard cohort design because the investigator manipulates patient care to estimate the effect of the intervention and from a pre/post design because it can identify trends in the outcome rate that existed before the intervention.

2.1. Identify Appropriate Questions

Time-series experiments are useful in two clinical situations: first, when the intervention produces a rapid and sustained impact on the outcome of interest. For example, a time-series analysis has been used to determine the impact of laparoscopic techniques on the rate of bile duct injuries after cholecystectomy *(7)*. Interventions that produce a delayed or gradual change may be much more difficult to capture. Using a single pre/post comparison, but including sufficiently long follow-up, a multiple time-series design can improve the robustness of the statistical comparison. The validity of a time-series analysis can be further improved by conducting a similar analysis on a comparison (control) cohort (thereby combining a time series design with a nonrandom control group design).

2.2. Define Appropriate Inclusion/Exclusion Criteria

As in all studies, inclusion and exclusion criteria must be balanced to ensure adequate comparability of the study group and generalizability of the study results. In a time-series analysis, broad inclusion criteria should generally be used to ensure that there is little room for the investigators to differentially enroll patients into a study (e.g., all consecutive patients undergoing laparoscopic cholecystectomy). Thereby, as long as the under-

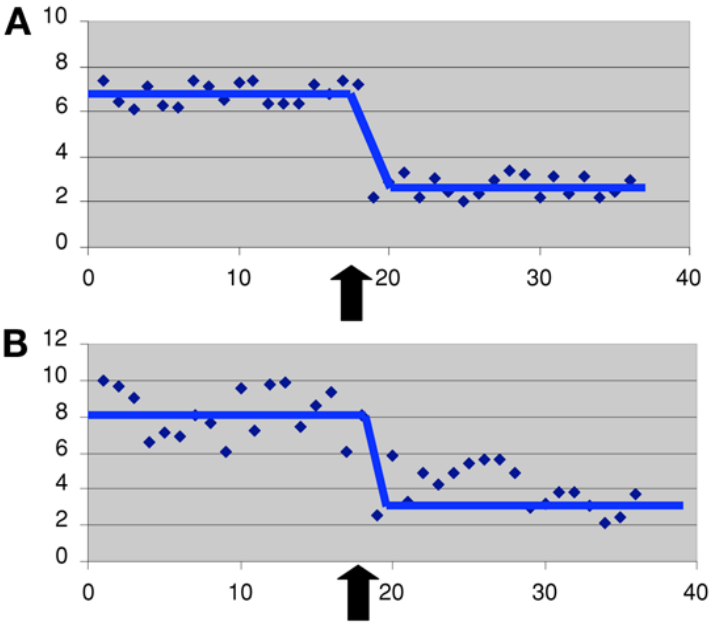

Figure 3: Impact of variation on stability in time-series trial. **(A)**. Time series with excellent precision in outcome assessment. **(B)** Time series with moderate precision in outcome assessment.

lying population does not change, patients included before and after the intervention should be similar. The exclusion criteria must include the patients' ability and willingness to remain in the study and have outcomes data collected for the length of the expected follow-up, because significant loss to follow-up is likely to bias the results of any longitudinal study. Clearly, it is also important to exclude patients from the study who would not have been candidates for procedures, including the cohort of potential study subjects preceding the introduction of the procedure. For example, there is a clear bias if outcomes from patients who are candidates for laparoscopic cholecystectomy are compared with outcomes for patients are candidates for open cholecystectomy, which may have broader eligibility criteria.

2.3. Estimate Sample Size

A vital feature of time series investigations is that there must be sufficient sample size to provide a stable estimate of the outcome incidence throughout the time series. In essence, the noise of random variation cannot be so great that it obscures the signal that you are trying to detect (the "true" outcome incidence rate). For example, in a study of risk adjusted mortality rates after cardiac surgery in the VA hospital system, investigators determined that 185 cases per 6-mo period were needed to produce a "statistical" stable (precise) estimate of surgical mortality *(8)*. Unfortunately, only one hospital during a single period achieved this case volume. Factors that influence the precision of statistical estimation are described elsewhere. To a degree, increasing the number of measurements (usually by increasing the follow-up time) and modern statistical methods for accounting for measurement error can help overcome moderate imprecision of the individual outcome rate estimates (Figure 3).

2.4. Collect Baseline Data

The number of baseline data collection points that need to be collected is in large part determined by the degree of temporal stability in the outcome rates before the intervention. As is often true in research, the exact number of data points is a balance between always wanting more data, but needing to consider incremental benefits and incremental costs of data collection.

2.5. Measure the Outcome

As in the case of the baseline data, data must be collected for a sufficient period to establish both a reliable postintervention baseline and to assess the durability of the response. It is particularly important to evaluate the potential that any observed effect merely regression to the mean.

2.6. Analysis of the Data

In a simple pre/post design, the outcome rate before and after the intervention is compared. This is also true for a true-series study, but a time-series analyses also compares the temporal changes within the pre- and postintervention periods. This is accomplished by fitting two multivariate regression models to the temporal trend in outcomes rates, one for the preintervention period and one for the postintervention period. If there is no effect of the intervention, the slopes of regression lines and their intercepts will be the same. A one-time effect will be reflected as an increase in the intercept of the regression line. Ongoing, longer term impacts will result in an acute change in the slope in the postintervention phase compared to the preintervention period (Figure 4).

2.7. Advantages of Time-Series Trial

Time-series experiments can allow the investigator to identify preexisting temporal trends in the outcome of interest and more effectively test causal influence. As shown in Figure 4, if the outcome of interest has been stable over several observation periods and then changes and persists at a new level at the time of the intervention, this provides strong inferential evidence of a treatment effect.

Time-series experiments have often been used to track quality of care and health care costs over time and assess the impact of practice changes (e.g., the impact of laparoscopic cholecystectomy on bile duct injury rates). They have also been applied to assess the impact of systemic changes on operative time, resource utilization, and throughput. They may also be particularly useful for single institution studies in which historical data are available to define precise base line values and temporal trends, thus improving on a simple pre/post intervention design without requiring a randomized control group. However, just like the pre/post design, a time-series analyses can usually benefit from adding comparison sites that are similar to the intervention sites except for the absence of the intervention. Also, pre/post and time-series studies should always include an evaluation of other changes that may have occurred at the study site at the same time as the intervention. Qualitative methods are often the preferred approach for collection this information on changes at the institution.

2.8. Disadvantages of the Time-Series Analysis

Time-series analyses are limited by the investigator's ability to completely control for potential confounders. Specifically, the population under study may change because of

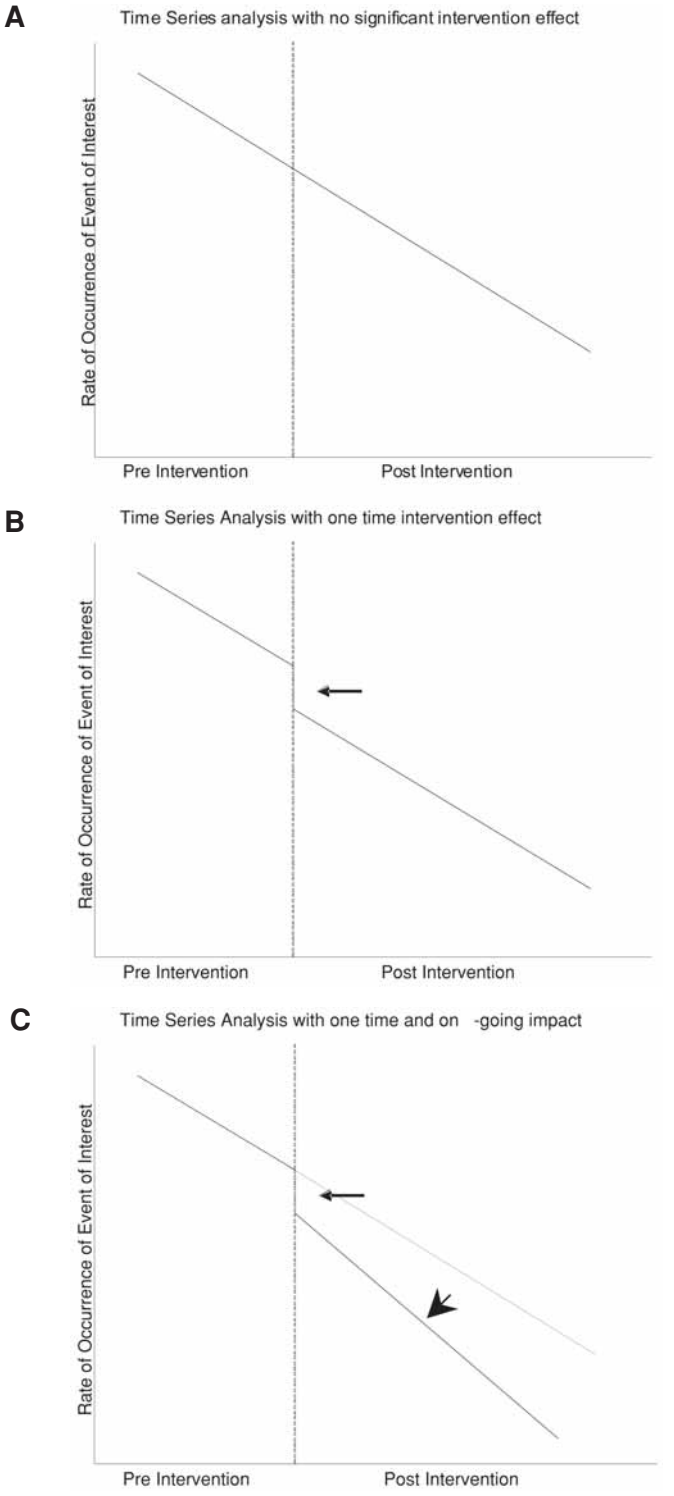

Figure 4: Time series analysis with no effect **(A)**, one time effect (**B**; small arrow, reduction in intercept of plot line), and ongoing impact (**C;** small arrow, reduction in intercept of plot line; large arrow, reduction in slope of graph).

a contemporaneous phenomenon (e.g., change in neighborhood demographics or socioeconomic status), which may bias the results of the study. This threat can be minimized through prospective data collection with established entry criteria, the use of multiple pre- and postintervention measurements, and serial qualitative evaluations of the study site.

Although increasing the number of measurements is likely to reduce the threat of an unrecognized confounder, it will usually substantially increase the cost and complexity of a time series experiment. Investigators may seek to identify outcomes that can be tracked though administrative data or other existing systems to reduce the cost of the intervention. Thus health care environments that care for a more stable group of patients over time and have comprehensive medical information system data (e.g., staff model health maintenance organizations, VA hospitals) may be excellent venues for this type of analysis.

3. COMPREHENSIVE COHORT TRIALS/PATIENT PREFERENCE TRIALS

As a result of the invasive nature of surgical interventions, patients may be reluctant to agree to random assignment. Consequently, the representativeness of surgical RCTs may be substantially compromised, thus making extrapolation of results to the general population concerning. In the comprehensive cohort study (CCS) or patient preference trial (PPT) designs, patients who decline to participate in the randomized portion of a trial continue to be followed in their chosen therapeutic arm (Figure 5) *(9,10)*. At the conclusion of the trial, comparisons are made for four groups of patients: patients randomized into intervention A, patients who selected intervention A, patients randomized into intervention B, and patients who selected intervention B. The comprehensive cohort trial differs from a traditional cohort study because only patients who are considered appropriate for either treatment arm are enrolled and all patients undergo a uniform treatment as would occur in an RCT. In addition, for the nonrandomized subjects, you are specifically examining the impact of patient preferences under circumstances that minimize the impact of physician recommendations, access and economic barriers, and other nonclinical confounders that may influence treatment decisions in usual clinical practice.

For example, the Coronary Artery Surgery Study comparing coronary artery bypass surgery with medical therapy included patients who accepted randomization (780 of 2099 patients approached) and patients who refused randomization (1315 patients) *(11)*. Patients who underwent surgery within 90 d of their evaluation were considered surgical patients, and the remaining patients were assigned to the medical management arm. These patients were followed over time and analyzed according to an intention to treat methodology as described in the following section.

3.1. Identify Appropriate Questions

Comprehensive cohort trial designs have been used to augment a randomized controlled trial when clinicians and patients are likely to have strong preexisting treatment preferences and the outcome does not rely on patient reported outcomes. As originally described, the comprehensive cohort study should usually be used to evaluate techniques that are also available outside of the study, although it has been applied to settings in which treatment is limited to within the study environment *(12)*. If the intervention is limited only to randomized patients, no meaningful comparison can be made to the group followed in the CCS.

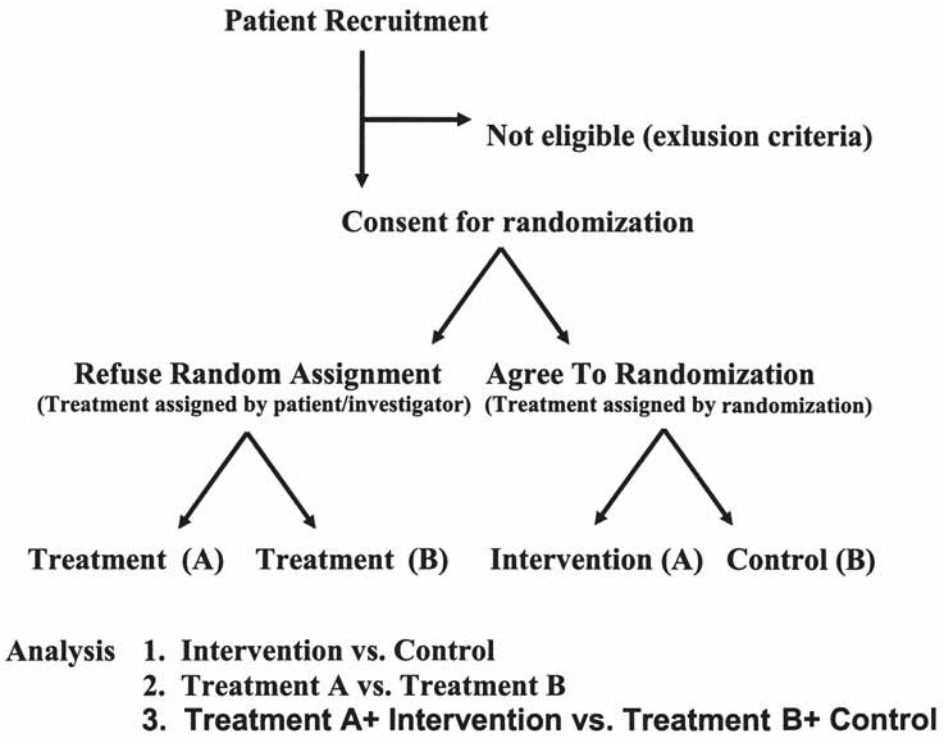

Figure 5: Analysis scheme for comprehensive cohort trial.

3.2. Define Appropriate Inclusion/Exclusion Criteria

The inclusion and exclusion criteria for a CCS or PPT should be determined by the RCT component of the study. Patients who chose interventions that are not included in either study arm should be excluded from the analysis.

3.3. Estimate Sample Size

Sample size should be determined primarily by the underlying RCT and estimates of the number of patients who will agree to randomization. As described below (Section 3.7.), the evaluation of a CCS proceeds in stages and adequate recruitment into the RCT arm is necessary to determine the principal treatment effect.

3.4. Allocate Subjects Between Groups

In a CCS, all eligible subjects are initially approached and consent for randomization is requested. Patients who refuse randomization are then asked to consent to be included in the CCS follow-up study and are allocated to treatment groups based on their preferences. In a PPT, patients are initially asked whether they have a strong preference for a specific treatment. Those patients without a strong preference are then asked to consent to randomization. Patients who refuse randomization or have a preexisting strong preference are then assigned to their preferred treatment arm. Unfortunately, treatment assignment may be difficult for patients who seek care elsewhere or who delay initiating therapy. CCS trials should be analyzed using an intention to treat methodology, and criteria for treatment assignment (e.g., surgery within 90 d of evaluation) should be specified.

3.5. Collect Baseline Data

For the most part, considerations for baseline data collection are similar to those for other quasi-experimental designs (comprehensive collection of factors that may influence risk of the outcome). However, it may be particularly relevant to collect baseline information on the patients' perspectives on the treatment options, such as the strength of their preference, their expectations or optimism regarding outcomes, and the reasons for their treatment selection.

3.6. Measure the Outcome

A predefined, objective outcome measure should be used when possible and appropriate. The confounding between patient preferences and patient reported outcomes can make analysis of subjective outcomes (e.g., pain or health-related quality of life) problematic. Therefore, physiologic or clinical outcomes are preferred (e.g., death, stroke, strength testing). Nonetheless, a "subjective" measure may still be preferable to an "objective" measure that is not very clinically or socially compelling (e.g., range of motion) especially when the nature of the comparison interventions are similar (such as two different major surgical procedures).

3.7. Analysis of the Data

It is recommended that CCS and PPT trials be analyzed sequentially. The first analysis should compare outcomes for the patients in the randomized portion of the trial. Next, patients in the nonrandomized arm should be examined to determine if the treatment effect is consistent or inconsistent in the preference allocation cohort. Finally, all patients may be considered in a single multivariate regression analysis, including an indicator variable for randomization status as a covariate. Using this technique, the independent effects of treatment and patient preference and the interaction between patient preference and treatment choice can be determined *(13)*.

3.8. Advantages of CCS/PPT Design

CCS/PPT designs offer two principal advantages. First, the ability to choose treatment assignment may improve recruitment into a clinical trial and thus increase the sample size. Unfortunately, the availability of a CCS/PPT may limit patients' desire to enter the RCT portion of the trial and will, in the end, increase the duration and costs of the enrollment process necessary to obtain a sufficient number of randomized patients. Second, CCS/PPT trials may enhance the external validity of the study's main findings. Because patients who decline randomization may represent the majority of patients seen in clinical practice, a consistent finding in both the randomized and nonrandomized cohorts can provide some reassurance regarding generalizability. Furthermore, if the results between the randomized and nonrandomized cohorts are inconsistent, then one can describe the direction and magnitude of the bias introduced through the self-determination of treatment (or physician selection of treatment).

3.9. Disadvantages of the CCS/PPT Analysis

The addition of a CCS/PPT study to an RCT is likely to increase the cost and complexity of the trial. A CCS/PPT study will be larger and the follow-up may be more difficult if patients seek treatment outside of the study centers, but are still included in the cohort. Furthermore, the threat of residual unmeasured confounding is an inherent threat to the

validity of the CCS/PPT even with state-of-the-art measures of baseline risk factors. Thus, although the CCS/PPT patients may be more representative of the general population, careful attention must be paid when adjusting for differences in baseline characteristics.

4. EMPIRICAL EVIDENCE COMPARING RANDOMIZED AND NONRANDOMIZED TRIALS

Although RCTs continue to be viewed as the gold standard for clinical research, a series of comprehensive evaluations has failed to demonstrate consistent differences in treatment effects found in high-quality randomized and nonrandomized investigations *(14–16)*. MacLehose and colleagues reviewed 14 articles involving 38 interventions in which the results from quasi-experimental studies were compared with those derived from RCTs. The authors concluded that there were no significant differences in the effect size or direction between high-quality quasi-experimental studies and RCT. However, in low-quality studies, the effect size appeared to be more extreme, but the direction varied in only one comparison. The low-quality studies were principally review articles that did not use appropriate meta-analytic techniques.

The difference in outcomes reported between randomized and nonrandomized trials may, sometimes, be just a reflection of the underlying patient characteristics. In a comprehensive review of RCT and non-randomized studies comparing coronary artery bypass grafting with angioplasty, Britton and colleagues determined that coronary artery bypass grafting was favored in the RCT and angioplasty in the nonrandomized cohort studies. However, after adjustment for patient characteristics, the differences were no longer statistically different *(15)*. Benson and Hartz reached a similar conclusion after examining 136 articles in 19 treatment areas. The effect estimates derived from RCT and nonrandomized trials were similar for 17 of the 19 conditions studied and there did not appear to be systematic bias in observational investigations *(15)*.

The impact of publication bias may differentially impact randomized and nonrandomized trials. Although adequately powered randomized trials that fail to demonstrate a significant difference between treatment arms are routinely published, it appears that nonrandomized studies demonstrating a similar conclusion are more often rejected and probably less often submitted for consideration by the authors as well. Thus, when RCT and nonrandomized trials comparing a single treatment are examined, the nonrandomized trials are more likely to demonstrate positive results. Perhaps, this publication inequality accounts for part of the perception that quasi-experimental and observational studies are intrinsically biased.

5. CONCLUSIONS

Quasi-experimental study designs offer surgical investigators a valuable tool to overcome many of the impediments to conducting a randomized clinical trial. The use of properly selected nonrandomized control groups can help to overcome threats to internal validity from temporal trends, surgical learning curve effects, regression to the mean, and the difficulty in obtaining equipoise among surgeons. Likewise, time-series analysis can be well suited for situations in which clinical practice outpaces research evaluation. Often combining these two methods (i.e., using both pre/post and contemporary comparisons) will be the optimal approach, making it possible to examine the impact of rapid clinical change in diverse patient populations and clinical settings.

The trade-off between RCT and a quasi-experimental study design is largely pragmatic. When feasible, an RCT is almost always preferred because it minimizes the risk that unmeasured confounding is biasing the studies conclusions. However, often a quasi-experimental study design can offer important insights into the care of surgical patients and can lead to more generalizable study results based on more representative patient populations. Surgical case-series reports continue to be very common; however, many of these case series could be readily redesigned to create rigorous and more scientifically-sound quasi-experiments. Quasi-experimental designs warrant careful consideration by surgical researchers and should be more widely used.

REFERENCES

1. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomized trials in surgery: problems and possible solutions. BMJ 2002;324:1448–1451.
2. Macklin R. The ethical problems with sham surgery in clinical research. New Engl J Med 1999;341:992–996.
3. Curry JI, Reeves B, Stringer MD. Randomized controlled trials in pediatric surgery: could we do better? J Pediatric Surg 2003;38(4):556–559.
4. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995:48:1503–1510.
5. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;347:81–88.
6. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomized and non-randomized clinical trials. BMJ 1998;317:1185–1190.
7. Flum DR, Dellinger EP, Cheadle A, Chan L, Koepsell T. Intraoperative cholangiography and risk of common bile duct injury during cholecystectomy. JAMA 2003;289:1639–1644.
8. Marschall G, Shroyer AL, Grover FL, Hammermeister KE. Time series monitors of outcomes: a new dimension for measuring quality of care. Med Care 1998;36:348–356.
9. Olschewski M, Scheurlen H. Comprehensive cohort study: an alternative to randomized consent design in a breast preservation trial. Methods Inf Med 1985;24:131–134.
10. Torgerson DJ, Sibbald B. Understanding controlled trials: what is a patient preference trial? BMJ 1998;316:360–361.
11. CASS Principal Investigators and their Associates. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Comparability of entry characteristics and survival in randomized and non-randomized patients meeting randomization criteria. J Am Coll Cardiol 1984;3:114–128.
12. Olschewski M, Schumacher M, Davis KB. Analysis of randomized and non-randomized patients in clinical trials using the comprehensive cohort follow-up study design. Controlled Clin Trials 1992;13: 226–239.
13. Olschewski M, Scheurlen H. Comprehensive cohort study: an alternative to randomized consent design in a breast preservation trial. Methods Inf Med 1985;24:131–134.
14. MacLehose RR, Reeves BC, Harvey IM, et al. A systematic review of comparisons of effect sizes derived from randomized and non-randomized studies. Health Technol Assess 2000;4:34.
15. Britton A, McKee M, Black N, et al. Choosing between randomized and non-randomized studies: a systematic review. Health Technol Assessment 1998;2:13.
16. Benson K, Hartz AJ. A Comparison of observational studies and randomized controlled trials. N Engl J Med 2000;342:1878–1886.