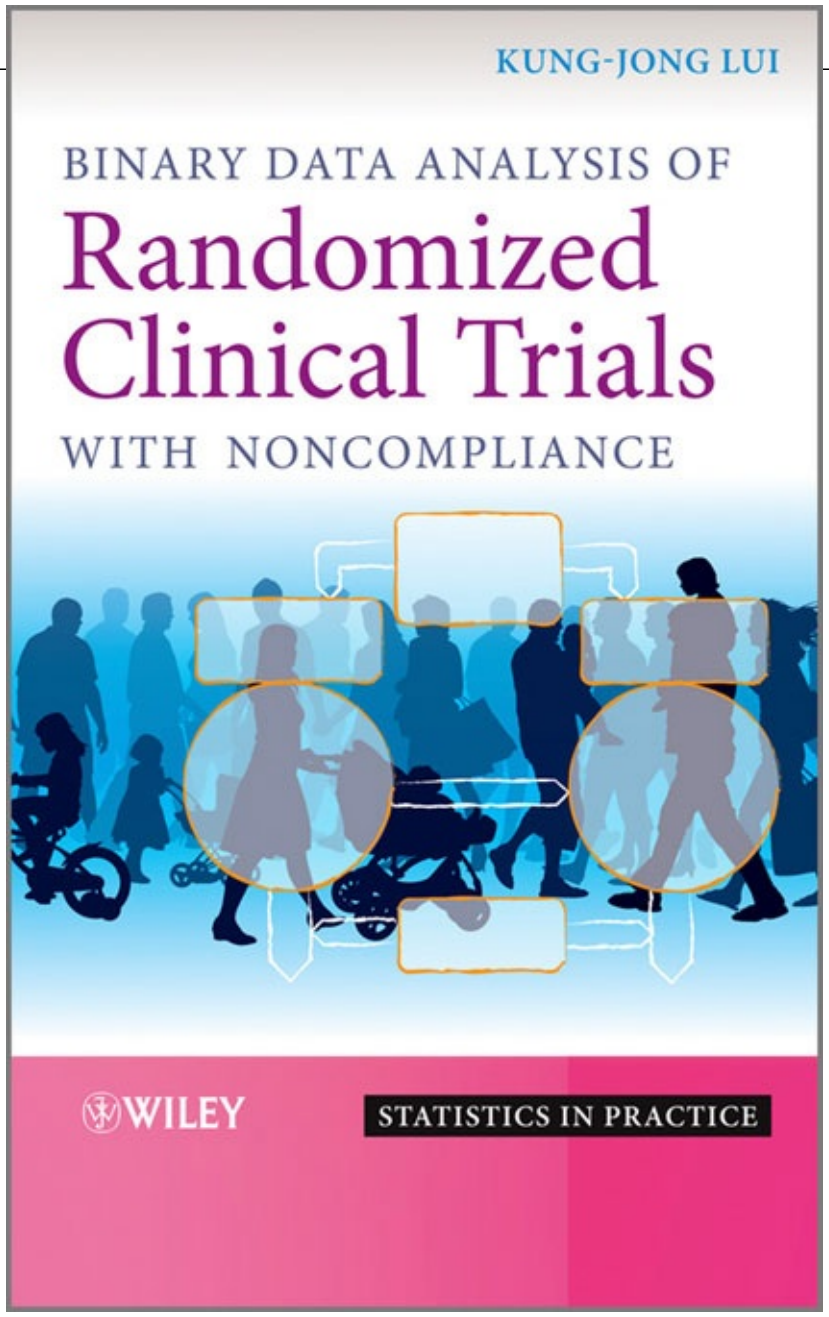
KUNG-JONG LUI

BINARY DATA ANALYSIS OF Randomized Clinical Trials WITH NONCOMPLIANCE

WILEY

STATISTICS IN PRACTICE


KUNG-JONG LUI
BINARY DATA ANALYSIS OF
Randomized
Clinical Trials
WITH NONCOMPLIANCE
WILEY
STATISTICS IN PRACTICE

Contents

Statistics in Practice

Series Advisors

Statistics in Practice is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceutics; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges. It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

Binary Data Analysis of Randomized Clinical Trials with Noncompliance

Kung-Jong Lui

Department of Mathematics and Statistics
San Diego State University, USA

WILEY
A John Wiley and Sons, Ltd., Publication

This edition first published 2011

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Lui, Kung-Jong, author.
Binary data analysis of randomized clinical trials with noncompliance/Kung-Jong Lui, Department of Mathematics and Statistics, San Diego State University, USA.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-66095-9 (cloth) – ISBN 978-1-119-99160-1 (epdf) – ISBN 978-1-119-99161-8 (obook) – ISBN 978-1-119-99390-2 (epub) – ISBN 978-1-119-99391-9 (mobi)
1. Clinical trials–Statistical methods. 2. Drugs–Testing–Statistical methods. I. Title.
[DNLM: 1. Randomized Controlled Trials as Topic–methods. 2. Medication Adherence. 3. Statistics as Topic–methods. QV 771]
RM301.27.L85 2011
615.5072$′$4–dc22
2010051072

A catalogue record for this book is available from the British Library.

Print ISBN: 978-0-470-66095-9
ePDF ISBN: 978-1-119-99160-1
oBook ISBN: 978-1-119-99161-8

ePub ISBN: 978-1-119-99390-2
Mobi ISBN: 978-1-119-99391-9

Dedicated to

Professors William G. Cumberland and Abdelmonem A. Afifi at UCLA, as well as Professor Daniel McGee at Florida State University

Preface

In a randomized clinical trial (RCT), it is quite common to encounter patients who do not comply with their assigned treatment due to ethical reasons, patient's decision or the feature of a study design (such as pre-randomized consent designs). Since noncompliance often occurs non-randomly, the commonly-used subgroup analyses, including as-treated (AT) analysis and as-protocol (AP) analysis, are well known to produce a possibly biased inference of treatment efficacy due to the incomparability of the underlying prognostic conditions for patients between two comparison groups. To alleviate this concern, the intent-to-treat (ITT) (or as-randomized (AR)) analysis has been often suggested for a RCT with noncompliance. However, the ITT analysis estimates the programmatic effectiveness rather than the treatment efficacy. Although ITT analysis may provide us with unbiased test for assessing the superiority of an experimental treatment to a standard treatment, the ITT analysis tends to underestimate the relative treatment effect in the presence of noncompliance under certain commonly-assumed conditions. Thus, how to assess the treatment efficacy in a RCT with noncompliance becomes practically useful and important.

The analysis of data for a RCT with noncompliance is generally quite complicated even for the simplest case of a simple noncompliance RCT, in which only patients assigned to the experimental treatment can have access to the experimental treatment. Furthermore, the frequent involvement of sophisticated numerical iterative procedures based on likelihoods to obtain parameter estimates makes this topic even more challenging and difficult for many clinicians and data analysts to appreciate. This book is to focus attention on the level which clinicians with one year of solid training in biostatistics can comprehend, and provides readers with a simple, systematic, and organized approach to study treatment effect for a RCT with noncompliance when the patient response is dichotomous and the noncompliance status is all-or-none in a variety of situations. This book adopts an instructive and easily-understood approach by using contingency tables to explicitly lay down the latent probability structure of observed data so that readers can easily visualize the logics and the ideas behind the development of the proposed test procedures and estimators in a one unified model frame. By contrast, when using the proportion difference (PD) to measure the relative treatment effect, we assume the structural risk additive model based on the model-based approach. While using the proportion ratio (PR) to measure the relative treatment effect, we assume the structural risk multiplicative model. Furthermore, this book presents all test procedures, estimators and sample size calculation procedures in closed forms. Readers may simply use a hand calculator to calculate all the test statistics, interval estimators or sample size calculation formulae without the need of employing any iterative numerical procedures in the situations considered here. For the easy access of the particular topic of reader's interest, this book is written in such a constructive structure that the underlying assumptions, notation, test procedures and formulae in each chapter are self-contained. Readers may directly refer to the particular chapter without the need of reading the details in all the preceding chapters, although I must admit that some assumptions, definitions in notation, and important notes are repeated to avoid confusions in narrative or ambiguities in formulae and findings. Through some real-life examples and computer-simulated data, readers can appreciate the practical usefulness of the test procedures and estimators discussed in this book. The exercises given at the end of each

chapter can further help readers better understand the underlying assumptions, the theory and limitations of the proposed test procedures and estimators, as well as other relevant issues and extensions. To facilitate the use of sample size determination presented here, we include in Appendix SAS programs that can be easily modified by readers to accommodate the situations in which they are interested. Despite the book generally adopting the principal stratification approach to account for the effect due to noncompliance, this book also briefly addresses use of the model-based approach (which is related to a quite general class of the structural mean models (SMMs) proposed elsewhere) and notes the relations of parameters and estimators between these two approaches. Because the discussion on the SMM is truly beyond the modest scope of this book, the SMM is not discussed in the book. Readers who are interested in this area may begin with reading a few key references regarding the SMMs cited here.

This book is intended for postgraduates, clinicians, biostatisticians and data analysts. This book can be used as supplemental material for an introductory-level course focusing on clinical statistics or experimental trials in Epidemiology, Psychology and Sociology. This book may also be used as a desk reference for clinicians or biostatisticians when they come across binary data in the presence of noncompliance. To clarify the main issues raised by noncompliance and strengthen the narrative, we explicitly define and discuss some common assumptions and terms encountered in a RCT with noncompliance, as well as include numerical examples to illustrate the bias of most commonly-used subgroup analyses in Chapter 1. Because testing superiority, non-inferiority and equivalence, interval estimation and sample size calculation are all the most fundamental statistical topics for analyzing clinical data, this book concentrates discussions on these when we use the PD, the PR and the odds ratio (OR) to measure treatment efficacy under various frequently-encountered situations. These include parallel groups design (Chapter 2), stratified sampling (Chapter 3), cluster sampling (Chapter 4), parallel sampling with subsequent missing outcomes (Chapter 5) and data in repeated binary measurements (Chapter 6). Clinicians and biostatisticians should find that this book is useful and handy.

I wish to express my indebtedness to my colleagues Drs. Richard Levine, Barbara Bailey and Kristin Duncan at San Diego State University and the five anonymous reviewers who generously provided valuable comments and suggestions on an early draft and outlines of contents of the manuscript. I also wish to thank my wife Jen-Mei, whose continued patience and understanding have endured throughout so many years and made the work much more pleasant than it otherwise would have been. I want to thank my brothers Dan-Yang, Kung-Yi and Kung-Jen for their encouragements in many years. Finally I want to express my deepest appreciation to my parents, Shung-Wu and Li-Ching for their endless love, spiritual support and guidance, which continue to last in my memory.

Kung-Jong Lui
San Diego

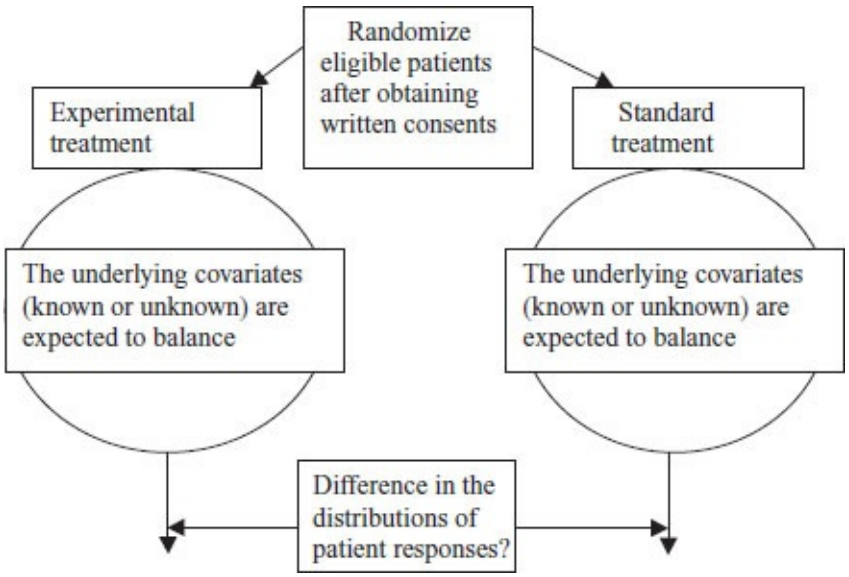
About the Author

Kung-Jong Lui is a professor in the Department of Mathematics and Statistics at San Diego State University. He obtained his Ph.D. in biostatistics in 1982, M.S. in biostatistics in 1979, M.A. in Mathematics in 1977, all from UCLA, and B.S. in Mathematics in 1975 at Fu-Jen Catholic University at Taipei, Taiwan. He has had 150 publications in peer-reviewed journals, including *Biometrics*, *Statistics in Medicine*, *Biometrical Journal*, *Computational Statistics and Data Analysis*, *Psychometrika*, *Journal of Biopharmaceutical Statistics*, *Drug Information Journal*, *Contemporary Clinical Trials*, *Journal of Applied Statistics*, *Statistical Methodology*, *Communications in Statistics, Theory and Methods*, *Science*, *Nature*, *Proceedings of National Academy of Sciences*, *Journal of Official Statistics*, *IEEE Transactions on Reliability*, *Environmetrics*, *Test*, *American Journal of Epidemiology*, *American Journal of Public Health*, *New England Journal of Medicine*, *Journal of the American Medical Association*, etc. He is the author of the book *Statistical Estimation of Epidemiological Risk* published by Wiley in 2004. He is an Associate Editor for *Biometrical Journal*. He is a Fellow of the American Statistical Association, a Fellow of the American College of Epidemiology, and a life member of International Chinese Statistical Association.

Randomized clinical trials with noncompliance: issues, definitions and problems of commonly used analyses

When comparing an experimental treatment with a standard treatment (or placebo), we often employ a randomized clinical trial (RCT), in which eligible patients (after obtaining their informed consents) are randomized to one of the two treatments under comparison. One of the most fundamental ideas behind use of the RCT is, as shown in Figure 1.1, that all (known or unknown) covariates affecting patients' responses are expected to balance through randomization. Thus, when there is a difference in the distribution of patient responses between two treatments under perfect compliance, we may attribute this to different treatments they receive between the two randomized groups.

Figure 1.1 Schema for a RCT.



However, noncompliance can often occur in a RCT. When a patient feels that the burden of taking his/her assigned treatment is not worth its perceived benefits, the patient may decide not to comply with his/her assigned treatment (Heitjan, 1999). Noncompliance can also occur as a result of a negative experience of taking a treatment, drug sharing among participated patients, an error in treatment administration by study staff, or even the feature of a pre-randomized study design (Zelen, 1979, 1982, 1986, 1990). Because noncompliance often occurs nonrandomly, simply excluding patients who do not comply with their assigned treatment from data analysis may produce a misleading inference. For convenience in the following discussion, we call the RCT with noncompliance, in which only patients assigned to an experimental treatment group can have access to the experimental treatment, the simple noncompliance RCT.

Example 1.1 Consider the simple noncompliance RCT, in which children who resided in 225 villages randomly selected out of 450 villages were assigned to the intervention group of receiving two large oral doses of vitamin A supplementation, while children who resided in the

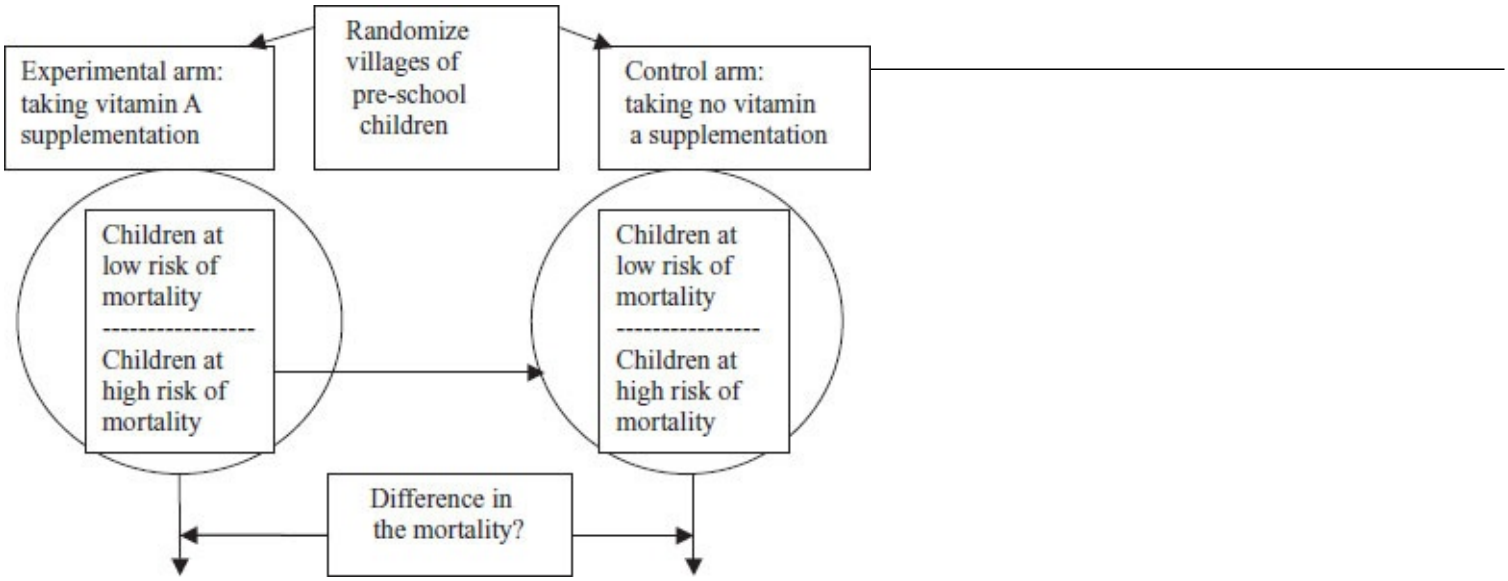
remaining 225 villages were assigned to the control group of receiving no vitamin A supplementation (Sommer and Zeger, 1991; Sommer, Tarwotjo and Djunaedi *et al*., 1986). Approximately 20 % of children assigned to the intervention group did not receive vitamin A supplementation, but children assigned to the control group were all assumed to receive no vitamin A supplementation. To investigate whether there was a self-selection bias due to noncompliance, we might compare the mortality rate of children who were assigned to the intervention group but received no vitamin A supplementation with the mortality rate of children who were assigned to the control group. We summarize these data in Table 1.1 (Sommer and Zeger, 1991). For the purpose of illustration, we ignore the intraclass correlation (which will be discussed in Chapter 4) of survival outcomes between children within villages here. When employing the commonly used two independent sample-proportion statistics to test the equality of mortality rates (Fleiss, 1981), we find strong statistical significance (p-value < 0.001) based on the data in Table 1.1. In other words, there is strong evidence that children who declined receiving vitamin A supplementation tended to be in poorer health or at a higher risk of mortality. To help readers easily appreciate the schema of this simple noncompliance RCT, we may use the schema as shown in Figure 1.2.

Table 1.1 The observed cell frequency and the corresponding cell proportion (in parenthesis) in preschool children who were assigned to the intervention group but did not receive the vitamin A supplementation versus those in preschool children who were assigned to the control group and assumed to all receive no vitamin A supplementation.

	Patients assigned to the intervention group but received no vitamin A supplementation	**Patients assigned to the control group**
Death	34 (1.4 %)	74 (0.6 %)
Survival	2385 (98.6 %)	11514 (99.4 %)
Total	2419 (100 %)	11588 (100 %)

Because noncompliance does not, as shown in Figure 1.2, occur randomly, we cannot directly compare the mortality in preschool children between the two randomized arms by simply excluding those children who did not comply with taking vitamin A supplementation from the experimental arm. This is because the underlying prognostic conditions on children between the two arms would not balance; the experimental arm would consist of children at the low risk of mortality and the control arm would consist of children at both low and high risks of mortality. Thus, if we included only children who complied with their assigned treatment in our analysis and found a reduction in the mortality rate of the experimental arm as compared with the control arm, this could be due to the reason that the children at a high risk of mortality were excluded from the experimental arm.

Figure 1.2 Schema for the simple noncompliance RCT of studying vitamin A supplementation to reduce mortality among pre-school children.



1.1 Randomized encouragement design (RED)

For certain treatments, such as flu vaccine or quitting smoking, it is not ethical to randomly assign high-risk patients to receive either the treatment or the placebo. Thus, to alleviate the ethical concern in application of the traditional RCT, the randomized encouragement design (RED) is often suggested (Multiple Risk Factor Intervention Trial Research Group, 1982; McDonald, Hui and Tierney, 1992; Zhou and Li, 2006; Jo, 2002).

> **RED** – Patients are randomly assigned to either the intervention group of receiving an encouragement to accept the experimental treatment or the control group of receiving no such encouragement.

Because of randomization, the underlying prognostic conditions between the intervention and control groups are expected to balance in a RED. Since we do not interfere with patients assigned to the control group to receive their usual medical treatment in a RED, there are no ethical issues involved. Note that the rate of taking the experimental treatment in the intervention group of receiving an encouragement is expected to be higher than that in the control group through the encouragement. Thus, when there is a difference in the proportion of patient responses between the two randomized groups, we may attribute this to the difference in the two treatments under comparison. Note that because patients may decide to take or decline the experimental treatment despite whether they receive an encouragement or not, the extent of noncompliance is generally large in a RED. Thus, the RED can be relatively inefficient to the traditional RCT for detecting a difference between two treatments, especially when the extent of noncompliance is not small. How to achieve a high compliance rate becomes a very important and critical issue in designing a good RED.

Example 1.2 If we employed the traditional RCT to randomly assign high-risk patients to receive either the flu vaccine or the placebo, we would withhold vaccination from some high-risk patients. This would raise the ethical concern. The RED has been employed to study the influenza vaccine efficacy in reducing morbidity by using computer-generated reminder for flu shots (McDonald, Hui and Tierney, 1992). Physicians were randomly assigned to either the intervention group of receiving a computer-generated reminder when a patient with a scheduled appointment was eligible for a flu shot or the control group of receiving no such reminders. Each

physician at the clinic cared for a fixed group of patients and his/her patients were then similarly classified. Since the study did not keep information on the clustering of patients by doctor, we ignore clustering for the purpose of illustration (Zhou and Li, 2006). We summarize these data in Table 5.1 (Chapter 5). Approximately 79 % of patients who were assigned to the intervention group of receiving reminders did not receive the flu vaccine, while approximately 14 % of patients who were assigned to the control group of receiving no reminders received the flu vaccine. There were also many patients with subsequent missing outcomes. How to obtain a consistent estimator of the flu vaccine effect on morbidity in the presence of a large percentage of noncompliance and a nonnegligible percentage of missing outcomes is likely to be of practical interest. We will discuss hypothesis testing and estimation of the treatment effect for a RCT with both noncompliance and subsequent missing outcomes in Chapter 5.

1.2 Randomized consent designs

When we implement a traditional RCT, the assignment of patients to a treatment completely depends on a chance mechanism after obtaining patients' informed consent. At the time of consent, neither physicians nor patients know exactly which treatment a patient will receive. This may compromise the relationship between physicians and patients (Zelen, 1990). Since physicians need to provide patients with all the relevant information on treatments, including the fact that they are not even sure which treatment can be the best to the patient, physicians may feel hesitated to enroll patients into a traditional RCT. This can cause the practical difficulty in recruiting patients into a RCT (Zelen, 1990). Furthermore, patients may originally agree to participate in a traditional RCT, but have reservation about continuing to participate or even decline the treatment once when the treatment is known. To account for these concerns, Zelen (1979, 1990) proposed the randomized consent design (or pre-randomized design), in which patients are randomly assigned to the treatments even before their consents are sought. After assigning an eligible patient to a treatment, physicians approach patients for consents and discuss potential risks, benefits, and treatment opinions. Patients will be given the assigned treatment if he/she is willing to accept the assigned one, and otherwise, the other. One important advantage of the randomized consent design over the traditional RCT is that the patient, at the time of consent, knows exactly which treatment he/she is going to receive (Zelen, 1990). By contrast, patients do not generally know exactly which treatment will be received in a traditional RCT. Based on whether noncompliance can occur in only one or both of the two randomized groups, the randomized consent designs can be classified as a single-consent randomized design (SCRD) and a double-consent randomized design (DCRD).

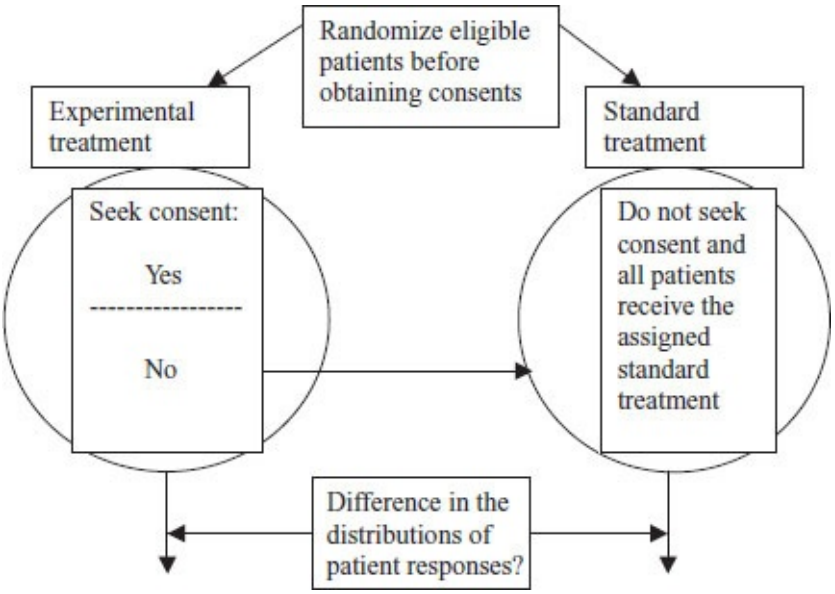
1.2.1 Single-consent randomized design (SCRD)

> **SCRD** – Patients assigned to an experimental treatment are asked for consents, while patients assigned to a standard treatment are not.

If a patient in the assigned experimental treatment group agrees to receive the assigned treatment, he/she will be given the experimental treatment and otherwise, the patient will be given the standard treatment. However, all patients in the assigned standard treatment group are assumed to all receive the standard treatment. We may use the diagram in Figure 1.3 to illustrate

the schema of the SCRD.

Figure 1.3 Schema for the single-consent randomized design.



When comparing an experimental treatment with the best available standard treatment, Zelen (1979, 1990) contended that the SCRD could be a useful alternative design to the traditional RCT. This is because patients assigned to the standard treatment receive the best available treatment to them and hence it should not involve ethical issues if we did not seek their consents. On the other hand, patients assigned to the experimental treatment could be allowed to switch the best standard treatment if they were not willing to accept the assigned (experimental) treatment. Thus, patients and physicians know exactly which treatment the patient will be given. Zelen (1990) provided an excellent discussion on when the randomized consent design can be more efficient than the traditional RCT through an increase of the enrollment rate of patients into a trial. Anbar (1983), Matts and McHugh (1993) as well as Brunner and Neumann (1985) also discussed estimation and testing hypothesis under the randomized consent design.

Example 1.3 The SCRD has been employed to study the extracorporeal membrane oxygenation (ECMO) on new infants having a diagnosis of persistent pulmonary hypertension (PPH) (Zelen, 1990). When infants were diagnosed with PPH, using the traditional RCT would require that the parents of an infant near death provide informed consent for an invasive surgical procedure (ECMO) which might not be even administered to their babies. This can raise an unnecessarily stressful burden to both parents and health administrators. The SCRD only required that parents whose infants were assigned ECMO be approached for giving consents because this was a deviation from the conventional therapy. Other practical applications of the randomized consent design can be found elsewhere (Zelen, 1990).

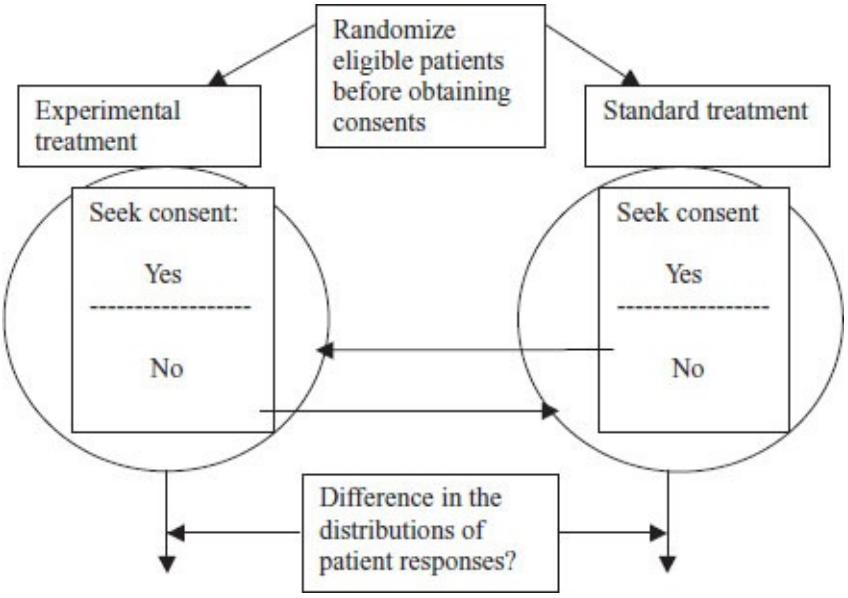
1.2.2 Double-consent randomized design (DCRD)

DCRD – Patients are randomly assigned to either an experimental treatment or a standard treatment. Patients are then approached for consents in both groups.

If a patient assigned to the experimental treatment group does not agree to accept the assigned treatment, he/she will be given the standard treatment. Similarly, if a patient assigned to the standard treatment group does not agree to accept the assigned treatment, he/she will be given the

experimental treatment. We may use the diagram shown in Figure 1.4 in to illustrate the schema of the DCRD.

Figure 1.4 Schema for double consent randomized design.



The SCRD is a special case of DCRD when only patients assigned to the experimental treatment are asked for consents. Because noncompliers in the assigned experimental treatment group do not necessarily represent the same subpopulation as noncompliers in the assigned standard treatment group, a direct comparison of patient responses by excluding those noncompliers from data analysis can be misleading due to the underlying prognostic conditions are no longer comparable between the two comparison groups. Some discussions on hypothesis testing and interval estimation under the DCRD appeared elsewhere (Anbar, 1983; Brunner and Neumann, 1985; Lui and Lin, 2003).

1.3 Treatment efficacy versus programmatic effectiveness

Before discussing the bias of an estimator for treatment efficacy, it is essential to clarify the definition of treatment efficacy in the presence of noncompliance to avoid the confusion noted elsewhere (Lui, 2009). Following Last (1988), we define the treatment efficacy as the treatment effect relative to a control (or placebo) among compliers who would fully comply with whatever their assigned treatment regimen. The treatment efficacy provides us with the useful information on the biological action of a treatment and is most interesting to clinicians. By contrast, we define the programmatic effectiveness as the treatment assignment effect relative to a control (or placebo) in a population consisting of both compliers and noncompliers. On the basis of the programmatic effectiveness, a drug with low treatment efficacy but an extremely high compliance rate can be more useful than a drug with high treatment efficacy but an extremely low compliance rate (Nagelkerke, Fidler and Bernsen *et al*., 2000). This is because a drug can be beneficial to only those patients who would accept the drug, and using the former with a very high compliance rate may save more patients than the latter with an extremely low compliance rate in practice. Thus, the programmatic effectiveness can be of interest and importance to health policy administrators. When subjects in a population are all compliers, the treatment efficacy and the programmatic

effectiveness are, by definition, identical. Note that the programmatic effectiveness can vary as compliance changes. A meta-analysis of empirical research showed an overall 26 % difference in response rates between patients with high and low compliance rates (Walter, Guyatt and Montori *et al.*, 2006; DiMatteo, Giordani and Lepper *et al.*, 2002). Unless the distribution of noncompliance for patients participated into a RCT is quite similar to that for patients of the targeted population, we may not be able to extrapolate the findings on the programmatic effectiveness from a particular RCT to the targeted population. Thus, we will focus our attention on the treatment efficacy, and use the terms treatment efficacy and treatment effect synonymously in this book.

1.4 Definitions of commonly used terms and assumptions

To estimate the efficacy of a treatment in the presence of noncompliance, we first make the stable unit treatment value assumption (SUTVA) (Rubin, 1978).

> **SUTVA** – There is no interference between patients; the treatment-received status of one patient does not influence the response or the treatment-received status of another patient (Sato, 2001; Matsuyama, 2002). Also, we often include consistency – the responses of patients remain identical regardless of possibly different forms or versions in administration of a treatment, as a part of the SUTVA as well (Ten Have, Elliott and Joffe *et al.*, 2004; Bellamy, Lin and Ten Have, 2007).

The SUTVA can sometimes be violated in practice. For example, consider the study of quitting smoking on the mortality of coronary heart disease (CHD) (Matsui, 2005; Multiple Risk Factor Intervention Trial Research Group, 1982). Smoking can have a direct effect on the outcome of a patient who smoked and an indirect effect on the outcome of a patient whose roommate smoked. An analysis without accounting for this indirect effect can lead us to a biased estimate of the effect due to smoking if both of these patients are included into the trial. Other examples about the violation of SUTVA can be found elsewhere (Cox, 1958; Bellamy, Lin and Ten Have, 2007). Also, for simplicity, we focus our attention on the situation in which there is only a single one form or version of administrating treatments and hence consistency is implicitly assumed to be satisfied throughout this book.

Say, we compare an experimental treatment with a standard treatment. Following Angrist, Imbens and Rubin (1996), we define for each patient the vector $(d(1), d(0))$ of his/her potential treatment-received status: $d(g) = 1$ if the patient assigned to treatment g (= 1 for experimental, and = 0 for standard) actually receives the experimental treatment, and $d(g) = 0$ otherwise. Therefore, we can divide our sampling population into four subpopulations, including compliers ($d(1) = 1$ and $d(0) = 0$), never-takers ($d(1) = d(0) = 0$), always-takers ($d(1) = d(0) = 1$), and defiers ($d(1) = 0$ and $d(0) = 1$). To allow the parameter representing the treatment efficacy to be identifiable, we commonly make the monotonicity assumption as well as the exclusion restriction assumption for always-takers and never-takers.

> **The Monotonicity Assumption** – We assume $d(1) \geq d(0)$ for all patients; or equivalently, there are no defiers.
>
> **The Exclusion Restriction Assumption** – The treatment affects a patient response only through the treatment which the patient actually receives and the treatment assignment itself does not affect the patient response.

Brunner and Neumann (1985) contended that a patient who refused a proffered treatment should only stay in the trial if he/she preferred the other treatment. In fact, it is very likely that a patient who is a defier may not even provide his/her written consent and enter into a RCT in practice. Thus, the monotonicity assumption should be plausible in most encountered RCTs, although one may find situations in which the assumption of no defiers does not hold (Bellamy, Lin and Ten Have, 2007; Ten Have, Elliott and Joffe *et al*., 2004). Based on the monotonicity assumption, if a patient assigned to an experimental treatment ($g = 1$) receives a standard treatment ($g = 0$), he/she must be a never-taker (i.e. a patient with $d(1) = 0$ must have $d(0) = 0$). Similarly, if a patient assigned to a standard treatment receives an experimental treatment, he/she must be an always-taker (i.e. a patient with $d(0) = 1$ must have $d(1) = 1$). However, if a patient assigned to an experimental treatment receives his/her assigned (experimental) treatment, he/she can be either a complier or an always-taker. Also, if a patient assigned to a standard treatment receives his/her assigned (standard) treatment, he/she can be either a complier or a never-taker. Because we cannot distinguish compliers from always-takers in the assigned experimental treatment or compliers from never-takers in the assigned standard treatment, the difference in the probabilities of response among compliers between the experimental and standard treatments is not directly estimatible from data without making some assumptions. This is actually the fundamental issue in estimation of treatment effect (among compliers) under a RCT with noncompliance.

The exclusion restriction assumption is likely to be reasonable in a double-blind study. Frangakis and Baker (2001) contended that the exclusion restriction assumption for always-takers and never-takers is probably to hold when noncompliance occurs soon after assignment. This is because, within the defined groups of always-takers and never-takers, different assignment results in the same extent of actual exposure to the experimental treatment (for always-takers) and the standard treatment (for never-takers). On the other hand, for example, in the RED studying the flu vaccine, the exclusion restriction assumption for always-takers might not necessarily hold (Hirano, Imbens and Rubin *et al*., 2000; Zhou and Li, 2006). This is because always-takers who received the flu shot regardless of their assigned group tended to be patients who were most likely at high risk for getting flu. If the flu reminder prompted the physician to take other medical treatments beyond the flu shot to improve the health outcomes on such patients, the exclusion restriction assumption for always-takers could be violated. Note that the exclusion restriction assumption is generally not testable without having the additional auxiliary information. Hirano, Imbens and Rubin *et al*. (2000) proposed a Bayesian approach and discussed sensitivity analysis to violation of the exclusion restriction assumption. Their results can depend, however, on their assumed specific form of the likelihood function and prior distribution. Note also that there is a subtle difference in the definition of compliers between the traditional RCT and RED. A complier in the former represents a patient who receives whatever treatment he/she is assigned to, while a complier in the latter represents a patient who will accept a treatment if he/she is assigned to an intervention group of receiving an encouragement and who will decline a treatment if he/she is assigned to a control of group of receiving no encouragement. Thus, the treatment efficacy defined in compliers between the traditional RCT and RED can be nonidentical. The compliers in a RED can be trial specific and hence the extrapolation of findings from a RED to the targeted population should also be treated with caution.

1.5 Most commonly used analyses for a RCT with

noncompliance

To analyze data in a RCT with noncompliance, the most commonly used approaches include as-protocol (AP) analysis, as-treated (AT) analysis, and intent-to-treat (ITT) analysis.

> **AP Analysis** – Patients are compared between those who comply with their assigned treatments. Patients who do not comply with their assigned treatments are excluded from data analysis.
>
> **AT Analysis** – Patients are compared according to the treatment they actually receive regardless of what their originally assigned treatment.

Because noncompliance often does not occur randomly, both the AT and AP analyses generally produce a biased inference of treatment efficacy. To clarify this point, we consider the following numerical examples.

Example 1.4 Consider comparing an experimental treatment with a standard treatment in a simple noncompliance RCT. Suppose that our population consist of two subpopulations: 70 % compliers (who fully accept whatever their assigned treatment) and 30 % never-takers (who always take the standard treatment regardless of whatever their assigned treatment). Suppose further that we randomly assign patients to either an experimental treatment ($g = 1$) or a standard treatment ($g = 0$). First, consider the case of equal treatment efficacy between the two treatments. Say, the conditional probabilities of death, given a complier for both the experimental and standard treatments, are given by: $P(death|complier, g = 1) = P(death|complier, g = 0) = 0.30$. Furthermore, because never-takers, who are randomly assigned to the two treatments under comparison, will take the same (standard) treatment, the conditional probabilities of death, given a never-taker, can be reasonably assumed to equal to each other between the two assigned treatment groups. This is actually the exclusion restriction assumption defined in the above for never-takers. We arbitrarily assume that these conditional probabilities are given by $P(death|never - takers, g = 1) = P(death|never - takers, g = 0) = 0.60$. If we randomly assigned 500 patients to each of the two treatments, we would obtain the expected frequencies as given in Table 1.2 (Exercise 1.1). When using the AP analysis excluding those patients who do not comply with their assigned treatments from the experimental treatment group, we obtain the hypothetical mortality data in Table 1.2a.

Table 1.2 The expected cell frequency and the corresponding cell proportion (in parenthesis) for the experimental and standard treatments under the simple noncompliance RCT as described in Example 1.4.

	Patients assigned to the experimental treatment		
	Compliers	Never-takers	Total
Death	105 (21 %)	90 (18 %)	195 (39 %)
Survival	245 (49 %)	60 (12 %)	305 (61 %)
Total	350 (70 %)	150 (30 %)	500 (100 %)
	Patients assigned to the standard treatment		
	Compliers	Never-takers	Total
Death	–	–	195 (39 %)
Survival	–	–	305 (61 %)
Total	–	–	500 (100 %)

Table 1.2a The expected cell frequency and the conditional cell proportion (in parenthesis), given the column total fixed, between patients complying with their assigned treatment for the AP analysis.

	Patients complying with the assigned experimental treatment	Patients complying with the assigned standard treatment
Death	105 (30 %)	195 (39 %)
Survival	245 (70 %)	305 (61 %)
Total	350 (100 %)	500 (100 %)

Table 1.2b The expected frequency and the conditional cell proportion (in parenthesis), given the column total fixed, between patients according to their actually received treatment for the AT analysis.

	Patients actually received the experimental treatment	Patients actually received the standard treatment
Death	105 (30 %)	285 (44 %)
Survival	245 (70 %)	365 (56 %)
Total	350 (100 %)	650 (100 %)

Based on these data, there is strong evidence that the mortality rate in patients complying with the assigned experimental treatment is lower than the mortality rate in patients complying with the assigned standard treatment (Exercise 1.2), although the underlying mortality rates among compliers between these two treatments are actually equal, as assumed in the example. The estimated proportion difference (PD) in the mortality rate between the two comparison groups in the AP analysis is −0.09 (= 105/350-195/500). Also, when using the AT analysis by comparing patients according to the treatment they actually receive, we obtain the data in Table 1.2b.

Again, there is strong evidence that the mortality rate in the experimental treatment is lower that that in the standard treatment (Exercise 1.3). The estimated PD in the mortality rate between the two comparison groups is −0.14. The above results illustrate the case in which the bias in inference can occur for both hypothesis testing and estimation when we employ the AP and AT analyses to study treatment efficacy. To alleviate this concern, the ITT analysis has been suggested.

> **ITT Analysis** – Patients are compared according to the treatment to which they are randomly assigned, despite what treatment they actually receive. Thus, the ITT analysis is also called as-randomized (AR) analysis (Heitjan, 1999).

When using the ITT analysis in the above example, we can easily see that the estimated PD is 0

(= 195/500-195/500) based on Table 1.2, and there is obviously no evidence against the underlying assumed condition that the two treatment effects are equal to one another. These illustrate that the ITT analysis is unbiased in both hypothesis testing and estimation when an experimental treatment effect is equal to a standard treatment effect. In Chapter 2, we will explicitly show why use of the ITT analysis is unbiased under the null relative treatment effect and certain assumptions. On the other hand, the ITT analysis can be biased in estimation of the relative treatment efficacy when the underlying two treatment effects are different. To illustrate this point, we consider the following numerical example.

Example 1.5 Consider the above simple noncompliance RCT in Example 1.4, in which the population consist of two subpopulations: 70 % compliers and 30 % never-takers. However, we now assume that the conditional probability of death $P(death|compliers, g = 1) = 0.30$ among compliers assigned to the experimental treatment is different from the conditional probability of death $P(death|compliers, g = 0) = 0.70$ among compliers assigned to the standard treatment. Thus, the assumed underlying PD among compliers in the mortality rate is −0.40. We assume that the conditional probability of death among never-takers remains the same as that given in the previous example. If we randomly assigned 500 patients to each of the two treatments, we would obtain the expected frequencies as given in Table 1.3 (Exercise 1.4).

Table 1.3 The expected cell frequency and the corresponding cell proportion (in parenthesis) for the experimental and standard treatments under the simple noncompliance RCT as described in Example 1.5.

	Patients assigned to the experimental treatment		
	Compliers	Never-takers	Total
Death	105 (21 %)	90 (18 %)	195 (39 %)
Survival	245 (49 %)	60 (12 %)	305 (61 %)
Total	350 (70 %)	150 (30 %)	500 (100 %)
	Patients assigned to the standard treatment		
	Compliers	Never-takers	Total
Death	–	–	335 (67 %)
Survival	–	–	165 (33 %)
Total	–	–	500 (100 %)

– denotes that the cell frequency is unobservable.

When using the AP analysis, we obtain the data in Table 1.3a. Again, there is strong evidence that the mortality rate in the experimental treatment is lower than that in the standard treatment using the data in Table 1.3a (Exercise 1.5). The estimated PD is −0.37 (= 105/350-335/500), which is slightly different from the underlying assumed PD = −0.40. On the other hand, when using the AT analysis, we obtain the data in Table 1.3b. There is also strong evidence that the mortality rate in the experimental treatment is lower that that in the standard treatment (Exercise 1.6). The estimated PD is −0.35 for the AT analysis. When using the ITT analysis in the above example, we can easily see that the estimated PD in the mortality rate is −0.28 (=195/500-335/500), which is larger than the underlying assumed PD= −0.40 by 30 % (= (|−0.28 + 0.40|/0.40). This illustrates that the ITT analysis can be biased in estimation of the relative treatment efficacy and the