WHO/HRP/EPI/1994
Distr: LIMITED

An Epidemiologic Approach to Reproductive Health

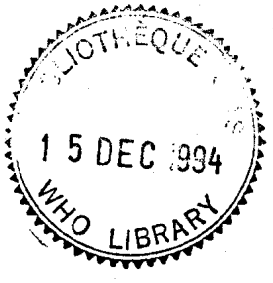
Centers for Disease Control
Atlanta, Georgia, U.S.A.

Family Health International
Research Triangle Park, North Carolina, U.S.A.

World Health Organization
Geneva, Switzerland

Editors

Phyllis A. Wingo
James E. Higgins
George L. Rubin
S. Christine Zahniser



WHO/HRP/EPI/1994
Distribution: LIMITED

Copies of this document are available on request from:

**The Special Programme of Research,
Development and Research Training in
Human Reproduction,
World Health Organization,
1211 Geneva 27,
Switzerland**

Editors

Phyllis A. Wingo, MS, Section Chief, Statistics and Computer Resources Branch, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia. Ms. Wingo served as the Managing Editor. She was instrumental in the development and production of this manual. She determined the emphasis of the material; assembled the team of authors, contributors, and reviewers who are expert in the field of reproductive health epidemiology; coordinated the development and integration of the text material; and provided technical editing and the final review. Ms. Wingo also prepared the current format and schedule for the two-week workshop that was designed from the material in the manual.

James E. Higgins, PhD, Director of Biostatistics and Quality Assurance, Clinical Research International, Research Triangle Park, North Carolina. Dr. Higgins collaborated with Dr. George Rubin to plan the present version of the manual. Using published reports, he developed many of the examples that are used in the present text and workshop. His extensive contributions and revisions produced the reference-like format and the self-help approach to learning epidemiologic research design and methodology. Dr. Higgins provided technical review of the present materials.

George L. Rubin, MBBS, FRACP, Director, Epidemiology and Health Services, Evaluation Branch, Department of Health, New South Wales, Sydney, Australia. Dr. Rubin was instrumental in developing and contributing to the first versions of the manual. Together with Christine Zahniser and Dr. Roger Rochat, he developed the manual and workshop for the first regional workshops. He collaborated with Dr. Higgins to plan the present version of the manual. Many of the examples incorporated in the present text and workshop were developed by Dr. Rubin from published reports.

S. Christine Zahniser, BSN, MPH, Nurse Epidemiologist, Pregnancy and Infant Health Branch, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia. Ms. Zahniser was instrumental in developing the concepts and in applying principles of instructional technology to the first versions of the manual. With Dr. Rubin and Dr. Rochat, she designed the format for the first regional workshops. Ms. Zahniser also developed many of the examples and practice exercises that are incorporated in the present text and workshop.

Partial funding for the development of these materials was provided by the U.S. Agency for International Development (USAID) through a Resources Support Services Agreement (RSSA) with the Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, and through a Cooperative Agreement with Family Health International (FHI).

Authors

Jose Becerra, MD, MPH, is coauthor of Chapter 3. Dr. Becerra is a medical epidemiologist, Pregnancy and Infant Health Branch, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Nancy J. Binkin, MD, MPH, is the author of Chapter 9. Dr. Binkin is a medical epidemiologist with the Centers for Disease Control and is detailed to the World Health Organization and Istituto Superiore di Sanita, Rome, Italy.

Susan Y. Chu, PhD, is the author of Chapter 6. Dr. Chu is an epidemiologist, Surveillance Branch, Division of HIV/AIDS, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia.

Elaine D. Eaker, ScD, is coauthor of Chapter 3. Dr. Eaker is the Assistant Director for Science, Division of Chronic Disease Control and Community Intervention, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Howard I. Goldberg, PhD, adapted the text for Chapter 7 from a manual prepared by Christopher Scott for the Demographic and Health Surveys Project of the Institute of Resource Development, Columbia, Maryland. Dr. Goldberg is Section Chief, Epidemiologic Research and Statistical Support Branch, Division of Chronic Disease Control and Community Intervention, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Joan M. Herold, PhD, is the author of Chapter 2. Dr. Herold is a demographer and Assistant Professor, School of Public Health, Emory University, Atlanta, Georgia.

James E. Higgins, PhD, is coauthor of Chapters 5, 8, and 10. Dr. Higgins is Director of Biostatistics and Quality Assurance, Clinical Research International, Research Triangle Park, North Carolina.

Carol J.R. Hogue, PhD, is the author of Chapter 1. Dr. Hogue also developed many of the examples that are incorporated in the text from published reports and provided extensive contributions to earlier versions of the manual. She is Director, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Edmond F. Maes, PhD, is the author of Chapter 4. Dr. Maes is an epidemiologist, Section Chief, Statistics Activity, Division of Environmental Hazards and Health Effects, Center for Environmental Health and Injury Control, Centers for Disease Control, Atlanta, Georgia.

George L. Rubin, MBBS, FRACP, is coauthor of Chapter 10. Dr. Rubin is Director, Epidemiology and Health Services, Evaluation Branch, Department of Health, New South Wales, Sydney, Australia.

Kenneth F. Schulz, MBA, is coauthor of Chapter 8. Mr. Schulz also developed many of the examples that are incorporated in the text from published reports and provided extensive contributions to earlier versions of the manual. He is Associate Director for International Activities, Division of STDs/HIV Prevention, Center for Prevention Services, Centers for Disease Control, Atlanta, Georgia.

Holly B. Shulman, MS, is coauthor of Chapter 5. Ms. Shulman is a mathematical statistician, Statistics and Computer Resources Branch, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Phyllis A. Wingo, MS, is coauthor of Chapters 8 and 10. Ms. Wingo is the Section Chief, Statistics and Computer Resources Branch, Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.

Contents

Chapter 3 Measures of Disease Frequency in Reproductive Health

Chapter 4 Epidemiologic Study Design

Chapter 5 Sample Size and Power

Chapter 6 Descriptive Studies

Chapter 7 Survey Sampling

Chapter 8 Randomized Clinical Trials

Chapter 9 Cohort Studies

Chapter 10 Case-Control Studies

Acknowledgements

Contributors

Elizabeth Fitch, BA, Graphics*
Valerie Ford, BA, Word Processing
Linda Heath, Word Processing
Audrey Pinto, BA, Writer-Editor*
Cheryl Tryon, MS, Training Development Specialist*
Anna Peed Washington, MEd, MAT, Education Consultant
Karen Williams, BA, Word Processing

Technical Reviewers

Robert J. Berry, MD, MPHTM*	Randomized Clinical Trials
Michael E. Dalmat, DrPH*	Developing a Research Proposal
Adele L. Franks, MD*	Descriptive Studies
Eileen M. Gentry*	Survey Sampling
Gary Grubb, MD‡	Descriptive Studies
Dennis Hurley, DrSc†	Cohort Studies, Case-Control Studies
Juliette Kendrick, MD*	Measures of Disease Frequency in Reproductive Health
Nancy C. Lee, MD*	Case-Control Studies
Beverly Martinez-Schnell, PhD*	Sample Size and Power
Herbert B. Peterson, MD*	Reproductive Health Epidemiology
Rebeca Ponce de Leon, MD, MPH‡	Measures of Disease Frequency in Reproductive Health
Audrey F. Saftlas, PhD*	Cohort Studies
Andrew Schmidt, BS*	Exercises
Lynne S. Wilcox, MD*	Epidemiologic Study Design
Pamela H. Wolf, DrPH§	Complete Text
Special Programme of Research, Development and Research Training in Human Reproduction World Health Organization Geneva, Switzerland	Complete Text

*Centers for Disease Control, Atlanta, Georgia, USA
†Clinical Research International, Research Triangle Park, North Carolina, USA
‡Family Health International, Research Triangle Park, North Carolina, USA
§National Institute for Child Health and Human Development, Washington, DC, USA

Preface

This manual has been prepared for clinicians, researchers, public health professionals, and other personnel who wish to conduct epidemiologic studies of reproductive health issues. It is intended to be used in a two-week workshop that includes an introduction to epidemiologic methods, the unique applications of these methods to reproductive health research, and the development of research proposals. The manual is not intended to replace the many excellent resources and texts presently available on these subjects. Course participants use the teaching materials to develop research proposals about reproductive health problems that are of interest to them or to the organizations they represent. Representatives from funding agencies attend the final session of the workshop to observe presentations of the research proposals prepared during the workshop and to make recommendations about further development of the projects and possible funding.

The evolution of this workshop and manual has been extensive. Many organizations and individuals have contributed to this project over a long period of time. In November 1980 while working in Asia, Drs. Linda Atkinson and Oscar (Bud) Harkavy, Program Officers at the Ford Foundation, and Dr. Roger Rochat from the Centers for Disease Control (CDC) discussed the need to support capacity-building in epidemiologic research in developing countries, particularly in the area of contraceptive safety (broadly defined to include the health consequences of not contracepting as well as maternal mortality). Discussions with health personnel at The United States Agency for International Development (USAID) (Jim Shelton, Tony Boni), the Ford Foundation (Lincoln Chen, Tony Measham), the Population Council (Jarrett Clinton), and the International Development Research Council (Marjorie Koblinsky) provided the impetus to increase the number and quality of contraceptive safety studies being implemented internationally. Training in epidemiology, research methodology, and proposal development were used to accomplish this objective. As a result, five contraceptive safety workshops were developed for presentation in Southeast Asia.

These first workshops were a collaborative effort by the Division of Reproductive Health at CDC, the Population Council, and the Ford Foundation. The objectives for these first workshops included the following:

- To increase the knowledge of health professionals in Southeast Asia about the principles of epidemiology.
- To assist workshop participants in developing research proposals for conducting contraceptive safety studies.
- To increase the number and quality of epidemiologic studies on contraceptive safety conducted in Southeast Asia.

Each of the first five workshops was scheduled to last one week. Didactic sessions and group exercises were the primary approaches used for the

presentation of materials. To supplement the lecture materials, a training manual was developed and included information on the principles of epidemiology and research design, examples of epidemiologic studies from the literature (from Southeast Asia whenever possible), case studies, and practice exercises. Other teaching materials included epidemiology textbooks, statistical monographs, and publications from the medical and epidemiologic literature on contraceptive safety. Participants worked in groups of 4 to 8 persons to develop research proposals on studies of contraceptive safety. The research proposals were presented on the last day of the workshop to a panel of health professionals familiar with contraceptive safety concerns in the region. Feedback was provided by the panel regarding the relevance of the problem selected, the suitability of the selected study design, and the feasibility of implementing the study. These presentations helped to formalize the learning experiences of the participants from the workshop and to publicize to potential funding agencies the research hypotheses and methodologies for contraceptive safety studies that the participants hoped to eventually conduct.

To evaluate the success of the workshop and to help the workshop participants assess their progress during the workshop, tests on the principles of epidemiology were administered on the first day and last days of the workshop. The same test was administered on both days. In general, test scores at the conclusion of the workshop were approximately twice as high as test scores on the pretest.

The first workshop was planned and developed during the summer of 1981 by George Rubin (CDC) and Christine Zahniser (CDC) and conducted on September 21-28, 1981, in Bangkok, Thailand. Andrew Fisher (The Population Council) managed the logistics, including the selection of the workshop participants. David Brandling-Bennett from the World Health Organization (WHO), Andrew Fisher, George Rubin, and Christine Zahniser were workshop instructors. The first workshop trained 12 workshop participants from four countries in Southeast Asia. Three additional workshops were conducted in Bangkok during February 1982, October 1982, and March 1983; additional instructors from CDC included Carlos Huezo and Peter Layde.

After the first Bangkok workshop, CDC received a request for a national workshop in Bangladesh. This workshop was a collaborative effort by CDC, the Program for the Introduction and Adaptation of Contraceptive Technology (PIACT), and The Population Council/International Center for Diarrhoeal Disease Research. David Grimes and Christine Zahniser conducted the first national workshop in Dhaka, Bangladesh, in March 1982. Atiqur Rahman Khan (PIACT), a regional resource person for the September 1981 workshop in Thailand, and Yusef Choudhury (PIACT) arranged the Bangladesh workshop. PIACT received grant monies from the Ford Foundation for contraceptive safety research in Bangladesh that ultimately were used to support research initiated during the workshop.

The series of national workshops in Bangkok stimulated substantial interest among the participants, private organizations, and government organizations such as USAID, WHO, the Ford Foundation, the Population Council, and the

Johns Hopkins Program for International Education in Gynecology and Obstetrics (JHPIEGO) for conducting additional workshops. Furthermore, the newly developed technical expertise of selected participants proved useful for implementing research and future workshops.

Beginning in 1983, Family Health International (FHI) provided technical assistance to an epidemiologic training program implemented by the Instituto de Investigacion Cientifica of the Universidad Juarez, Durango, Mexico. With support from FHI and the United Nations Family Planning Assistance (UNFPA), the Instituto trained and funded investigators from Mexico and other countries in Latin America to conduct epidemiologic studies of reproductive health issues. The training was provided annually to approximately 10 students during a three-week workshop in Durango. Materials used in the workshop evolved from those developed by the Division of Reproductive Health at CDC and were prepared in a self-instructional format in both English and Spanish.

As an outgrowth of the epidemiology training program supported by FHI and UNFPA at the Instituto, the Mexican Interuniversity Group for Epidemiologic Research in Reproductive Health (GIMIESAR) was formed in 1984. GIMIESAR is based in Durango and includes representatives from Mexican universities and medical schools who received training in the Durango workshops. GIMIESAR has embellished the training materials originally developed in collaboration with FHI and has been responsible for organizing and presenting the annual workshop since 1985. Later workshops were coordinated by the Mexican Institute of Social Security (IMSS), WHO, and CDC and were conducted during March and November 1990. Jose Becerra, Daniel Hernandez, Octavio Mojarro, Holly Shulman, and Phyllis Wingo from CDC were workshop instructors.

The first regional workshop in Africa was conducted in Mombassa, Kenya, in 1983 by Ronald Burkman, Michael Dalmat, David Grimes, Peter Lamptey, Japheth Mati, Hamid Rushwan, and Kenneth Schulz. For this workshop, the instructional materials were expanded to include randomized clinical trials, surveys, and sample size and power estimation. The length of the workshop was extended to two weeks, a recommendation that had been made during the Southeast Asia workshops. JHPIEGO managed the administrative functions for this workshop and paid all participant costs. Representatives from potential funding agencies participated in a grant application review process during the last two days of the workshop, and three research projects were funded. The results of a randomized clinical trial of the efficacy of prophylactic antibiotics administered when an intrauterine device was inserted were recently published. The need to train additional reproductive health professionals resulted in a national workshop conducted in 1985 by Barbara Janowitz, John Repke, George Rubin, Kenneth Schulz, and Charles Warren.

Another regional workshop in Africa was sponsored by the Special Programme on Human Reproduction (WHO) and was conducted in Yaounde, Cameroon, in September 1987 by Robert Anda, Nancy Binkin, Barbara Maciak, and Phyllis Wingo. Boniface Nasah (Centre Universitaire pour la Science et la Sante and Centre Hospitalier Universitaire) facilitated the presentation of the workshop.

The collaboration among CDC, FHI, and WHO that resulted in the preparation of the present version of the manual evolved from this workshop. Carol Hogue supported the completion of this project through steady encouragement and the provision of program resources and protected time.

The first national workshop in China, a collaborative effort between the CDC and the Beijing Medical College, was conducted in October 1983 by Charles Chen, Carol Hogue, George Rubin, and Roger Rochat. A participant at the fourth regional workshop in Thailand, Qiao Geng-Mei (Beijing Medical University), was instrumental in arranging the first China workshop. The second workshop in China was requested through WHO and conducted by Jonathan Liff, Wong-Ho Chow, Raymond Bain, and Roger Rochat from Emory University, Atlanta, Georgia, in 1986. For both workshops, the manual was translated into Chinese. All lectures were conducted in English and simultaneously were translated into Chinese.

The first national workshop in Indonesia was coordinated by the Indonesian National Family Planning Board (BKKBN), Yayasan Kusuma Buana (YKB), the Ford Foundation of Indonesia, and the CDC. The workshop was conducted in January 1987 by Nancy Lee, Edmond Maes, and Phyllis Wingo. An Indonesian researcher, Joedo Prihartono (YKB), a participant at the first regional workshop in Thailand, arranged for the participants to attend the workshop and managed the administrative functions of the workshop. The manual was translated into Bahasa Indonesian. The proposal to conduct a case-control study of the possible association between hepatocellular adenoma and the use of oral contraceptives received funding for further development. Although participants ultimately established a surveillance system for identifying biopsy-proven cases of hepatocellular adenoma, the case-control study has not yet been implemented.

A total of 14 workshops have been conducted in eight countries during 1981-1990, including six regional and eight national workshops. More than 300 clinicians, researchers, and public health professionals have participated. CDC continues to receive requests for regional and national workshops on epidemiologic approaches to reproductive health. This training program should complement the Field Epidemiology Training Program supported by CDC and WHO as well as the training program for clinicians that is supported by The Rockefeller Foundation. The paucity of medical and epidemiologic literature on reproductive health issues in developing countries indicate that this training program is unique in at least three aspects:

- This program lasts two weeks while most epidemiology training programs last 6 months to 2 years.

- This program provides the training in the developing countries whereas most other programs provide training at institutions in developed countries. Conducting short-term workshops in reproductive health epidemiology at regional or national levels may allow a greater number of individuals to benefit from this training.

- When representatives from funding agencies attend the presentations of the proposals developed during the workshop, workshop participants have the opportunity to obtain funding to conduct their research.

Roger Rochat
Phyllis Wingo
Christine Zahniser

April 30, 1991

1 Reproductive Health Epidemiology

Introduction

Scope of Study

This workbook is designed to train public health professionals to use epidemiologic methods to answer questions about reproductive health. It serves as the text for the Centers for Disease Control, Family Health International, and the World Health Organization's Reproductive Health Epidemiology Workshop and should provide the user with the necessary skills to develop a protocol for conducting epidemiologic research. Specific skills include basic quantitative measures used in reproductive health epidemiology, epidemiologic study design (descriptive and analytic), sample size and power estimation, survey sample design, and scientific statements of research questions and hypotheses.

Epidemiology defined

Epidemiology can be defined as the study of patterns of human diseases, health, and behaviors. Epidemiologists answer research questions by classifying individuals into one or more discrete groups and assessing the differences between these groups.

Reproductive health epidemiology

Reproductive health epidemiology applies this system of study to questions about maximizing the reproductive health of men and women. Human reproductive health begins with sexual growth and development that is manifest at puberty, continues throughout life for men, and ends at menopause for women. Reproductive health is influenced by fertility and by decisions related to sexual activity, pregnancy, and contraception.

History

Origins of Reproductive Epidemiology

Early Advances

Reproductive epidemiology dates back to 19th century Vienna when Ignaz Semmelweis discovered that deaths from puerperal (childbed) fever were higher among women whose babies were

delivered in the hospital by medical students than among women whose babies were delivered by midwives. He correctly attributed this difference to the midwives' practice of washing their hands between deliveries (MacMahon and Pugh, 1970).

Modern reproductive epidemiology has evolved substantially during the 20th century. The advent of birth and death registries, which were established in the United States during the early part of this century and in Europe during the 18th and 19th centuries, gave health officials a means of identifying risk factors for maternal and infant mortality. Public health measures could then be designed to reduce these factors. For example, milk stations, which provided milk to nursing mothers, were established when a relationship was found between infant mortality, sanitation, and nutrition (Holland et al., 1984). This and similar measures led to a 95% decrease in maternal mortality in the United States from 1915 to 1965 (Hogue CJR, personal communication). In Sweden, the infant mortality rate declined from 200 deaths per 1,000 live births in the 1750s to less than 20 per 1,000 in the 1950s (United Nations, 1963).

Influence of Demography and Epidemiology

Stages of demographic transition

Reproductive epidemiology also has origins in demography, a discipline that evolved during the 19th century from studies of parish registries in England. Historical changes in the population (demographic transitions) and related changes in the patterns of health and disease (epidemiologic transitions) directly affect mortality, fertility, birthrates, and other measures of reproductive health. These changes also affect the health and status of women, children, and the family.

The original theory of demographic transition (Notestein, 1945) describes three stages of population growth that accompanied economic development in Western countries.

High growth potential. During this stage, rates for births and deaths, although high, are at similar levels, so population growth is minimal. If mortality were to decline at this stage without a concurrent decline in fertility, the size of the population would increase rapidly.

Transition. This stage begins with a decline in mortality while fertility remains high, and shifts to a decline in fertility until both rates are at similarly low levels. During the first part of this stage, the high growth potential is realized; during the latter part, population growth declines.

Incipient decline. During this final stage, rates for births and deaths are at low and relatively stable levels. Sometimes, however, fertility rates decrease to levels lower than death rates and produce a decline in population.

Although demographic transition provides a perspective for interpreting historical changes in Western populations, the theory does not completely describe or explain patterns of population change in non-Western societies or those in developing countries where external factors have affected a dramatic decline in the mortality rate without a concomitant decline in the birthrate (Hauser and Duncan, 1959; Notestein and Segal, 1963).

Stages of epidemiologic transition

Epidemiologic transition theory describes changes in patterns of health and disease by focusing on mortality and fertility rates and on the interaction of social, economic, demographic, and health variables (Omran, 1974). The three stages of epidemiologic transition parallel and influence the three stages of demographic transition.

Age of pestilence and famine. During this stage, the prevalence of endemic diseases is high, nutrition is poor, and infectious diseases and famine are rampant. The rates for births and deaths are high, and population growth minimal. Extended family structures with large family size, multiple-generation households, and home-centered lifestyles are dominant. Women function as mothers with no rights or responsibilities outside the home (Omran, 1974).

Age of receding pandemics. At this stage, disease and famine decrease, mortality rates decline, birthrates increase, and populations grow. Large extended families are predominant, especially in rural areas, but nuclear family units become more common in urban centers. Women begin to become involved in activities outside the home (Omran, 1974).

Age of degenerative and man-made diseases. During this period, social, economic, and environmental conditions improve; infectious diseases and conditions related to poor nutrition decline. Birth and death rates are low, and population size is stable. Chronic diseases (e.g., cardiovascular disease, cancer, stroke, diseases caused by occupational exposures, etc.) are the primary causes of death. Small nuclear families become the norm. *Women are increasingly emancipated from traditional roles and become better educated and more career-oriented* (Omran, 1974).

Status Of Reproductive Health

Substantial variation in maternal and infant mortality rates exist throughout the world today. Among the ten most populous countries, maternal mortality rates vary from 1,500/100,000 live births to less than 15/100,000 live births, a 100-fold range. Infant mortality varies from 140 to 5 per 1,000 live births (Table 1.1).

In the United States, epidemiologists are seeking explanations for the diminishing decline in U.S. maternal and infant mortality rates, while these rates continue to decrease in other developed countries. They are also concerned about the excess mortality among minorities in the United States. For example, infant mortality is twice as high among blacks as among whites, and maternal mortality is more than two times higher among black women than among white women (Hogue et al., 1987; Rochat et al., 1988).

Population growth rates also vary considerably. At the current rate of growth, three of the ten most populous countries—Bangladesh, Nigeria, and Pakistan—will double in population in less than 30 years (Hatcher et al., 1989). By comparison, the two most populous countries with the lowest growth rates—Japan and the United States—will double in 133 and 99 years, respectively. In Nigeria and Bangladesh, the two countries with the highest total fertility rates, births per 1,000 females aged 15 to 19 years exceed 200.

In the ten most populous countries, the prevalence of contraceptive use among married women aged 15 to 44 years varies from less than 10% to more than 80% (Hatcher et al., 1989). In all but one of these countries, 10% or fewer of the couples of reproductive age use condoms, which provide both contraception and protection from HIV and other sexually transmitted infections.

Table 1.1

Reproductive Health Around the World

	Population Estimate Mid-1988 (in millions)	Crude Birth Rate (births per 1000 population)	Crude Death Rate (deaths per 1000 population)	Natural Increase (annual %) (births minus deaths per 100 population)	Years for Population to Double (at current rate of increase)	Total Fertility Rate (TFR) (# children per woman by age 49)	% Decline in TFR 1970-75 to 1985-90	Infant Mortality Rate (deaths in 1st year of life per 1000 live births)	Maternal Mortality Ratio (maternal deaths per 100,000 live births)	Life Expectancy at Birth - 1988 (in years)	Human Suffering Index[1] (maximum score = 100)	Status of Women[2] (maximum score = 100)
China	1087	21	7	1.4	49	2.1	55	44	44	66	50	58.5
India	817	33	13	2.0	35	3.7	31	101	400-500	54	61	43.5
U.S.S.R	286	20	10	1.0	68	2.4	0	25	..	69	19	77.0
United States	246	16	9	0.7	99	1.9	5	11	14	75	8	82.5
Indonesia	177	27	10	1.7	40	3.5	36	88	357-800	58	62	46.5
Brazil	144	28	8	2.0	34	3.5	26	63	87-154	65	50	54.5
Japan	123	11	6	0.5	133	1.8	14	5	16	78	11	68.5
Nigeria	112	46	17	2.9	24	7.1	0	124	1500	47	80	29.0
Bangladesh	110	43	17	2.7	26	5.5	21	140	623	50	79	21.5
Pakistan	108	43	15	2.9	24	5.3	18	125	400-600	54	73	28.0
Mexico	84	30	6	2.4	29	4.0	38	50	92	66	47	61.5
United Kingdom	57	13	12	0.2	408	1.8	14	9	12	75	12	74.5
Egypt	53	38	9	2.8	24	4.3	22	93	269	59	55	38.0
Turkey	53	31	9	2.2	32	3.7	33	92	207	63	55	52.5
World	5128	28	10	1.7	40	3.3	25	81	390	63	55	..

(Source: Hatcher et al., 1989)

Table 1.1 (continued)

	% Married Women Ages 15-44 Using Any Method of Birth Control	Percent Couples of Reproductive Age Using											
		Pills	IUDs	Female Sterilization	Male Sterilization	Condoms	Injectables	Withdrawal	Natural Family Planning	Vaginal Methods	Other Traditional	Access to Birth Control (maximum score = 100)
China	81	5	32	30	9	3	—	—	—	1	1	84
India	39	5	4	26 (combined)		4	—	—	—	—	..	64
U.S.S.R	75	..	..	..	..	..	..	..	..	..	..	..
United States	68	14	5	17	11	10	—	1	3	7	..	83
Indonesia	46	14	13	3	—	2	10	..	..	4	..	67
Brazil	65	25	1	27	1	2	—	5	4	1	..	56
Japan	64	1	4	8	3	43	—	1	4	..	1	63
Nigeria	5	1	1	—	—	—	—	—	—	—	3	21
Bangladesh	25	5	1	8	2	2	1	1	4	—	1	64
Pakistan	8	1	1	2	—	2	1	1	—	—	—	29
Mexico	53	10	11	19	1	2	3	..	..	..	8	72
United Kingdom	83[3]	24	7	14	14	17	—	6	2	3	—	97
Egypt	30	17	8	2	—	1	—	—	1	1	1	37
Turkey	62	9	9	1	—	5	—	30	1	3	3	39
World	51	7	10	13	5	5	1	4	4	1	1	..

Table 1.1 (continued)

		Abortion		Breastfeeding			Adolescent Pregnancy & Determinants			
	Status of Abortion I - V[4]	# of Abortions per 1000 Live Births	% Children Ever Breastfed	% Breastfeeding at 3 Months	% Breastfeeding at 6 Months	Births per 1000 Females Ages 15-19	Singulate Mean Age at Marriage for Females	% Females Marrying Before Age 20	% Females Ages 15 and Over Enrolled in Secondary School	% Females Illiterate Ages 15 and Older
China	I	490	..	56	55	12	23	4	35	34
India	II	247	98	90	79	41	19	57	20	25
U.S.S.R.	I	2080	..	..	..	16	..	27	77	2
United States	I	422	58	36	22	53	22	8	92	1
Indonesia	IV	..	97	93	92	31	16	37	24	42
Brazil	IV	250-1200[5]	91	56	31	81	22	18	35	28
Japan	II	382-1492	75	..	..	3	25	1	93	1
Nigeria	III	..	100	100	89	213	19	44	16	77
Bangladesh	III	45	99	98	97	237	16	63	6	87
Pakistan	III	..	97	..	91	141	21	38	7	87
Mexico	IV	..	87	78	52	80	18	19	49	..
United Kingdom	II	223	51	15	9	28	23	5	84	..
Egypt	III	..	95	91	86	99	21	22	45	75
Turkey	I	256	90	83	53	67	18	16	28	..
World		300-500	..	..	..	..	..	..	..	..

These data highlight the urgent need worldwide for improved family planning and maternal and child health services. The availability of existing contraceptives should also be improved, and new methods should be developed and tested. Epidemiologic research has an important role in improving each of these areas.

Uses of Epidemiologic Methods in Reproductive Health Practice

Testing, intervention, and evaluation

Epidemiologic methods are used to define reproductive health problems, to elucidate the causes of these problems, to test interventions, and to evaluate programs (Table 1.2). Problem definition involves describing the population affected, the etiology of the health problem, identifying the alterable risk factors, and conducting ongoing surveillance to detect trends in the problem. Reduction of risk factors through intervention depends on accurate assessment of the comparative safety and efficacy of proposed interventions and treatments. Analytic epidemiology is used to test such interventions. Epidemiologic methods and results are used to assess whether a program is based on the appropriate interventions and treatments and whether the interventions and treatments are being properly used. Cost-benefit analysis is applied to determine whether interventions make the best use of available resources.

Example 1

Shunyi Risk Approach Project in Perinatal Health

The Shunyi Risk Approach Project in Perinatal Health (Yan et al., 1989), which was conducted in Shunyi County, People's Republic of China, shows how epidemiologic methods are applied to problem definition, intervention and testing, and program evaluation. The project used the *risk approach* in an attempt to improve perinatal health services. It began in 1983 and continued for five years.

Problem definition. Investigators collected data on 1,914 pregnant women and their 1,928 infants, and on 50 cases of perinatal mortality. Among other problems, the investigators found that 151 per 1,000 women suffered from hypertensive disorders during pregnancy, and 1 per 1,000 suffered from eclampsia. The perinatal mortality rate for

the infants born to these women was elevated. For example, the perinatal mortality rate for infants born to women who experienced mild to severe hypertension that improved during pregnancy was 4.6 per 1,000 deliveries, or more than twice the rate for pregnancies uncomplicated by these disorders. The rate for infants born to women with hypertensive disorders that worsened or remained severe throughout pregnancy was 10.8 per 1,000 deliveries.

Table 1.2

Uses of Epidemiologic Methods in Reproductive Health Practice

Uses	Problems Addressed
Problem definition	Population affected Risk factor identification Surveillance
Intervention and testing	Comparative safety of treatments Comparative efficacy of treatments
Program evaluation	Risk approach Cost-effectiveness

Intervention and testing. To reduce the incidence of pre-eclampsia, eclampsia, and perinatal mortality associated with hypertensive disorders of pregnancy, the investigators initiated a number of interventions in 1985. They educated patients about the need for rest, proper nutrition, and the signs and symptoms of eclampsia. High-risk women were informed about the need for weekly or biweekly blood pressure measurements. The investigators also provided training to health care providers and taught village doctors to measure blood pressure and to check their equipment. Practitioners in the township hospital were taught how to diagnose and treat hypertensive disorders, to make proper referrals to the county hospital, and to follow a set protocol for monitoring all pregnant women.

Program evaluation. The investigators conducted epidemiologic surveillance to assess the impact of these interventions. From

1984 to 1986, the incidence of preeclampsia and eclampsia decreased from 1.8% to 0.4%, and the perinatal mortality rate for pregnancies complicated by these disorders declined from 10.8 per 1,000 deliveries to 0. The team of investigators concluded that the interventions had been highly successful in reducing maternal and infant morbidity and mortality related to hypertensive disorders of pregnancy.

Example 2

Surveillance of maternal smoking and pregnancy outcome

Another example of epidemiologic problem definition and intervention is the surveillance of maternal smoking and pregnancy outcome and the study of effective smoking cessation programs.

Problem definition. Over the last three decades, a relationship between maternal smoking and pregnancy outcome has been established through numerous analytic epidemiologic studies. For example, MacMahon et al. (1966) reported that infants born to women who smoked were approximately 200 g lighter than infants born to nonsmoking women. Since then, virtually all studies of maternal smoking and birthweight have confirmed this finding. A dose-response relationship also exists between the number of cigarettes smoked and infant birthweight, regardless of gestational age at birth (Hogue et al., 1987). Descriptive epidemiologic studies have revealed that one in three women in the United States are smokers when they conceive, and one in four women continue to smoke throughout pregnancy (Prager et al., 1984). Worldwide, the prevalence of smoking among women varies considerably. In general, the prevalence of smoking increases with economic development. Using analytic epidemiology, Kleinman et al. (1988) estimated that 10% of infant mortality in the United States could be eliminated if maternal smoking were no longer a risk factor. In the United States, epidemiologic surveillance of maternal smoking has been conducted through the analysis of birth certificates in Missouri since 1979, through the Pregnancy Risk Assessment Monitoring System in six states since 1988, and in the future for all states adopting the revised birth certificate of 1989.

Intervention and testing. Experimental epidemiology has been used to test the effectiveness of smoking cessation counseling to help women who want to stop smoking and to assess the impact on

improved pregnancy outcome for women who do stop smoking. Well-conducted clinical trials have found that some women can stop smoking with counseling assistance. The results of these trials show that the birthweights of the infants born to women who stop smoking by the fifth month of pregnancy are similar to the birthweights of the infants born to women who never smoked during their pregnancy (Sexton and Hebel, 1984; Windsor et al., 1985).

Program Evaluation. Using analytic epidemiologic studies as the data source, Marks et al. (1990) estimated that a national smoking cessation program could save over $5 in care for low-birthweight infants for every $1 spent on the intervention, and that program costs would amount to $69,542 for each of the 338 deaths prevented.

Scope of Epidemiologic Research in Reproductive Health

Epidemiologists have conducted investigations into virtually every aspect of reproductive health, including sexual development, sexual activity, contraception, contraceptive methods, fertility, unintended pregnancy, induced abortion, maternal and infant morbidity and mortality, male and female problems with the reproductive tract, and the delivery of maternal and child health and family planning services. In this workbook, we have selected examples from epidemiologic studies that have been conducted in many countries to illustrate the usefulness of epidemiology in answering important public health questions on reproductive health issues. During the workshop, we will discuss this research to provide a framework for developing research protocols. The examples we provide should not be considered comprehensive or exhaustive because epidemiologic methods can be applied to many areas of human reproductive health.

References

Hatcher RA, Kowal D, Guest F, et al. Contraceptive technology: international edition. Atlanta: Printed Matter Inc., 1989.

Hauser PM, Duncan OD. Demography as a body of knowledge. In: Hauser PM, Duncan OD, eds. The study of population: an inventory and an appraisal. Chicago: University of Chicago Press, 1959.

Hogue CJR, Sappenfield W. Smoking and low birth weight: current concepts. In: Rosenberg MJ, ed. Smoking and reproductive health. Littleton, Massachusetts: PSG Publishing Company, 1987:97-102.

Hogue CJR, Buehler JW, Strauss LT, Smith JC. Overview of the National Infant Mortality Surveillance (NIMS) Project—design, methods, results. Public Health Rep. 1987;102:126-138.

Holland WW, Detels R, Knox G, eds. Volume 1: history, determination, scope, and statistics. In: Oxford textbook of public health. New York: Oxford University Press, 1984:1-12.

Kleinman JC, Pierre MB, Madans JH, Land GH, Schramm WF. The effects of maternal smoking on fetal and infant mortality. Am J Epidemiol 1988;27:274-282.

MacMahon B, Alpert M, Salber EF. Infant weight and parental smoking habits. Am J Epidemiol 1966;82:247-261.

MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown & Company, 1970:35.

Marks JS, Koplan JP, Hogue CJR, Dalmat ME. A cost-benefit/cost-effectiveness analysis of a smoking cessation program for pregnant women. Am J Prev Med [In Press].

Notestein FW. Population—the long view. In: Schultz TW ed. Food for the world. Chicago: University of Chicago Press, 1945.

Notestein FW, Segal S. The problem of population control. In: Hauser PM ed., The population dilemma. Englewood Cliffs, New Jersey: Prentice-Hall, 1963.

Omran AR ed. Community medicine in developing countries. New York: Springer Publishing Co., 1974.

Prager K, Malin H, Spiegler D, Van Motts P, Placek PJ. Smoking and drinking behavior before and during pregnancy of married mothers of live-born infants and stillborn infants. Public Health Rep 1984;99:117-127.

Rochat RW, Koonin LM, Atrash HK, Jewett JF. Maternal mortality in the United States: report from the maternal mortality collaborative. Am J Obstet and Gynecol 1988;72:91-97.

Sexton M, Hebel JR. A clinical trial of change in maternal smoking and its effect on birth weight. JAMA 1984;251:911-915.

Windsor RA, Cutter G, Morris J, et al. The effectiveness of smoking cessation methods for smokers in public health maternity clinics: a randomized trial. Am J Public Health 1985;75:1389-1392.

Yan RY, McCarthy BJ, Ye HF, et al. The risk approach in perinatal health: Shunyi County, People's Republic of China. Atlanta, Georgia: Centers for Disease Control (World Health Organization Collaborating Center in Perinatal Care and Health Service Research in Maternal and Child Health), 1989; DHHS publication no. (HHS)89-8412.

United Nations. The situation and recent trends of mortality in the world. Population Bulletin of the United Nations, No. 6. New York, 1963.

Learning Objectives - Chapter 2

After completing this chapter you should be able to:

1. Write an abstract of a research proposal.
2. Use the literature to identify a research problem.
3. Define a research problem.
4. Write a justification for a proposed study.
5. Write an ultimate goal for a proposed study.
6. Write research objectives for a proposed study.
7. Write research questions or hypotheses for descriptive and analytic research.
8. Distinguish four types of study designs.
9. Recognize parts of each of the following that must be included in all research plans and outlined in a research proposal:

   research methods
   analysis of data
   interpretation of findings
   reporting the findings
   logistics
   work schedule

10. Design a research proposal following guidelines given.

2 Developing a Research Proposal

Introduction

Research is conducted in response to a problem, such as an epidemic or an unexpected health outcome, or in response to a need to plan or change a program or a course of action, to test a hypothesis, or to further study recent research findings. Inquiries from politicians or journalists and scientific or medical curiosity also stimulate research.

A protocol is developed as a guide for a study.

Having decided to pursue a research project, the researcher needs to develop a plan or protocol as a guide for the study. A plan written to seek approval for research from a supervisor or organization is called a research proposal.

The format and content of research proposals vary widely. This chapter defines the standard information usually included in a research proposal. The order in which this information is presented can vary.

The Proposal Abstract

Abstract

The abstract is a summary of the proposed study. Although it usually appears as the first section of the proposal, it is not written until all other sections are completed. The abstract should be brief—approximately 200 words—and should answer as many of the following questions as possible:

- What is the problem to be studied?
- What are the research questions or hypotheses?
- What are the expected implications of the study?
- Who will conduct the study (data collection and/or analysis)?
- When will the study be conducted (data collection and/or analysis)?
- Where will the data collection be conducted?

- What methods will be used to collect and/or analyze the data?
- What resources are required to conduct the study?

Examples 2.1 and 2.2 provide hypothetical abstracts for study proposals.

Example 2.1

Abstract for a Case-Control Study of Adolescent Pregnancy

Adolescents constitute a higher percentage (15%-20%) of pregnant women admitted to clinics in Central state than in the southern parts of the country. They also account for a disproportionately high rate of morbidity, maternal mortality, and perinatal mortality in the state. Because of this situation, the Ministry of Health wishes to determine the major differences in attitudes toward and practice of contraception between younger and older women.

This descriptive, case-control study in Central state will be conducted between January and December 1995 by the Ministry of Health with technical assistance from the Centers for Disease Control. It will compare the knowledge, attitudes, and practices of fertility control among (1) primigravidae less than 18 years of age and their husbands and (2) primigravidae 18 years of age and older and their husbands.

A sample size of 800 couples will be selected for each group from antenatal clinics (assuming a rate of ignorance of contraception of 70% in the under 17 group and 60% in the 17 and over group, a power of 90%, and a confidence level of 0.1). These couples will be interviewed by highly trained interviewers using precoded questionnaires.

Information from the study will help the government and officials responsible for managing health care delivery systems to find appropriate ways of preventing pregnancies in this high-risk group and thereby reduce the costs to the health care delivery system. Information will be disseminated through:

- A seminar on adolescent pregnancy problems sponsored by the Congressional Committee on Health.
- Health Management Board.
- Women's health associations.
- Educators.
- Community leaders.

The total cost of this project is estimated at $47,000 to $55,000.

Example 2.2

Abstract for a Randomized Clinical Trial of Malaria Prophylaxis in Pregnancy

Pregnancies complicated by malaria are becoming an increasingly important problem in Country Y and in many other developing countries. Adverse outcomes include fetal death as well as maternal mortality and morbidity. Routine prophylaxis against malaria in pregnancy might help to prevent these adverse outcomes, and yet the value of prophylaxis has not been established in Country Y.

A randomized clinical trial will be conducted by a research team from State University during 1991-1992. The study will compare the use of weekly prophylaxis against malaria for 2,500 pregnant women with the weekly use of a placebo for 2,500 pregnant controls. Patients will be enrolled from the obstetric population of State University and Western University clinics.

The research hypothesis of this study is that weekly prophylaxis will reduce the incidence of intrauterine fetal death from 5% to 3%. The sample size will provide 90% power with a p value of 0.05.

Although the primary focus of the study is to reduce intrauterine fetal death, other complications of malaria, such as morbidity, anemia, and congenital malaria, will also be studied.

The findings will be incorporated into the existing program plans of the Ministry of Health.

Problem Identification and Definition

The literature review

This section of the proposal depends upon information from previous research and the literature. It should therefore contain the literature review. Before writing this section, you should:

1. Search the literature thoroughly for information about the problem you are researching. Establish whether or not others are currently engaged in similar or related research.

2. Classify key literature on the subject. It is often useful to do this in tabular form (Table 2.3).

3. Identify critical areas for research (i.e., shortcomings of previous studies or areas where no research has been or is being done).

After careful examination, the medical and epidemiologic literature may be used to identify and define the problem under study.

Table 2.3

Literature Review of Selected Studies of Infertility After Induced Abortion

Study Location and Time	Design	Abortion Procedure*	Controlled Variables	Results
Japan 1971	Cross-sectional	Mostly D&C	Age, duration of marriage	Gravidity was similar for those reporting and not reporting previous induced abortions.
Yugoslavia 1966-1972	Abortion cohort	Mostly VA	Contraceptive use	Pregnancy-to-conception intervals were significantly shorter after induced abortion.
Greece 1974	Case-control	Illegal D&C	Age, parity, education	For secondary infertility >18 months, relative risk (RR) = 1.1 (95% CI = 0.6, 1.9).
		Illegal D&C	Age, parity, education, spontaneous abortion history	RR = 12.5 (95% CI = 2.3, 66.9)
Denmark 1967-1975	Pregnancy cohort	Mostly VA	Age, parity, socioeconomic status, regular menstrual cycle, contraceptive use	For pregnant women, the outcome of the preceding pregnancy did not significantly affect the length of the subsequent interpregnancy interval.
Washington 1976-1978	Case-control	Mostly VA	Age, gravidity, race or ethnic status, marital status, socioeconomic status	For secondary infertility >12 months, RR = 1.3 (95% CI = 0.1, 2.4)
		Mostly VA	Age, gravidity, race or ethnic status, marital status, socioeconomic status, spontaneous abortion history	RR = 1.2 (95% CI = 0.4, 3.0)
Hawaii 1970-1979	Abortion cohort	Mostly VA	Age, gravidity, menstrual regularity, contraceptive use	Several analytic techniques resulted in the conclusion that induced abortion had not impaired ability to conceive.

*D&C = dilatation and curettage; VA = vacuum aspiration.

(Hogue et al., 1982)

Problem identification for analytic epidemiologic research

Identify the problem: State why you think the problem requires study. Problem identification for analytic epidemiologic research must proceed from the following three conditions:

1. Indicate the discrepancy between the real or observed situation (*what is*) and the ideal, desired, or theoretical situation (*what should be*).

2. Indicate the alternative solutions or explanations for the discrepancy.

3. Indicate which of the alternatives you believe is the most likely to be correct and why.

Problem identification for descriptive research

For exploratory or descriptive research, only the first condition is required. Conditions 2 and 3 may or may not apply.

Example 2.4 illustrates three hypothetical problem identification statements. The *what is* is indicated in bold text, and the *what should be* is noted in italics.

Example 2.4

Three Problem Identification Statements

Infant Mortality (Analytic Study): **Infant mortality is higher in Village A than in neighboring villages.** *There should be no difference in infant mortality between these villages.* Possible explanations are that Village A has (a) lower nutrition levels, (b) lower levels of breast-feeding, (c) poorer sanitary conditions, (d) lower maternal literacy. Our knowledge of lifestyles in these villages leads us to believe that infant mortality is higher in Village A because sanitary conditions are poorer than those in surrounding villages.

Increase in Vasectomies (Analytic Study): **Thirty vasectomies were performed in region A over the past six months. This number constitutes an abnormal increase,** since *the region usually averages two vasectomies a month.* This increase may be due to support for this procedure from local community leaders or it may be due to a new promotional campaign undertaken by family planning clinics in Region A. Our team believes the primary influence is the new promotional campaign.

Adolescent Fertility (Descriptive Study): **Fertility is higher among adolescents than among women 18 to 24 years of age in City X.** Public

Example 2.4 (continued)

health officials felt *the fertility rate for adolescents should be lower.* In response to this concern, we propose to study the factors associated with adolescent pregnancy in order to identify high-risk groups and to begin formulating hypotheses regarding interventions for further study.

Define the problem: Summarize current research and list issues needing further study. Problem definition may include information on:

- Magnitude. What is the incidence and prevalence of the problem?
- Time frame. When does it occur? Is it current?
- Geographic area. Where does the problem generally occur?
- Population. Does the problem affect certain groups of people? If so, what are their characteristics?
- Why? What are the probable reasons for the problem? Is there agreement or conflict over these reasons?
- Solutions. What solutions have already been tried? How successful have they been? What untried solutions might there be?
- Unanswered questions. What parts of the problem need further research?

Example 2.5 illustrates a hypothetical problem identification and definition.

Example 2.5

Problem Identification and Definition
Family Planning Services in Country X

Problem Identification

- Country X has a shortage of trained health professionals to provide intrauterine devices (IUDs) to women in rural regions.
- Therefore, the government initiated a program in 1985 to provide two months of training in family planning to traditional midwives.
- A study will be conducted to determine IUD retention rates for village women who received IUDs from rural midwives and those who received IUDs from physicians or nurse-midwives in a clinic setting.

Example 2.5 (continued)

Problem Definition

Trained health professionals are currently (time frame) in short supply (magnitude of the problem) in rural areas of Country X (geographic area). Consequently, women residing in rural areas (population) have not adopted effective methods of birth control.

In many developing countries, and particularly in rural areas, where there are few trained health providers, auxiliary personnel serve as health care providers. Thus, since 1985, traditional midwives have been providing family planning services to rural women in Country X (solution).

The traditional midwives are well accepted as paramedics in rural areas. Their traditional tasks of delivery and care of the mother and child permit them to easily provide family planning services to those mothers who wish them (elaboration).

The effect of traditional midwives on people's decisions to use contraceptives and their continued use in rural Country X has never been evaluated (unanswered questions).

In addition, this problem needs to be studied in the context of previous studies about IUD insertion performed by rural midwives versus insertion by physicians and trained nurse-midwives.

Justification

Justify the proposed study

The justification of the research topic is an important part of any research proposal. Research is often expensive and time consuming, and most funding agencies are reluctant to support studies unless the results have direct program implications. When funds are limited (as they almost always are), it is especially important for the research investigator to justify the proposed study carefully. In doing so, the investigator should place himself or herself in the position of the approving official and should consider what criteria would be used to select the study if there were funds to support only one of several proposed studies.

In writing the justification, it is usually helpful to consider the following questions and then arrange the answers to these questions into a few concise paragraphs:

- Is the problem a current and timely one? In other words, does the problem exist now? Current problems are more likely than past problems to receive funding.

- Does the problem have life-threatening or serious morbidity consequences? Poor surgical technique during sterilization can have life-threatening or serious morbidity consequences for the patient, whereas occasional spotting from IUD use generally does not have serious consequences.

- Does the problem affect, or potentially affect, a large number of people? Some problems, such as thromboembolism from contraceptive pill use, are life threatening, but of all the people who use oral contraceptives, relatively few are affected. In countries where sterilization is widely used, other problems, such as anesthesia overdose, tetanus, and intraperitoneal hemorrhage, may affect a large number of people.

- Does the problem relate to on-going program activities? That is, does the problem have implications for current programs? For example, a study comparing failure rates and complications of different IUDs is not likely to have major program implications in a country where the IUD is not commonly used.

- Does the problem have broad social, economic, political, or health implications? Some studies may impact many different activities. For example, using nonmedical personnel to provide contraceptive methods may lower maternal mortality and fertility rates and thus have broad social, economic, and political ramifications.

- Is the problem viewed as a concern by many different people? A research problem that evokes the concern of many different people—administrators, politicians, health professionals, the general public—is more likely to receive priority funding than one that only a small group of researchers view as a concern.

- Have many studies already addressed the problem? For some reproductive health issues study has been extensive, and much is already known about the etiologies of the problems. For

example, the complications and failure rates of different IUDs have been widely studied. Would another IUD study add significant new information?

Example 2.6 illustrates problem justification.

Example 2.6

Problem Justification: Prophylactic Use of Antibiotics at the Time of Intrauterine Device Insertion

The Ministry of Health and the Women's Development Movement have completed negotiations with an international donor for a $5 million loan that will be distributed over the next 5 years. The funding will be used to double the existing network of family planning clinics in an effort to increase contraceptive prevalence from 12% to 35% by 1996.

Considerable controversy exists over whether or not to promote the use of IUDs in existing and future rural and urban family planning clinics because IUD use is associated with an increased incidence of pelvic inflammatory disease (PID) (Faulkner and Ory, 1976; Kaufman et al., 1980; Vessey et al., 1981). Two studies were conducted in Country X— one at the National Teaching Hospital in 1980 and one in three provincial hospitals in 1981. The findings indicated that between 28% and 35% of women examined in the obstetrics and gynecology wards had PID and suggested that PID may be a prevalent condition. PID is expensive and difficult to treat in most clinics. If untreated, PID can lead to infertility, a problem that is thought to be widespread in certain areas of Country X.

Thus, before the Ministry of Health can promote IUD use in these areas, methods for decreasing IUD-associated PID need to be identified. If IUDs can be used safely and if one-third of the new users of contraceptives used IUDs, the costs associated with family planning visits for resupply of oral pills and other contraceptives could be reduced by as much as 50%. Service providers could spend more time promoting family planning among unserved high-risk women and thus increase the likelihood of achieving the country's goal for contraception.

A recent study conducted in the United States concluded that IUDs were associated with PID only at times of insertion or reinsertion (Burkman et al., 1981). Consequently, antibiotics given prophylactically to women at the time of IUD insertion may decrease the incidence of IUD-associated PID. However, no research has been conducted to test this hypothesis. The randomized clinical trial to be described is designed to test this hypothesis in rural and urban service delivery clinics in Country X.

Goals and Objectives

Ultimate research goals

Before a study is actually designed, the study's ultimate, long-term, public health goal and its immediate, specific research objectives are written. Ultimate goals should be stated in terms of the potential impact or public health purpose of the study or service delivery program. Although goals are not as detailed as research objectives, they must be clear.

Goals are stated in terms of:

- Broad social, economic, or health concerns.
- Change in policy decisions, service delivery programs, or individual health behavior.
- Populations that may be affected.

Three hypothetical goals are presented in Example 2.7.

Example 2.7

Three Ultimate Goals

By 1995: To decrease by 40% the percentage of clinic or hospital deliveries among women who are at low risk for pregnancy-related complications.

By 1999: To decrease by 20% the death-to-case rate for vasectomies among rural males in Country X.

To improve use of health facilities by farmers in Province A.

Research objectives

Research objectives describe what will be demonstrated, tested, evaluated, confirmed, or compared. They communicate:

- What you plan to do.
- Who will do it.
- To whom it will be done.
- When it will be done.
- Where it will be done.
- What you hope to learn.

Example 2.8 provides details of hypothetical research objectives for three study designs.

Example 2.8

Research Objectives

Cohort Study

The Department of Maternal and Child Health will conduct a cohort study over a two-year period (1993-1995) to compare the risk of complications associated with postpartum tubal sterilization and interval sterilization in 2,000 women living in Central City.

What:	Cohort study
By whom:	Department of Maternal and Child Health
To whom:	2,000 women
When:	1990-1992
Where:	Central City
Purpose:	Risk of complications of postpartum and interval sterilization

Surveillance System

The Centers for Disease Control (CDC) will establish a surveillance system in Division A by the end of 1995 to monitor all pregnancy-related deaths among women aged 15 to 44 years.

What:	Surveillance system
By whom:	Centers for Disease Control
To whom:	Women aged 15 to 44 years
When:	End of 1995
Where:	Division A
Purpose:	Monitor all pregnancy-related deaths

Descriptive Study

A descriptive study will be conducted by the Health Department of District X from June to December 1992 to determine the demographic and social characteristics of the 4,300 women obtaining family planning services in government health clinics.

What:	Descriptive study
By whom:	Health Department
To whom:	Women obtaining family planning services in government health clinics
When:	June to December 1992
Where:	District X
Purpose:	Determine demographic and social characteristics

Research Questions or Hypotheses

All proposals should contain a formal and explicit statement of the research question(s) to be studied or the research hypothesis(es) to be tested. Whether to use questions or hypotheses depends on the type of research. Exploratory or descriptive epidemiologic research does not involve hypothesis testing; it is based on underlying questions. The research questions must be formally stated with clarity, specificity, and appropriate inclusiveness. Example 2.9 provides examples of research questions.

Example 2.9

Research Questions
Exploratory or Descriptive Epidemiology

1. Why does Village Y have half the fertility rate of other populations in the region? Is the lower fertility rate due to more effective use of contraception, different breast-feeding practices or marriage patterns, or other unspecified social practices?

2. What are the levels of maternal and infant mortality in major cities of Country X?

Analytic epidemiologic research is designed to make predictions about the relationships between variables and therefore tests hypotheses. All proposals for analytic research must explicitly state the hypothesis(es) (Example 2.10).

The hypothesis

A hypothesis is a statement (not a question) about an expected relationship between one or more independent variables and one dependent variable. The statement should proceed logically from the prior problem identification. In addition to stating the hypothesis(es), the proposal should also indicate:

- Under what conditions the hypothesis is expected to be true.
- All potential intervening variables that may affect the dependent variable.
- Operational definitions for all variables included in the hypothesis(es).

Example 2.10

Research Hypotheses
Analytic Epidemiology

1. Children who receive a sex-education course before entering their teens are more likely to use contraceptives at first intercourse than children who do not receive such a course.
2. In Country X, Catholic women have higher fertility rates than non-Catholic women.
3. Lower rates of smoking are expected among pregnant women who attend the prenatal clinic in District X than among women who do not attend.

At this point in the proposal, the null hypothesis is not stated. The null hypothesis is appropriate only in the section of the proposal that discusses the analytic methods to be used in the study.

Study Design

Descriptive and analytic research

The study design or research design is determined, in part, by the primary purpose of the research. Therefore, the proposal should first indicate whether the study is descriptive or analytic.

A descriptive study is used when additional information is needed before being able to formulate specific hypotheses. Descriptive studies provide accurate baseline data on the occurrence or prevalence of a characteristic or event related to a health problem, and on the people who are affected and how they are affected.

An analytic, or explanatory, study is used to explain the relationship between two or more variables by testing causal hypotheses that specify the relationship between the variables.

Four study designs used in epidemiologic research

Once the primary purpose of the research is identified, we can select the study design. The type of design we choose is influenced by the purpose, the cost, and the nature of the problem or variables to be studied. Possible designs include:

- Cross-sectional design. Current or retrospective information is collected at one point in time from a sample of subjects from the target population. This design is appropriate for descriptive purposes because it does not always lend itself to temporal arrangement of independent and dependent variables. This type of study is discussed in Chapter 6.

- Experimental design (randomized clinical trials). The researcher manipulates the independent variable or study factor and controls allocation of subjects to the exposure under study. This design is ideal for analytic epidemiologic research; it is discussed in Chapter 8.

- Cohort design. Information is collected on the study population at one point in time. Then, at a later point in time the subjects are examined again to measure the outcome of interest. The temporal aspect of this design makes it the most appropriate nonexperimental design for analytic epidemiologic research. This design is discussed in Chapter 9.

- Case-control design. Typically a retrospective design that compares a group of cases and controls to examine the effect of a current or previous risk factor. The case-control design may also be a prospective design in which cases and controls are enrolled prospectively soon after a health problem is diagnosed or identified. The design is used for both descriptive and analytic research purposes; it is especially useful if the outcome or dependent variable is a rare event. This design is discussed in Chapter 10.

Methods

Epidemiologic research involves a careful and systematic observation of people (subjects) and events. The methods used for such observation affect the quality of the data. Therefore, the investigator must provide a thorough description of the methodology for selecting the subjects and for collecting the data. The content of the proposal's methods section will vary depending on the purpose of the research and study design, but this section should specify the study

population, the type of data to be collected, and the data collection and quality control procedures. The Methods section should present step-by-step instructions for carrying out the research. A brief outline for the Methods section follows:

1. Define the population, including political, geographic, social, economic, and demographic identifiers.

2. Describe the sampling process, if applicable.

    - Identify the type of sample (e.g., simple random, systematic, cluster, multistage, nonprobability).

    - Specify the sample size calculations.

    - Describe the random assignment procedure for clinical trials.

3. Define the type of data to be collected.

    - Define cases, controls, and comparison groups.

    - List all variables (i.e., independent, dependent, control, exposure, treatment, outcome, confounders, effect modifiers) and state the conceptual and operational definitions (Example 2.11).

Example 2.11

Hypothetical Example of Conceptual and Operational Definitions for the Term Contraceptive Use

Current contraceptive use could be conceptually defined as any use of birth control by either the subject or her partner during the last month. Operationally, current contraceptive use could be defined as a subject indicating that she had used within a four-week period one of the contraceptive methods listed on an interview card. The list of methods could include tubal ligation; vasectomy; contraceptive pills; IUD; injection; condom; diaphragm; vaginal tablets, creams or jellies; rhythm; Billings method; and withdrawal.

4. Describe the data collection procedure.

   - Indicate data collection method(s) (e.g., structured or unstructured interview; focus groups; self-administered questionnaire; direct observation of behavior; service statistics; medical chart review; vital records, census data, or other secondary sources).

   - Describe the data collection instrument (e.g., questionnaire, medical records' abstract form, etc.). A first draft of the data collection instrument is typically developed after the research project has been approved.

     The proposal may, however, indicate that individuals with experience in developing and using similar data collection instruments will be asked to critique and improve the draft. Other aids for developing a questionnaire, may also be mentioned (e.g., focus groups for clarifying concepts and terminology). If the study uses a preexisting instrument, a copy may be appended to the proposal.

   - Discuss consent of participants and how it will be obtained. If an Informed Consent form is needed, a copy should be included in the proposal.

   - Discuss confidentiality of the data and, if needed, how it will be maintained.

   - Discuss human subjects review, if applicable.

5. Describe the procedures used to control data quality. The research proposal should include a discussion of any activities that are planned for maximizing the validity and reliability of the data.

   - Pretest the data collection instrument. Field test on a limited basis in an area outside the study area (or without involving the study subjects). All the study procedures should be followed, including sampling, data collection (e.g., interviewing), supervising, coding, data entry, editing,

and a limited analysis. Pretesting is useful to modify the data collection instrument as well as other data collection procedures.

- Reinterview subgroups of respondents. This is a common technique for testing the reliability of the instrument.

- Train interviewers and supervisors for data collection. Initially, data collection teams should be closely observed by a field coordinator or supervisors. Often, data collection does not go as planned and many problems can affect the validity or reliability of the data. When these problems occur, interviewers or other data collectors should consult a supervisor or the field coordinator so that decisions about how to proceed take into consideration the impact of the decision(s) on the entire study.

- Describe plans for data control. Meticulous attention to detail is required on the part of supervisors so that 1) all forms are completed according to the predesignated specifications, 2) errors are corrected, and 3) no forms are lost. The forms should be sent to a central location where they are counted and processed for tabulation.

- Indicate multiple sources of information. Some studies may be structured to allow for more than one source of data. An interview, for example, may be supplemented with a medical record to obtain a medical history. More than one source for the same information provides an excellent opportunity to check the validity of the primary source.

- Describe all other data quality checks. For example, more than one question may be structured to ask the same information of the interviewee. Responses to these questions can be compared for consistency.

Analysis Plan

The analysis answers the research questions

The analysis provides answers to the research questions. All proposals for epidemiologic studies contain plans for analysis. The analysis plan and data collection methods are so interdependent that defining one without the other is difficult. Although the analysis depends on the type of data collected, how the data are collected depends on the type of analysis anticipated. Sample size, for example, is a function of the type of analysis that will be performed. The sampling design is also frequently determined by the analytic needs. The analysis plan should deal with data preparation issues as well as data analysis.

Preparation of the Data

Before the actual analysis, the data must be checked for errors and put into a form that will allow it to be manipulated accurately and efficiently:

Tabulation. Indicate whether the data will be tabulated by hand, computer, or some other method.

Coding. The process of coding translates verbal responses into numerical codes that will facilitate data manipulation. Indicate whether coding is necessary and who will do it. If any of the key variables in the study are obtained with open-ended questions, the need to code the responses to these questions may be mentioned.

Editing or cleaning the data. Editing ensures that no question on a questionnaire is omitted erroneously, that no illegal codes have been used, and that logical inconsistencies in the recorded responses are noted. Data may be edited in the field during the collection phase or in a central office after the fieldwork is completed. Data may be edited by manually reviewing the questionnaires or forms on which responses were originally recorded, by using computer programs that find errors and inconsistencies in the data, or by reviewing tabulations produced by the computer. Computer editing may be structured to check each record as it is entered into the computer (this may be done in the field) or after all the records have been entered into

the computer. The proposal should briefly state how the editing will be carried out.

Analysis of the Data

Any combination of the following may be required.

Variable transformations. These may include:

- Collapsing response categories for a particular variable into broader ones (e.g., age may be recorded by single years but collapsed into five-year age groups for a particular analysis).
- Creating new variables (e.g., create a variable denoting premarital conception by comparing age at marriage with age at first birth).
- Counting the responses to a number of questions (e.g., create a score that indicates the number of correct responses in a set of true and false questions).
- Constructing a scale or index that combines the responses to two or more questions (e.g., create a socioeconomic score using mother and father's education, father's occupation, and family income).
- Creating temporary mathematical transformations by converting original numerically scaled values of a variable to a different scale (e.g., square root, quadratic, or logarithmic) to better meet the assumptions of a particular statistical method.

Descriptive statistics. Descriptive statistics are used to describe data quantitatively. They may be univariate, bivariate, or multivariate:

- Univariate statistics include proportions, percentages, ratios, frequency distributions, and any graphic presentations. Other univariate descriptive statistics measure central tendency (e.g., mean, median, mode), deciles, quartiles, and

measures of dispersion (e.g., range, mean deviation, standard deviation, coefficient of variation).

- Bivariate and multivariate statistics are used to describe the associations between variables. Such statistics are called measures of association and include lambda, gamma, Pearson's correlation coefficients, relative risks, odds ratios, and others.

Even if the research objective is to test a hypothesis or generalize sample characteristics to a target population, a description of some basic quantitative characteristics of the sample may be of interest. More importantly, if a total population is under study instead of a sample, descriptive statistics will be the form used to report the study results.

Inferential statistics. Most epidemiologic studies, and certainly those that are based on population samples, use statistical inference methods, which allow conclusions about a population to be made from results obtained in a sample. If the study is based on a sample, you should use not only descriptive statistics that describe characteristics or associations in the sample but also inferential statistics that estimate the effect of sampling error on the ability to infer population characteristics and associations from sample findings. Measures used in statistical inference include confidence levels, confidence intervals, and tests of statistical significance.

Table shells. A table shell contains all the elements of a data analysis table except the data. Some researchers find constructing table shells useful in planning the data collection instrument and in visualizing how the data will be organized for analysis. Table shells may be included in the proposal or in an appendix.

Plans for Interpretation

Plans to interpret the data should be included in the proposal.

Although data have not been collected or analyzed yet, the literature review and study design provide guidelines and constraints for interpreting the research results. The proposal should describe plans to interpret the results. Considerations include:

- Generalizability. The generalizability of a study is a function of sampling and analysis procedures. The proposal should indicate the target population and any other populations (in time or space) to which the results can be generalized.

- Limitations. No study is flawless. All studies will have some weaknesses, for example, in the sample selection, questionnaire design, measurement, or analysis. The researcher's task is to keep these weaknesses at a minimum, to identify the limitations that do exist, and to inform the reader as to how the limitations preclude generalizability or how the problems may be overcome in future studies.

- Potential contributions. The proposal should also discuss the merits of the study, such as timeliness, public policy implications, contribution to scientific knowledge, and public health contribution.

Plans to Report Research Findings

Indicate how research findings will be disseminated.

The proposal should indicate what reports and other means of disseminating research findings are planned. Any or all of the following are appropriate for disseminating the results of the study.

- Progress reports
- Final report
- Publications
- Seminars, workshops, and conferences
- Discussions with policymakers and program managers

Some of the questions that should be addressed when discussing dissemination of study results are:

- What specific parts of the research or data will be covered?
- At what stage in the study will the results be written, and by whom?
- How much time will be required to prepare the materials?
- Who will receive these materials?

Logistics

Describe resources, personnel, facilities, and budget

The logistics are the resources, personnel, facilities, and budget required for the study. The proposal should indicate the anticipated cost of the study, the source of these funds, and how the funds will be allocated. The discussion about logistics should include:

- A description of the resources and facilities available for the study. For example, computer facilities, secretarial help, office space, library facilities, and vehicles. Indicate whether other institutions will contribute resources and what proportion of the principal investigator's time will be devoted to the study (e.g., 100%, 60%, 20%). Many funding agencies prefer joint research projects and look favorably on proposals that show contributions from the applicant's home institution or other organizations.

- Any anticipated difficulty in obtaining scarce professional skills. Consultants or an advisory committee might be used if this need exists.

- A brief management plan that indicates who will be responsible for the budget, staff, field operations, data processing, analysis, and other components of the project. If several departments or institutions are collaborating on the project, indicate who will have overall responsibility for the project and what would be the roles or contributions of the different departments or institutions.

- A clearly outlined, realistic budget that lists each cost item and its components. For example, the cost of 10 interviewers might be shown as: Interviewers (10 @ $5 per day × 30 days) $1,500. Remember that different funding agencies have different rules for what they can fund in a project. (See World Health Organization, *Preparing a Research Project Proposal: Guidelines and Forms*, pages 15-17).

Arrange the cost items under major headings:

Salaries and benefits. Personnel under this heading might include:

- Project director

- Researchers
- Consultants
- Field supervisors
- Interviewers
- Computer programmer
- Keypunchers and coders
- Clerical staff
- Other

Supplies and equipment. Items under this heading might include:

- Reproducing questionnaires or other forms
- Office supplies
- Telephone costs
- Mailing costs
- Computer time or purchase
- Report printing and distribution

Travel. This is a very limited category and should include only travel necessary to complete the study and to initially distribute the results. These costs might include:

- Vehicle rentals
- Gasoline
- Lodging for interviewers during fieldwork

Miscellaneous costs. Separate cost by year if the study will require more than one year. Also if the study will have a long duration, you may want to include a line item for inflation (A hypothetical proposal budget is shown in Example 2.12).

Work Schedule or Timeline

The steps and their sequence in the entire research process should be outlined. A corresponding calendar should indicate the amount of time each step will require (Example 2.13). The steps might include:

Example 2.12

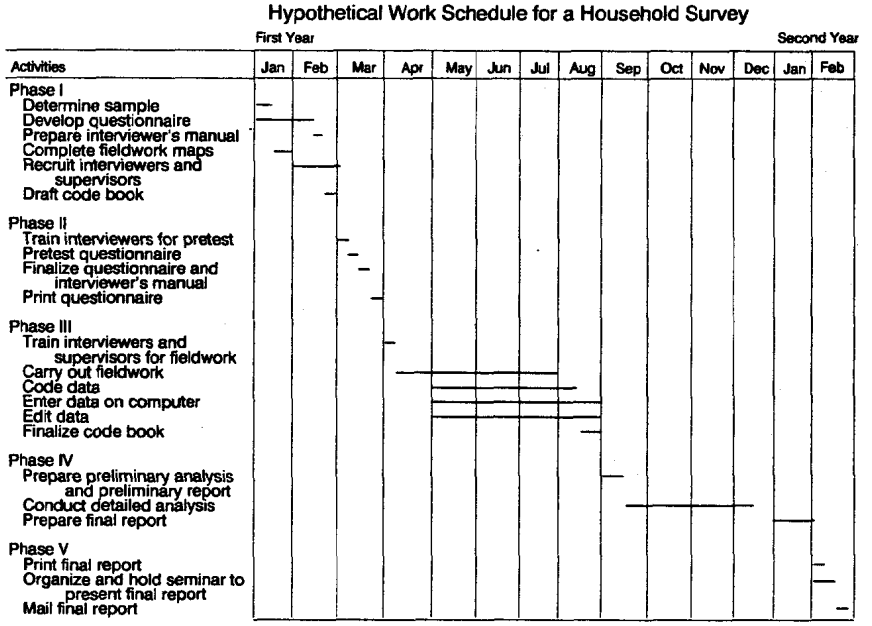
Hypothetical Budget for a Reproductive Health Survey

Personnel Salaries and Wages	PESOS
1 Study Director @ 6 mos. x 141,000/mo.	846,000
1 Statistical Consultant @ 0.5 mo. x 130,000/mo.	65,000
1 Field Coordinator @ 3.5 mos. x 117,500/mo.	411,250
4 Supervisors @ 3 mos. x 75,000/mo.	900,000
16 Interviewers @ 3 mos. x 55,000/mo.	2,640,000
1 Programmer/Analyst @ 2 mos. x 117,000/mo.	234,000
1 Coder @ 2 mos. x 94,000/mo.	188,000
2 Keypunchers @ 2 mos. x 50,000/mo.	200,000
1 Secretary @ 3 mos. x 40,000/mo.	120,000
2 Drivers @ 2.5 mos. x 42,000/mo.	210,000
Subtotal	5,814,250
Supplies and Equipment	
2 IBM-XT computers @ 730,380/ea.	1,460,760
1 IBM wide carriage printer	296,100
Computer paper and supplies	20,000
Printing	
2500 Individual questionnaires	250,000
7000 Household questionnaires	210,000
500 Reports	190,000
Office supplies	90,000
Subtotal	2,516,860
Local Travel	
2 Vehicle rentals @ 10 wks. x 50,000/wk.	1,000,000
Gasoline for 10 wks.	200,000
Subtotal	1,200,000
Miscellaneous	80,000
Subtotal	80,000
Total Project Cost	9,611,110

- Selecting the sample
- Drafting the questionnaire
- Training interviewers and supervisors
- Pretesting the questionnaire
- Revising the questionnaire
- Printing the questionnaires
- Carrying out fieldwork (interviews)
- Coding the data
- Keypunching the data
- Editing the data
- Tabulating the data
- Analyzing the data
- Writing the final report
- Printing the final report
- Presenting the research findings at a conference

Example 2.13

Hypothetical Work Schedule for a Household Survey



Bibliography for the Proposal

The proposal must include a bibliography that contains all the sources cited in the text of the proposal (these citations will be found primarily in the problem identification and justification or in the literature review). Important references that were not cited in the text may also be listed in the bibliography, including methodologic sources. Various styles exist for listing references in a bibliography. Any published journal or book can be used as an example for a bibliographic style (e.g., this manual). The important point is to be consistent and use the same style for each entry in the bibliography. Every reference should be checked against the original publication to ensure correct information.

Appendices to the Proposal

Special documents should be in the appendices.

Many people believe that appendices are *extras* and that they are not read by proposal reviewers. This is not true. Documents such as curriculum vitae and questionnaires receive close scrutiny by reviewers. Therefore, careful attention should be given to the organization and presentation of documents that are not placed in the text of the proposal. These appended documents may include:

- Curriculum vitae of principal investigators
- Information on institutional affiliation of researchers
- Sample of data collection instrument
- Informed consent form
- Letters of endorsement for the study
- Other information relating to the study

Practice Exercises

1. Example 2.14 is part of a research proposal. After reading the example, write an abstract in the spaces provided. Your abstracts should answer these questions: 1) What is the research problem? 2) What is the research question or hypothesis? 3) What are the major expected implications? 4) Who will conduct the study? 5) When? 6) Where? 7) What methods will be used to collect and analyze the data? 8) What resources are required? If the example does not provide the information to answer one of these questions, provide your own hypothetical answer.

Example 2.14

Excerpts from a Proposal for a Descriptive Study of Prenatal Care in Puerto Rico

Goal of the Study

The purpose of this study is to provide a descriptive analysis of the determinants of prenatal care in Puerto Rico by using data collected in the Puerto Rico Fertility and Family Planning Assessment (PRFFPA) survey.

This islandwide survey provides a unique data set to describe receipt of prenatal care in Puerto Rico. The initial descriptive analysis presented in this study should provide an impetus for more specific and in-depth studies concerning the determinants, components, quality, and benefits of prenatal care in Puerto Rico. This study may also offer insight into prenatal care behavior of Puerto Rican women who live in the United States.

Problem Identification

Research findings suggest that further substantial reduction in neonatal mortality is possible only if there is a decrease in the incidence of low-birthweight infants (Institute of Medicine, 1985). The general consensus in the United States and in other industrialized countries is that prenatal care can have a positive impact on an infant's birthweight and thus improve the infant's chance for survival.

However, all women do not receive prenatal care equally. In the United States, factors associated with a lack of appropriate prenatal care include low socioeconomic status, race other than white, unmarried status, very young or very old age at pregnancy, and first or third or higher order of birth.

In Puerto Rico, the infant mortality rate decreased dramatically, from a rate of 93.4 in 1945 to 20.9 in 1975. Between 1975 and 1983, however, infant mortality declined only slightly to a rate of 17.3. This rate was higher than the infant mortality rate for any U.S. state (Rigau-Perez, 1986). Between 1980 and 1983, 9% of all live-born infants in Puerto Rico had low birthweights (Rigau-Perez, 1986; Becerra, 1989). In 1983, 6.8% of all live-born infants in the United States had low birthweights (5.7% of white births, 12.6% of black births) (National Center for Health

Example 2.14 (continued)

Statistics, 1983).

To date, the only information on prenatal care in Puerto Rico has come from birth certificates. In 1980, investigators analyzed the prenatal care information on the birth certificates, namely, the month prenatal care began and the total number of visits (Vazquez and Vazquez, 1982; Vazquez et al., 1983). Results indicate that 99% of women received some form of prenatal care, but only 63% began receiving care in the first trimester, and only 37% had at least ten visits. In comparison, U.S. vital statistics for the same period show that 73% of U.S. women began receiving prenatal care in the first trimester, and 53% had at least ten prenatal visits.

Justification

The PRFFPA, which collected information on many sociodemographic variables, offers a unique opportunity to further define and understand the level of prenatal care and the determinants of prenatal care in Puerto Rico. With more complete information on who receives and who does not receive care, prepared decision makers will be better able to respond to the discrepancies in prenatal care services. A clear description of who lacks care is necessary to effect policy decisions that will close the gaps in receipt of prenatal care.

As services and care become more equitable, we expect to see a lower incidence of low-birthweight infants (especially if care includes services such as identifying high-risk women, advising pregnant women against smoking, and focusing on good nutrition) and thus a lower neonatal mortality rate. A reduction of low-birthweight infants would lower the yearly costs of providing neonatal intensive care. Prenatal care programs can be a cost-effective approach to decreasing the incidence of low birthweight (Institute of Medicine, 1985).

Design

This study will be a descriptive analysis of the determinants and usage of prenatal care in Puerto Rico. It is based on a retrospective cross-sectional survey.

Methods

A. Definition of Terms

- Prenatal Care: At least one visit to a doctor, clinic, or other source for prenatal care during entire pregnancy.
- Late prenatal care: First prenatal visit occurs after the first trimester.
- Inadequate care: Care that begins after the first trimester or less than seven visits to a prenatal care provider.
- Low birthweight: Birthweight that is less than 2,500 g (5.5 pounds), a criterion recommended by the American Academy of Pediatrics in 1935 and adopted by the World Health Organization in 1948. Low-birthweight infants may be premature (gestation less than 37 weeks) or small-for-gestational age.

Example 2.14 (continued)

- Very low birthweight: Birthweight that is less than 1,500 g.
- Parity: The number of live births a woman has had.

B. Population

The population under study will be all women of reproductive age (15 to 49 years old) who have had at least one live birth and who were interviewed during the 1982 PRFFPA.

C. Data Source

This study will use data from the 1982 PRFFPA, a joint collaboration between the University of Puerto Rico, the Puerto Rico Department of Health, and the Division of Reproductive Health, Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control (CDC). Financial support came from the National Institute of Child Health and Human Development, the Bureau of Community Health Services, and the CDC.

1. Sampling and Sample Size. The survey sample was chosen to be representative of the entire island. The sample was a two-stage, disproportionate, stratified cluster; housing units were the sampling units. These units did not include institutional housing, such as prisons, military reservations, and college dormitories. The study subjects were all women aged 15 to 49 years who lived in the selected housing units. The sample size necessary for the study was calculated to be 3,000 women. The survey obtained 3,175 completed interviews.

   The present study will consider only women who have had at least one live birth. Of the 3,175 women interviewed, 2,012 women (63%) had had at least one live birth.

2. Data Collected. The individual questionnaire included basic social, economic, and demographic information and information on fertility, family planning, and maternal and child health. Each individual respondent was asked for complete retrospective histories on fertility, marriage, migration, education and employment, and contraceptive use since 1978.

3. Data Collection. The data collection procedure for the PRFFPA involved two steps: 1) a household questionnaire was completed for each of the 30 housing units within 150 sectors, and 2) an individual questionnaire was completed for all women aged 15 to 49 years in each housing unit. For the household questionnaire, a total of 3,877 completed interviews were obtained from 4,500 units less 418 vacant units (95.0%), and for the individual questionnaire, a total of 3,175 completed interviews were obtained from 3,493 eligible respondents (90.9%).

Example 2.14 (continued)

4. Data Quality. Previous data quality checks for the PRFFPA survey involved: 1) examining the internal consistency of data from the household questionnaire and the individual questionnaire for age and marital status, 2) examining the internal consistency of the marriage and birth histories from the individual questionnaires, and 3) comparing the PRFFPA data, including age, marital status, fertility, and contraceptive use, with data from vital statistics, census, and other surveys.

The proposed study will carry out a validity check for the prenatal care data. Prenatal care data from the PRFFPA survey (specifically for the year 1980) and the results of Vazquez's analysis of birth certificate data for that same year will be compared.

Interpretation

The sampling design of the study will allow the results to be generalizable to all Puerto Rican women aged 15 to 49 years in 1982. The findings may also be generalizable to Puerto Rican women living in the United States because these women move back and forth between the United States and Puerto Rico.

A limitation of this study is that quality of care is unknown. In addition, because the present study will be limited to information from the PRFFPA survey, we will not be able to consider why women chose not to receive prenatal care, chose to receive care late, or chose not to continue care.

This study will be limited to women who have had one live birth. Prenatal care practices of women having other pregnancy outcomes, such as stillbirth or spontaneous abortion, will not be considered in this study. Thus, our results will not be representative of all pregnancies or persons eligible to receive prenatal care.

(Hammett, 1988)

1.1 What is the research problem?

1.2 What is the research question or hypothesis?

1.3 What are the expected implications?

1.4 Who will conduct the study?

1.5 When will the study be conducted?

1.6 Where will the data be collected?

1.7 What methods will be used to collect and analyze the data?

1.8 What resources are required?

2. For each of the following statements, indicate whether a problem is clearly identified. Does the statement include both a *what is* and a *what should be* component. If the statement is an analytic research problem, indicate alternative solutions for the problem and state the preferred solution. For a descriptive research problem, provide the appropriate responses.

   2.1 Several different family planning methods have been distributed through rural clinics. The comparative continuation rates for these methods vary substantially in the Region Z villages. Based on previous research, we would expect no variation in the continuation rates for the methods used in this region. Local health workers believe most of the variation is due to various side effects from the different methods. The Central Health Bureau believes the variation is related to traditional beliefs associated with different forms of contraception. The real reason(s) for the variation should be determined.

   (a) Is this a clear problem statement? Yes/No
   (b) Identify the *what is*:
   (c) Identify the *what should be*:
   (d) Are alternative explanations provided? Yes/No
   (e) Is one explanation identified as the most likely?

2.2 Five women had their IUDs removed last month because of excessive bleeding.

(a) Is this a clear problem statement? Yes/No
(b) Identify the *what is*:
(c) Identify the *what should be*:
(d) Are alternative explanations provided? Yes/No
(e) Is one explanation identified as the most likely? Yes/No

2.3 All hospitals that admit women for complications following induced abortions should accurately note the reason for admission on the medical record. However, because of the legal and ethical problems associated with abortion, information is frequently not recorded correctly.

(a) Is this a clear problem statement? Yes/No
(b) Identify the *what is*:
(c) Identify the *what should be*:
(d) Are alternative explanations provided? Yes/No
(e) Is one explanation identified as the most likely? Yes/No

2.4 Men should not die from infection following a vasectomy.

(a) Is this a clear problem statement? Yes/No
(b) Identify the *what is*:
(c) Identify the *what should be*:
(d) Are alternative explanations provided? Yes/No
(e) Is one explanation identified as the most likely? Yes/No

3. Select three reproductive health problems that are important in your country. Write a problem identification statement for each:

(a) Health problem:

Problem identification statement:

(b) Health problem:

Problem identification statement:

(c) Health problem:

Problem identification statement:

4. Example 2.15 presents a hypothetical problem. Describe the following parts of the problem definition:

Example 2.15

Problem Definition for a Family Planning Study

In Country X, the population growth rate is 3.8% per year. Because of recent economic trends, the average desired family size has fallen from 7.2 children in 1970 to 4.2 in 1980. Until the 1960s, the traditional segments of the population (80% of the total population) practiced postpartum abstinence and breast-feeding for two years after giving birth. Today, fewer than 30% of women with newborn infants breast-feed for as long as 12 months, and 75% resume sexual activity three to five months after delivery. Modern methods of birth control have not yet compensated for the deterioration of traditional child-spacing practices. In 1980, contraceptive practice was estimated to be 12%.

(a) Magnitude

(b) Time frame

(c) Geographic area

(d) Population

(e) Why?

(f) Solutions

(g) Unanswered questions

5. Write a problem definition for each of the three reproductive health problems that you identified as important in your country.

   (a) Health problem:

   Problem definition:

   (b) Health problem:

   Problem definition:

   (c) Health problem:

   Problem definition:

6. Example 2.16 describes a proposed study. Which form(s) of justification does the author use? Circle all justifications you find in the description below.

   (a) Is the problem timely and current?
   (b) Does it have serious consequences?
   (c) Does it affect a large number of people?
   (d) Does it have implications for current programs?
   (e) Are there broad social, economic, or health implications?
   (f) Who is concerned about the problem?
   (g) What advances have already been made toward solving it?

Example 2.16

Justification for a Young Adult Reproductive Health Survey in Chile

Adolescent fertility has become an important concern in Latin America for various reasons, including a decline in age at first pregnancy, an increase in unintended pregnancies, and a relatively high rate of premarital conceptions among young women. This concern is tied to evidence that women who have their first birth at a young age will subsequently have shorter birth intervals, more unplanned pregnancies and illegitimate births, and lower income levels than women who have their first birth at older ages. The health consequences of adolescent pregnancy are a growing concern worldwide. The evidence from surveys confirms that mortality rates are higher for infants and children born to young mothers. For example, a study using World Fertility Survey (WFS) data from 29 countries showed that the infant mortality rate was 33% higher for infants born to mothers less than 20 years old than for infants born to mothers aged 20 to 29 years.

Despite the public health community's awareness of the health risks and social problems associated with early pregnancy and childbearing, representative sample surveys designed to document the attitudes toward sexual activity and sex education, history of sexual experience, and use of contraception among young people in Latin America have been rare. Many investigators have drawn samples of adolescents and young adults from clinic or school populations, groups that are not representative of the total population of young people. To address the need for representative data, Mexico City, Guatemala City, and the island of Jamaica reproductive health surveys using representative samples of young men and women were recently conducted.

In Chile, statistics reveal that the dramatic decline in age-specific fertility rates over the past two decades did not occur among 15 to 19 year olds. For example, women aged 15 to 19 years had fertility rates of 71.8 per 1,000 in 1969; 61.1 in 1979; and 64.1 in 1984. These fairly stable rates contrast sharply with the 24% decline in fertility rates among women aged 25 to 29 years and the 29% decline among women aged 30 to 34 years over the same period. This situation indicates a need for information on the reproductive knowledge and practices of the young adult population in Chile.

Example 2.16 (continued)

A more significant change among young Chilean adults in recent years has been the pronounced increase of births among unmarried adolescent women. In 1970, 44% of the infants born to younger women (<20 years old), were born to unmarried women. By 1985, this percentage had steadily increased to 55%.

To develop programs that address these critical fertility issues in Chile, specific information on sexual activity, reproductive knowledge and attitudes, and contraceptive use among young adults is needed to accurately describe this population.

(Valenzuela, 1988)

7. Using one of the three previously defined reproductive health problems, write a justification for the proposed research. If the exact statistics are not available, substitute an "X" for the real figures.

   (a) Health Problem:

   Justification:

8. Write a public health objective or an ultimate goal for each of the three reproductive health problems.

   (a) Health problem:

   Ultimate goal:

(b) **Health problem:**

Ultimate goal:

(c) **Health problem:**

Ultimate goal:

9. Read the examples below and pick out the key elements that go into the research objectives for each case.

   9.1 From 1993 to 1995, the State University will conduct a case-control study in the northern region to determine whether an association between pelvic inflammatory disease and IUDs exists among women attending private clinics.

   (a) What:
   (b) By whom:
   (c) To whom:
   (d) When:
   (e) Where:
   (f) Purpose:

   9.2 In 1993, a cohort study of 1,000 women will be conducted in Province X to determine if the risk of morbidity is greater among women whose tubal sterilizations were performed by nurse-midwives than among women whose tubal sterilizations were performed by physicians. The study will be conducted by the Ministry of Health with technical assistance from the United Nations Development Program.

   (a) What:
   (b) By whom:
   (c) To whom:
   (d) When:
   (e) Where:
   (f) Purpose:

   9.3 From August to December 1993, a descriptive study will be conducted in Division A by the National Health Council. The study will determine the types of complications that occur among women who had tubal sterilization surgery performed in mobile clinics. A total of 500 women who had sterilization surgery from January to June 1993 in Division A will be included in the study.

   (a) What:
   (b) By whom:
   (c) To whom:
   (d) When:
   (e) Where:
   (f) Purpose:

10. Using one of your reproductive health problems and the ultimate goal stated in Exercise 8, write three different statements of research objectives for the problem.

State reproductive health problem and ultimate goal:

(a) Research objectives:

(b) Research objectives:

(c) Research objectives:

11. For each of the three reproductive health problems selected earlier, decide on the primary purpose of the study and write the research question or hypothesis.

    11.1 Health problem:

        (a) Primary purpose:

        (b) Research question or hypothesis:

    11.2 Health problem:

        (a) Primary purpose:

        (b) Research question or hypothesis:

11.3 Health problem:

(a) Primary purpose:

(b) Research question or hypothesis:

12. From Exercise 11, select one problem that has a research hypothesis (not a question) and indicate:

(a) Conditions:

(b) All potential intervening variables:

(c) Operational definitions of variables:

13. Using your three reproductive health problems, decide which of the four study designs is appropriate for each problem.

    (a) Health problem:

        Type of study design:

    (b) Health problem:

        Type of study design:

    (c) Health problem:

        Type of study design:

14. Circle true (T) or false (F).

(a) T/F Resources already available for the study, such as office space and secretarial help, need to be discussed in the proposal.

(b) T/F Statistical analysis methods need to be outlined in the proposal.

(c) T/F Detailed time lines make a proposal too long.

(d) T/F Proposals should include plans for checking for errors.

(e) T/F Interviewer training, including supervision, should be outlined in the proposal for a survey.

(f) T/F The data collection instrument is usually included in the Methods section of the proposal.

15. Multiple choice. Select one response.

15.1 All of the following are generally included in the Methods section of the proposal except:

(a) Sample selection description
(b) Definition of variables
(c) Hypothesis formulation
(d) Data collection procedure
(e) Quality control procedure

15.2 All of the following are generally included in the Analysis section of the proposal except:

(a) Data preparation plans
(b) Scales, indices, transformations
(c) Descriptive or analytic statistics
(d) Dummy tables
(e) Summary of results

15.3 All of the following are generally included in the Logistics section of the proposal except:

(a) Resources and facilities
(b) Personnel
(c) Materials
(d) Travel
(e) Justification of costs

15.4 All of the following are generally included in the Timeline section of the proposal except:

(a) Problem identification
(b) Pretest questionnaire
(c) Data editing
(d) Data analysis
(e) Write final report

15.5 All of the following are generally included in the Interpretation section of the proposal except:

(a) Generalizability
(b) Limitations
(c) Bibliography
(d) Potential contributions
(e) Public health implications

15.6 All of the following are generally included in the Plans to Report Findings section of the proposal except:

(a) Reports
(b) Publications
(c) Seminars
(d) In-house meetings
(e) Discussions with policymakers

Suggested Answers to Practice Exercises

1. Write a proposal abstract.

   1.1 What is the research problem? Puerto Rico has a higher infant mortality rate than any state in the United States, and a higher proportion of low-birthweight infants are born in Puerto Rico than in the United States. Because prenatal care can have a positive effect on birthweight and infant survival, we need to determine the extent to which prenatal care services are used and the subgroups in the population at greatest risk for inadequate care.

   1.2 What is the research question or hypothesis? What is the level of prenatal care in Puerto Rico and what are the characteristics of women who are at greatest risk for no care or insufficient care?

   1.3 What are the expected implications? Improved levels of prenatal care, which in turn would bring about a reduction in incidence of low-birthweight births.

   1.4 Who will conduct the study? Data collection: University of Puerto Rico, Puerto Rico Health Department, and the Centers for Disease Control. Data Analysis: No information is given in the example, provide your own hypothetical *who*.

   1.5 When will the study be conducted? Data collection: 1982. Data analysis: No information is given in the example, provide your own hypothetical *when*.

   1.6 Where will the data collection be conducted? The island of Puerto Rico.

   1.7 What methods will be used to collect and analyze the data? Cross-sectional, islandwide survey of 3,175 women aged 15 to 49 years who live in the sampled housing units. (Other methods of data collection, to be used as needed, can be found in the Methods section.)

   1.8 What resources are required? Not available in example given. Provide your own hypothetical cost or other resources needed.

2. Identify a problem.

2.1 (a) Yes
(b) Continuation rates for family planning methods vary substantially in Region Z villages.
(c) No variation in continuation rates should exist for the methods used in Region Z.
(d) Yes
(e) No

2.2 (a) No
(b) Five women had their IUDs removed last month because of excessive bleeding.
(c) Not stated
(d) No
(e) No

2.3 (a) Yes
(b) Hospital admissions for complications following induced abortion frequently are not recorded on medical records.
(c) This information should be accurately recorded on medical records.
(d) Legal and ethical problems associated with abortion are preventing this information from being accurately recorded.
(e) No

2.4 (a) No
(b) Not stated
(c) Men should not die from infection following a vasectomy.
(d) No
(e) No

4. Describe parts of the problem definition.

(a) Countrywide high rate of population growth
(b) Current
(c) Country X
(d) Traditional segments of the population are more affected; that is, 80% of total population.
(e) Postpartum abstinence and breast-feeding use to last two years. Today, fewer than 30% of mothers breast-feed newborns for as long as 12

months, and 75% of women resume sexual activity from 3 to 5 months after delivery.

(f) Modern methods of birth control should be substituted for the traditional abstinence and breast-feeding.

(g) Why is the prevalence of contraceptive use only 12%?

6. Justifying a proposed study.

a, b, c, d, e, and f

9. Key elements of research objectives.

9.1 (a) A case-control study
(b) State University
(c) Women attending private clinics
(d) 1993 to 1995
(e) Northern region
(f) To determine whether an association exists between pelvic inflammatory disease and IUDs among women attending private clinics

9.2 (a) A cohort study
(b) Ministry of Health with technical assistance from the United Nations Development Program
(c) 1,000 women obtaining tubal sterilization
(d) 1993
(e) Province X
(f) To determine if the risk of morbidity is greater for women obtaining sterilization performed by nurse-midwives versus sterilization performed by physicians

9.3 (a) A descriptive study
(b) National Health Council
(c) 500 women who obtained a tubal sterilization from mobile clinics in Division A from January to June 1993
(d) August to December 1993
(e) Division A
(f) To determine the type of complications occurring after the sterilization procedure

14. True or false.

    (a) T Logistics

    (b) T Analysis Plan

    (c) F Work Schedule or Timeline. No, Example 2.13 shows that a detailed timeline can be put on one page.

    (d) T Analysis Plan

    (e) T Methods

    (f) F Methods. No, the data collection instrument is not included in the body of the proposal, but may be placed in an Appendix.

15. Multiple choice.

    15.1 (c) Methods

    15.2 (e) Analysis Plan. The proposal is the plan to do a study, and therefore there are no results yet.

    15.3 (e) Logistics

    15.4 (a) Work Schedule or Timeline. Problem identification takes place before writing the proposal and is included in the proposal. The time line outlines steps that will be taken to complete the study.

    15.5 (c) Interpretation

    15.6 (d) Plans to Report Research Findings

References

Becerra J, DeSaliceti NP, Smith JC. Evaluation of a regionalized perinatal care system through linked infant birth and death certificates-Puerto Rico: 1980-1984. PR Health Sci J 1989;8(2):253-258.

Blalock, HM. Social statistics. New York: McGraw-Hill, 1972.

Burkman RT, The Women's Health Study. Association between intrauterine devices and pelvic inflammatory disease. Obstet Gynecol 1981;57:269-276.

Cooper HM. Integrating research: a guide for literature reviews. 2nd ed. Newbury Park, California: Sage Publications, 1989.

Faulkner WL, Ory HW. Intrauterine devices and acute pelvic inflammatory disease. JAMA 1976;235(17):1851-1853.

Fisher A, Laing J, Stoeckel J. Handbook for family planning operations research design. New York: The Population Council, 1983.

Fowler FJ. Survey research methods. Newbury Park, California: Sage Publications, 1988.

Hammett T. [Unpublished thesis for a Master of Public Health Program]. Prenatal care in Puerto Rico (1979-1982) as determined from the Puerto Rico Fertility and Family Planning Assessment of 1982. [1988]. Located at: Emory University, Health Sciences Center Library, Atlanta, Georgia.

Hogue CJR, Cates W, Tietze C. The effects of induced abortion on subsequent reproduction. Epidemiol Rev 1982;4:66-94.

Institute of Medicine, Committee to study the prevention of low birthweight. Preventing low birthweight. Washington, DC: National Academy Press, 1985.

Isaac S, Michael WB. Handbook in research and evaluation. San Diego: Edits Publishers, 1981.

Kaufman DW, Shapiro S, Rosenberg L, et al. Intrauterine contraceptive device use and pelvic inflammatory disease. AM J Obstet Gynecol 1980;136:159-162.

Kiritz NJ. Program planning and proposal writing. Los Angeles: The Grantsmanship Center, 1979.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, California: Wadsworth, Inc., 1982.

Locke LF, Spirduso WW, Silverman SJ. Proposals that work: a guide for planning dissertations and grant proposals. Newbury Park, California: Sage Publications, Inc., 1988.

National Center for Health Statistics. Vital statistics of the United States, I, natality. Washington, D.C.: DHHS Publications No. (PHS) 87-1113, U.S. Government Printing Office, 1983.

Omran AR. Community medicine in developing countries: theory and practice 1972. New York: Spring Publishing Co., 1974.

Rigau-Perez JG. Los objetivos de salud para Estados Unidos en 1990 y su aplicación a Puerto Rico. La salud durante el embarazo y el primer año de vida en Puerto Rico: progreso hacia los objetivos nacionales de salud para 1990 (VII). Boletín Asoc Med PR 78(6):244-252.

Selltiz C, Wrightsman LS, Cook SW. Research methods in social relations. New York: Holt, Rinehart and Winston, 1976.

Valenzuela MS. [Unpublished protocol]. Young adult reproductive health survey: for the urban area of the metropolitan region of Santiago, Chile. [1988]. Located at: University of Chile, Western Campus-Faculty of Medicine, Department of Public Health, Santiago, Chile.

Vazquez JL, Vazquez M. El cuidado prenatal en Puerto Rico. Rev Salud Publica PR 1982-1983(4):9-14.

Vazquez JL, Velasco E, Morales Z, Ramirez de Jarellano AB. Cesarean childbirth in Puerto Rico: a world record. PR Health Sci Journal 1983;2(2):59-64.

Vessey MP, Yeates D, Flavel R, McPherson K. Pelvic inflammatory disease and the intrauterine device: findings in a large cohort study. Br Med J 1981;282:855-857.

World Health Organization. Preparing a research project proposal: guidelines and forms. Geneva: Special Programme of Research, Development and Research Training in Human Reproduction, WHO, 1990.

Learning Objectives - Chapter 3

After completing this chapter you should be able to:

1. Identify the following terms:

   rate
   proportion
   case-fatality rate
   ratio
   odds
   proportional mortality ratio
   point prevalence rate
   incidence rate
   incidence density
   standardized mortality ratio

2. Identify reasons for adjusting rates.

3. Interpret adjusted rates.

4. Identify measures of reproductive health:

   crude birthrate
   general fertility rate
   general pregnancy rate
   age-specific birthrate
   total fertility rate
   induced termination of pregnancy ratio
   induced termination of pregnancy rate I
   induced termination of pregnancy ratio II
   maternal mortality ratio
   fetal death rate
   fetal death ratio
   perinatal mortality rate
   perinatal mortality ratio
   infant mortality rate
   neonatal mortality rate
   postneonatal mortality rate

3 Measures of Disease Frequency in Reproductive Health

Introduction

To understand the occurrence and pattern of reproductive health outcomes in populations, epidemiologists must use statistical and epidemiologic methods to compare disease frequencies among individuals and populations. Standardized definitions of health outcomes are necessary to make comparisons across time and for specified periods of time. This chapter introduces quantitative measures used to assess reproductive health outcomes.

To measure the amount and distribution of diseases, health outcomes, or deaths in a population, investigators need to relate individuals with a specified disease or outcome to a population base. Investigators must also know the size and source (e.g., hospital patients, community sample, census tract) of the population from which the cases are drawn and the time period in which the information was collected. When these components are combined as a rate or ratio, we can compare disease frequencies in two or more groups of individuals.

Rates and Ratios

Rates

Rate defined

A rate is a measure of the frequency of some event in a defined population. A rate is composed of a numerator (number of events); a denominator (average population at risk for the event); a specified time in which events occur; and a multiplier (10^n ; $n \geq 1$), which converts a fraction or decimal to a whole number.

Proportions

Most rates are proportions—the numerator is a subset of the denominator. For instance, the case fatality rate is the number of deaths associated with a health problem divided by the number of

persons diagnosed with the problem during a specified time period. The result is a fraction that is usually expressed per hundreds, per thousands, etc.

(3.1.1)

$$\text{Case Fatality Rate}^{1} = \frac{\text{Number of deaths related to health problem}}{\text{Number of persons diagnosed with health problem}} * 10^{n}$$

The numerator and denominator should reflect a similar population. If the numerator is confined to a certain age, sex, or racial group, the denominator should be similarly restricted (Example 3.1).

Example 3.1

Rate of Oral Contraceptive Use Among Mexican-American Women in the United States-Mexico Border Survey, 1979

To determine the rate of oral contraceptive (OC) use among Mexican-American and non-Hispanic white women, we must calculate the rate of use for a defined population and time period. In this example, the population is the U.S.-Mexican border and the time period is the calendar year 1979.

Research Question

What proportion of Mexican-American and non-Hispanic white women in the U.S.-Mexico Border Survey used OCs in 1979?

$$\frac{\text{575 Mexican-American women used OCs}}{\text{1,255 Mexican-American women in the sample}} * 100 = 45.8\%$$

$$\frac{\text{482 non-Hispanic white women used OCs}}{\text{798 non-Hispanic white women in the sample}} * 100 = 60.4\%$$

(Smith et al., 1983)

Ratios

Ratio defined

A ratio is an expression of the relationship between a numerator and a denominator regardless of the population base from which the numerator and denominator are derived. The numerator of a ratio is not necessarily a subset of the denominator. For example, the maternal mortality ratio

[1] The asterisk (*) that appears in the formulas throughout this manual denotes multiplication.

is calculated by dividing the number of maternal deaths regardless of pregnancy outcome (i.e., ectopic pregnancy, fetal death, live birth) by the number of live births in a defined population. This measure of disease frequency is a ratio and not a rate because the numerator includes maternal deaths that are not a subset of the denominator. Ideally, the maternal mortality rate should include all pregnancies (not just live births) in the denominator. In practice, measuring all pregnancies in a population is difficult, so the maternal mortality ratio is used as a proxy for the more accurate maternal mortality rate. Unfortunately, most textbooks refer to the maternal mortality ratio as the maternal mortality rate.

Maternal mortality rate

A ratio is simply the value obtained by dividing one quantity by another. Therefore, a rate is also a ratio but a ratio is not always a rate. For example, a frequently cited ratio is the proportion of persons with disease relative to the proportion of persons without disease. This ratio is called the odds of disease:

Odds of disease ratio

(3.1.2)

$$\text{Odds of Disease} = \frac{\text{Proportion of persons with the health problem}}{\text{Proportion of persons without the health problem}}$$

Another ratio is the proportion of persons exposed to a risk factor relative to the proportion of unexposed persons. This ratio is called the odds of exposure. Because the numerator is not a subset of the denominator, we cannot call an odds ratio a rate.

Odds of exposure

Proportional mortality (PM) measures the relative importance of a specific cause of death relative to all deaths in a population (3.1.3). It is an estimate of the proportion of lives that may be saved by preventing a given cause of death.

Proportional mortality

(3.1.3)

$$\text{PM} = \frac{\text{Number of deaths from a specific cause within a specified time period}}{\text{Total number of deaths within the specified time period}} * 10^n$$

The proportional mortality ratio (PMR) is the ratio of two PMs. For instance, the proportional mortality of a given cause of death may be compared in two different populations (Example 3.2).

Proportional mortality ratio (PMR)

However, the PMR can be misleading. In this example, the PMR of 2.4 could have resulted because France had a higher infant mortality rate from birth defects than did Mauritius.

Example 3.2

Proportional Mortality Ratio of French and Mauritian Infants Who Died of Birth Defects

Of the 6,257 native French infants who died in 1986, 1,303 had birth defects. The proportional mortality for the infants with birth defects is 20.8% (1,303/6,257 = 0.208 * 100).

Of the 463 native Mauritian infants who died in 1987, 40 had birth defects. The proportional mortality for the infants was 8.6% (40/463 = 0.086 * 100).

The proportional mortality ratio (PMR) is calculated as:

$$PMR = \frac{20.8}{8.6} = 2.4$$

(World Health Organization, 1988)

However, a lower infant mortality in France would have produced the same result, particularly among infants who died of causes other than birth defects. In fact, in France only 80 out of every 10,000 live-born infants died within the first year of life, and 17 of every 10,000 live-born infants died from birth defects. In Mauritius, 242 of every 10,000 live-born infants died within the first year of life and 21 of every 10,000 died from birth defects. Because the denominator of the PM does not reflect a population at risk, we cannot describe the risks of the individuals in the population who are dying from a health problem.

Prevalence and Incidence Rates

Point prevalence rate (PR)

The measures of disease most frequently used in epidemiology are prevalence and incidence. The point prevalence rate (PR) is the proportion of the population that has the health problem under study.

(3.2.1)

$$PR = \frac{\text{Number of existing cases of the health problem at a specified point in time}}{\text{Total population}} * 10^n$$

We include all persons who have the health problem at a specified point in time in the numerator, regardless of the length of time the individuals have had the problem. The denominator includes the total population—all persons diagnosed with the health problem and all persons unaffected by the health problem (Example 3.3).

Example 3.3

Prevalence of Smoking Among Women in Puerto Rico in 1982

In the 1982 Puerto Rico Fertility and Family Planning Assessment, women 15 to 49 years old were asked, "Do you currently smoke?" The point prevalence of smoking was:

$$PR = \frac{\text{Women currently smoking at time of interview}}{\text{Total women in the sample}} = \frac{487}{3{,}175} = 15.3\%$$

(Becerra et al., 1988)

Incidence rate (IR)

In contrast, the incidence rate (IR) is the number of new cases of a health problem that occurs in a population *at risk* within a specified period of time (Example 3.4).

(3.3.1)

$$IR = \frac{\text{Number of new cases of disease during a specified period of time}}{\text{Population at risk}} * 10^n$$

Cumulative incidence rate

The incidence rate shown in Example 3.4 is sometimes referred to as a cumulative incidence rate. This rate is based on the assumption that the entire population at risk is followed from the beginning of the observation period to the end of the observation period.

However, we may not always be able to record observations for the specified period of time. For example, participants may enter a study at different points in time, or some participants may not be available for follow-up. Consequently, the duration of observations will not be uniform for all participants. When the observation periods are not uniform, we must use a more precise estimate of the incidence of a disease in a population. This

Example 3.4

Eight-Year Incidence of Vasectomy in the Framingham Offspring Study

During an eight-year follow-up period, forty-four men who participated in the Framingham Offspring Study had had a vasectomy.

$$IR = \frac{\text{Number of men who had had a vasectomy}}{\text{Number of men at risk for vasectomy}} = \frac{44}{397} = 11.1\%$$

(Hubert et al., 1987)

Incidence density (ID)

estimate, the incidence density (ID), is calculated by using the exact amount of time each study participant is followed; it is considered the instantaneous rate at which a disease develops in a population:

(3.4.1)

$$ID = \frac{\text{Number of new cases of disease during a given period of time}}{\text{Total person - time of observation}} * 10^n$$

For the ID and the IR, the numerator is the number of new cases of disease in a population during a given period of time. The denominator for the ID, however, is the sum of the total time that each individual contributes to the observation period and remains disease-free. The units in the denominator must reflect time (i.e., person-days, person-months, person-years) (Figure 3.5).

In calculating an incidence rate, we must precisely define the denominator. The denominator should include only those people who are *at risk* of developing the health problem. Therefore, people who have the health problem or who have had the problem, as well as people who cannot develop the health problem, should not be included in the denominator. Sometimes we may have difficulty determining who should not be included in the denominator, but the magnitude of the problem depends on the proportion of the total population not at risk. If persons not at risk are included in the denominator, the incidence rate will underrepresent the true incidence of the health problem. In some situations, we must remove people who are not at risk because they make up a large proportion of the population. For example, we should remove women who have

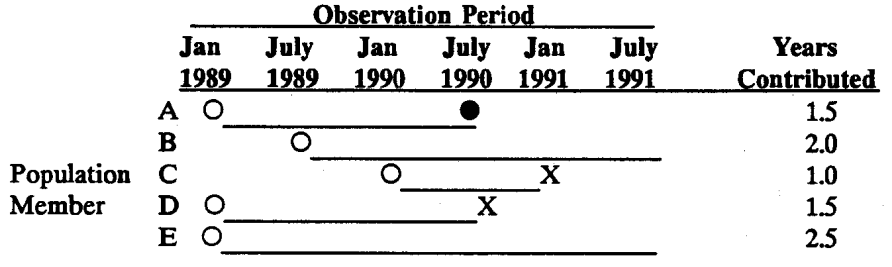
had a hysterectomy from the denominator when calculating incidence rates for endometrial cancer because these women are not at risk for this type of cancer (Nolan et al., 1982; Pokras, 1989).

In Figure 3.5, the total amount of time contributed by all participants is the sum of all person-years, or 8.5 years. Two occurrences of the health problem occurred during the follow-up period. Therefore, the incidence density is:

$$ID = \frac{2 \text{ cases}}{8.5 \text{ person-years}} * 100 = 24 \text{ per } 100 \text{ person-years}$$

Figure 3.5

Incidence Density



○ Initiation of observation
● Loss-to-follow-up: study subject died from causes other than health problem or moved out of study area.
X Disease of interest detected

An underlying assumption in calculating the ID is that the risk of developing the health problem is constant over time across age groups. Because this may not be true, we must divide the time period into smaller segments and stratify the population by age. By dividing and stratifying the population, our calculation may better reflect the ID over time and age. Certain age groups may also be more prone to being lost to follow-up. If loss to follow-up occurs equally among the various exposure groups, our bias problems will be minimized. However, because this is unlikely, we must carefully inspect the ID by time and age to assess the extent of the differences between our study groups.

Loss to follow-up

Adjusted Rates

Crude mortality rate

If the distribution of characteristics in a population (i.e., categorization by age or sex) and the mortality rates specific for each one of the characteristic categories (i.e., age-specific mortality rates) are known, we can always express the overall, or crude mortality rate as:

(3.5.1)

$$\sum_{i=1}^{I} c_i * r_i = \frac{\text{Total number of deaths}}{\text{Total population at risk}}$$

where c_i = the proportion of the population in each category, r_i = category-specific mortality rates, and I = number of categories or strata.

When populations are different from each other (i.e., by age or race distribution), we should not use crude rates to compare them. Rather, we must *adjust* or standardize the rate to remove the effect of differences between populations for comparison purposes. Age is the variable most often used for adjustment because of its effect on morbidity and mortality, although we can use other variables.

Direct and indirect methods

Standard population

Basically, two methods are used to adjust for differences in demographic composition of two populations: the direct method and the indirect method (Mausner and Bahn, 1974). Regardless of the method used, we must identify a standard population. The standard population may be one of the populations being compared, a combination of the comparison populations, a different population altogether, or a theoretical population. Because we use the process of standardization to remove the effect of differences between populations for comparison purposes, the standard population we select is not important.

Direct method. Using this method of adjustment, we multiply the proportions of the standard population in each category (i.e., age groups) by the corresponding age-specific mortality rates of each of the two comparison populations. These products are added up for each population.

(3.5.2)

$$\text{Direct Adjusted Rate} = \sum_{i=1}^{I} C_i * r_i$$

where C_i = the proportion of the standard population in each category, r_i = category-specific mortality rates in the study population, and I = number of strata.

Using this method, we obtain rates that would be expected in the standard population if the age-specific rates prevailed for each of the comparison populations. The directly adjusted rate is meaningless by itself because its absolute value will always depend on the standard population; it is useful, however, in comparing another similarly adjusted rate using the same standard population. In essence, the two comparison populations are similarly weighted by the adjusted characteristic (age).

In Example 3.6, the direct method was used to correct or adjust for differences in the age distribution of the two comparison populations, Argentina and Mexico. If only the crude rates had been cited, researchers would have concluded that Argentina (9.2 deaths per 1,000) had a higher death rate than Mexico (6.4 deaths per 1,000). However, after adjusting the rates, we are presented with a different picture: Mexico has a higher death rate (7.3 deaths per 1,000) than Argentina (6.6 deaths per 1,000). The difference is because Mexico has a younger population than Argentina and age-specific death rates are much lower in younger people than in older people. Therefore, when differences in the age distribution are considered, researchers would have concluded that Mexico had a slightly higher death rate than Argentina.

Indirect method. For this adjustment procedure, we multiply the category-specific mortality rates of the standard population by the corresponding number of individuals in each category in the study population. These products are then added up for each population. The result is usually expressed as a ratio of the observed number of deaths in a study population to the expected number of deaths if the study population had the category-specific mortality rates of the standard population. This ratio is called the standardized mortality ratio (SMR).

Standardized mortality ratio (SMR)

Example 3.6

Age-Adjusted Death Rates Using the Direct Method
Argentina and Mexico, 1982

Age in years	Argentina* Death % (a)	Argentina* Death Rate‡ (b)	Mexico* Death % (c)	Mexico* Death Rate‡ (d)	Standard† % (e)	Argentina Expected Deaths‡ [b*e/100]	Mexico Expected Deaths‡ [d*e/100]
< 1	2.4	33.2	3.3	38.0	3.0	1.01	1.15
01-04	9.3	1.5	13.2	2.5	12.1	0.18	0.30
05-14	18.9	0.5	28.9	0.8	26.1	0.14	0.20
15-24	16.6	1.2	20.4	2.3	19.3	0.22	0.45
25-34	15.0	1.7	13.2	3.8	13.7	0.23	0.52
35-44	11.9	3.7	8.4	6.1	9.4	0.35	0.57
45-54	10.6	9.3	5.7	9.8	7.1	0.66	0.70
55-64	8.0	20.4	3.7	17.7	4.9	1.00	0.87
≥ 65	7.2	64.6	3.2	58.0	4.4	2.82	2.52
Total§	100.0	(9.2)‖	100.0	(6.4)‖	100.0	6.60	7.28
						Adjusted Rates	

* Males only; midyear population estimates (Argentina = 14,501,000; Mexico = 36,647,000).
† Combined Argentina and Mexico population.
‡ Per 1,000 population.
§ Totals may not exactly add up due to rounding error.
‖ Average mortality weighted by the proportion of the population in each age category (crude death rates).

(Adapted from Pan American Health Organization, 1989)

(3.6.1)

$$\text{Standardized Mortality Ratio} = \frac{\text{Observed number of deaths}}{\text{Expected number of deaths}} * 100$$

Suppose we want to compare the survival of infants in different hospitals who weigh less than 2,500 grams with the national average (the standard population). Because the survival of a newborn weighing 1,000 grams is dramatically different from the survival of a newborn weighing 2,000 grams and because hospitals have different

referral patterns for high-risk newborns, we can adjust for birthweight. In this case, the number of births and deaths per birthweight category (usually in 500 gram increments) in a given hospital will probably be smaller than the aggregate national figure. Consequently, if we were to use the direct method, our estimates of birthweight-specific mortality for each hospital would be unstable and unreliable. The alternative is to use the indirect method to estimate the ratio of the observed number of deaths in a study hospital to the expected number of deaths if the hospital population had the birthweight-specific mortality of the national average.

In Example 3.7, Hospital A has the highest neonatal mortality in the country among newborns who weighed less than 2,500 grams. However, if we adjust for birthweight, we may reach a different conclusion. For example, if Hospital A had the same neonatal mortality experience for each birthweight category as the national average, we would expect to see 275 neonatal deaths, or 75 more deaths than the 200 that actually occurred. Therefore, after we adjust for birthweight, Hospital A has better neonatal survival than the national average.

Example 3.7
Indirect Standardization of Neonatal Mortality

In 1985, 1,000 infants who weighed less than 2,500 grams were delivered in Hospital A. The hospital reported that 200 of these infants died. The national neonatal mortality average for infants weighing less than 2,500 grams was 126 deaths per 1,000 live births. Hospital A was identified as having the highest neonatal mortality in the country. A study will be conducted to investigate this matter.

Birthweight	Live Births in Hospital A (a)	National Mortality (b)	Expected Deaths [a*b]	Observed Deaths
500- 999 g	300	0.750	225	168
1000-1499 g	250	0.125	31	20
1500-1999 g	250	0.060	15	10
2000-2500 g	200	0.020	4	2
Total	1000		275	200

$$\text{SMR} = \frac{200}{275} = 72.7\%$$

Reproductive Health Rates and Ratios

WHO definitions

The definitions and formulas that we present in this chapter were adopted from the World Health Organization (WHO) International Classification of Diseases (ICD-9) definitions used in perinatal statistics (Health Care Financing Administration, 1980; Chiswick, 1986) and adapted, reviewed, and approved by representatives of the American Academy of Pediatrics (AAP), the American College of Obstetricians and Gynecologists (ACOG), the American Medical Association, the American Medical Record Association, and the Centers for Disease Control (1988). Adherence to standard and commonly agreed upon definitions is necessary to make meaningful comparisons. Recently the WHO finalized the ICD-10 definitions (see Appendix 2).

Live birth

The definition of a live birth is *the complete expulsion or extraction from the mother of a product of human conception, irrespective of the duration of pregnancy which, after such expulsion or extraction, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles whether or not the umbilical cord has been cut or the placenta is attached.*

Fetal death

A fetal death is defined as *death prior to the complete expulsion or extraction from the mother of a product of human conception, fetus and placenta, irrespective of the duration of pregnancy; the death is indicated by the fact that, after such expulsion or extraction, the fetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles. Heartbeats are to be distinguished from transient cardiac contractions; respirations are to be distinguished from fleeting respiratory efforts or gasps. This definition excludes induced terminations of pregnancy.*

Induced termination of pregnancy

Induced termination of pregnancy is defined as *the purposeful interruption of an intrauterine pregnancy with the intention other than to produce a liveborn infant, and which does not result in a live birth. This definition excludes management of prolonged retention of products of conception following fetal death.*

Live Birth Measures

The crude birthrate measures the impact of fertility on population growth by relating the total number of births to the total population (usually approximated by the midyear estimate) in a given year.

(3.7.1)

$$\text{Crude Birthrate} = \frac{\text{Number of live births to women of all ages during a calendar year}}{\text{Total estimated midyear population}} * 1{,}000$$

The general fertility rate is also a measure of fertility that relates the number of births to women aged 15 through 44 years (childbearing ages). For this measure, the denominator is the population-at-risk.

(3.7.2)

$$\text{General Fertility Rate} = \frac{\text{Number of live births to women of all ages during a calendar year}}{\text{Estimated midyear population of women aged 15 through 44 years}} * 1{,}000$$

The general pregnancy rate is more inclusive than the general fertility rate. In addition to live births, this rate includes fetal deaths and induced terminations of pregnancy in the numerator.

(3.7.3)

$$\text{General Pregnancy Rate} = \frac{\text{Number of live births + number of fetal deaths + number of induced terminations of pregnancy during a calendar year}}{\text{Estimated midyear population of women aged 15 through 44 years}} * 1{,}000$$

Age-specific birthrate and the total fertility rate are rates that relate the number of births to women in a specific age group to the total number of women in that age group.

(3.7.4)

$$\text{Age - Specific Birthrate} = \frac{\text{Number of live births to women within a specific age group during a calendar year}}{\text{Estimated midyear population of women within the same age - specific group}} * 1{,}000$$

(3.7.5)

Total Fertility Rate = The sum of the age-specific birthrates of women at each age group 10-14 through 45-49 years. Because 5-year age groups are used, we multiply the sum by 5. This rate can also be calculated using single years of age.

= Sum of the Age-Specific Birthrates * 5

Maternal Mortality Definitions and Measures

Maternal death

These measures are designed to indicate the likelihood that a pregnant woman will die from complications of pregnancy, childbirth, or the puerperium period. Maternal death is defined as *the death of a woman from any cause related to, or aggravated by, pregnancy or its management (regardless of duration or site of pregnancy).* This definition does not include accidental or incidental causes.

A death that occurs during pregnancy or after a pregnancy is terminated from causes unrelated to the pregnancy, including complications and management of the condition, is not considered a maternal death. Nonmaternal deaths may result from accidental causes, such as an automobile accident or gunshot wound, or incidental causes, such as a concurrent malignancy.

Direct obstetric death. The death of a woman from obstetric complications of pregnancy, labor, or the puerperium; from interventions, omissions, or treatment; or from a chain of events resulting from any of these factors.

Indirect obstetric death. The death of a woman from a previously existing disease or a disease that develops during pregnancy, labor, or the puerperium. The disease(s) are not from direct obstetric causes, although the physiologic effects of pregnancy were partially responsible for the death.

Theoretically, the population at risk (the denominator) should include all pregnant women in a given time period. Because it is difficult to ascertain fetal deaths and pregnancies that are induced terminations, we usually use the number of live births in the denominator.

(3.7.6)

$$\text{Maternal Mortality Ratio} = \frac{\text{Number of deaths attributed to maternal conditions during a specified time period}}{\text{Number of live births during the same time period}} * 100{,}000$$

WHO currently recommends including maternal deaths that occur within 42 days of the end of the pregnancy. Some countries use other time periods (i.e., within one year) but are urged to use the WHO and the national definitions.

Although international comparisons are difficult to make because of variable reporting practices, we know that wide differences in maternal mortality exist worldwide (World Health Organization, 1985) (Example 3.8).

Example 3.8

Maternal Mortality Ratios in the World, 1985

In 1985, more than 500,000 maternal deaths occurred worldwide. Approximately 99% of these deaths occurred in developing countries. In 1985, the maternal mortality ratio averaged 640 per 100,000 live births in Africa, 420 in Asia, and 270 in Latin America. In 1985, the U.S. maternal mortality ratio was 7.8 per 100,000 live births.

(World Health Organization, 1985)

Induced Termination of Pregnancy Measures

Induced termination of pregnancy ratio I. Like the maternal mortality ratio, the population at risk for the measures of induced termination of pregnancy can be estimated by the number of live births in a time period as a surrogate measure of all pregnancies. Because this is not the actual total population at risk, the measure is considered a ratio.

(3.8.1)

$$\text{Induced Termination of Pregnancy Ratio I} = \frac{\text{Number of induced terminations occurring during a specified time period}}{\text{Number of live births occurring during the same time period}} * 1{,}000$$

Induced termination of pregnancy rate. This measure uses live births, induced terminations of pregnancy, and fetal deaths in the denominator.

(3.8.2)

$$\text{Induced Termination of Pregnancy Rate} = \frac{\text{Number of induced terminations occurring during a specified time period}}{\text{Number of induced terminations + live births + reported fetal deaths during the same time period}} * 1{,}000$$

Induced termination of pregnancy ratio II. This is the probability that women of reproductive age will have an induced termination of pregnancy.

(3.8.3)

$$\text{Induced Termination of Pregnancy Ratio II} = \frac{\text{Number of induced terminations occurring during a specified time period}}{\text{Female population aged 15 through 44 years}} * 1{,}000$$

Fetal Mortality Measures

Fetal death measures *indicate the likelihood that pregnancies in a population group would result in fetal death. . . . The population at risk for fetal mortality is the number of live births plus the number of fetal deaths in a year.* To obtain the fetal death ratio, we can use the number of live births as the population at risk. Because birthweight can be more accurately measured than gestational age, we recommend that this measurement include specific birthweight (e.g., fetal deaths of at least 1,000 g).

(3.8.4)

$$\text{Fetal Death Rate} = \frac{\text{Number of fetal deaths during a specified time period}}{\text{Number of fetal deaths + number of live births during same time period}} * 1{,}000$$

(3.8.5)

$$\text{Fetal Death Ratio} = \frac{\text{Number of fetal deaths during a specified time period}}{\text{Number of live births during same time period}} * 1{,}000$$

Perinatal Mortality Measures

These measures *combine fetal deaths and live births which survive only briefly (up to a few days or weeks) on the assumption that similar factors are associated with the losses. The population-at-risk is the total number of live births plus fetal deaths, or the number of live births alone. Perinatal mortality measures can vary by age of fetus and the infant.* In general, perinatal mortality measures should be reported by specific birthweight rather than gestational age.

(3.8.6)

$$\text{Perinatal Mortality Rate} = \frac{\text{Number of deaths of infants less than \_\_ days old + number of fetal deaths (birthweight at least \_\_ g) during a specified time period}}{\text{Number of live births + number of fetal deaths (birthweight at least \_\_ g) during same time period}} * 1{,}000$$

(3.8.7)

$$\text{Perinatal Mortality Ratio} = \frac{\text{Number of deaths of infants less than \_\_ days old + number of fetal deaths (birthweight at least \_\_ g) during a specified time period}}{\text{Number of live births during same time period}} * 1{,}000$$

When gestational age is used to calculate perinatal mortality rates and ratios, instead of birthweight, we generally use the following accepted categorizations:

(3.8.8)

Perinatal Period I = Number of deaths of infants less than seven days old; gestation of 28 weeks or more.

(3.8.9)

Perinatal Period II = Number of deaths of infants less than 28 days old; gestation of 20 weeks or more.

(3.8.10)

Perinatal Period III = Number of deaths of infants less than seven days old; gestation of 20 weeks or more.

Infant Mortality Measures

Infant deaths

Measures of infant mortality are intended *to show the likelihood that live births with certain characteristics will die during the first year of life.* In general, the population at risk are live births that occur during the time period specified for the study. Infant deaths include *any death at any time from birth up to, but not including, one year of age (364 days, 23 hours, 59 minutes from the moment of birth)* (Example 3.9).

(3.9.1)

$$\text{Infant Mortality Rate} = \frac{\text{Number of infant deaths (neonatal + postneonatal) during a specified time period}}{\text{Number of live births during the same time period}} * 1{,}000$$

(3.9.2)

Neonatal Period = Birth through 27 days, 23 hours, 59 minutes.

(3.9.3)

Postneonatal Period = End of 28th day through 364 days.

(3.9.4)

$$\text{Neonatal Mortality Rate} = \frac{\text{Number of neonatal deaths during a specified time period}}{\text{Number of live births during the same time period}} * 1{,}000$$

(3.9.5)

$$\text{Postneonatal Mortality Rate} = \frac{\text{Number of postneonatal deaths during a specified time period}}{\text{Number of live births during the same time period}} * 1{,}000$$

Example 3.9

Infant Mortality Rates, Worldwide, 1985

In 1985, approximately 40,000 infants died *each day* worldwide. In the United States, 40,000 infants died throughout the same year. Clearly, substantial disparities in infant mortality rates exist among countries. In 1985, Afghanistan, Mali, Sierra Leone, Malawi, Guinea, Ethiopia, Somalia, Mozambique, Burkina Faso, and Angola had infant mortality rates higher than 140 per 1,000 live births. On the other hand, Sweden, Finland, Japan,Finland, Japan, Switzerland, Norway, Netherlands, Denmark, France, Canada, and Hong Kong had infant mortality rates lower than 10 per 1,000 live births.

(UNICEF, 1987)

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F If the numerator of a rate is confined to a certain age, the denominator should relate to the general population.

   (b) T/F The numerator of a rate is always a subset of the denominator.

   (c) T/F The multiplier term in a rate could be 100,000.

   (d) T/F Ratios measure the relationship of one health outcome to another.

   (e) T/F The proportional mortality expresses the number of deaths from a specific cause relative to the total number of people in the population.

   (f) T/F All ratios are rates.

   (g) T/F The point prevalence rate is the proportion of the population that has an outcome or a risk factor at a specific point in time.

   (h) T/F The incidence rate is the number of new cases that develop in a population at risk during a specific time interval.

   (i) T/F The cumulative incidence rate assumes that the entire population is followed from the beginning of the study to the end of the study.

   (j) T/F Incidence density measures the exact amount of time each study participant is followed.

   (k) T/F Incidence density is an appropriate measure when the risk of developing the outcome of interest changes over time.

   (l) T/F Incidence density is useful when study subjects die from causes unrelated to the study question before the study is completed.

   (m) T/F The denominator in incidence density includes the entire population of the area.

(n) T/F If women who have had hysterectomies are included in calculating incidence density, our result will be an overestimation of the rate of endometrial cancer.

(o) T/F Comparisons using crude rates are accurate when populations have similar distributions with respect to a risk factor.

(p) T/F Adjusted rates are commonly used to correct for age differences between populations.

2. Define the following rates:

(a) Crude birthrate:

(b) Maternal mortality rate:

(c) Perinatal mortality rate (use Perinatal Period I):

(d) Infant mortality rate:

3. Circle true (T) or false (F).

(a) T/F A surrogate for the population at risk is used when we measure the induced termination of pregnancy rate.

(b) T/F The induced termination of pregnancy rate gives the probability of an induced termination of pregnancy among women in a specific age group.

(c) T/F WHO recommends no time limitation on maternal mortality ratio calculations for geographic areas within the United States.

(d) T/F The perinatal period I is number of deaths of infants less than seven days, gestation of 28 weeks or more.

(e) T/F Measures of infant mortality show the chances that live-born infants will die within the first two years of life.

(f) T/F The total fertility rate is the sum of age-specific birthrates.

(g) T/F The denominator of the general fertility rate uses the total midyear population.

4. Data were collected in Michigan from 1950 to 1964 on the incidence of Down's syndrome by birth order and maternal age (Example 3.10).

Example 3.10

Maternal Age and Birth Order
Direct Adjustment

Birth Order	Population at Risk*	Maternal Age 20-24 Yrs p	Expected Cases	Maternal Age Years < 20 p	Expected Cases
1	731,000	.00043		.00047	
2	725,000	.00046		.00035	
3	569,000	.00040		.00020	
4	358,000	.00038		.00044	
5+	443,000	.00026		.00000	
Total	2,826,000	Crude Rate: .00043	Adjusted Rate:	Crude Rate: .00043	Adjusted Rate:

* Michigan Standard

(Kleinbaum and Kleinbaum, 1979)

In this example, birth order is the confounding factor. Note that because the maternal age rates increment in different directions by birth order, the results are difficult to interpret.

(a) Complete the chart by computing the expected number of cases and the adjusted rates for both maternal age strata:

Age 20-24: Adjusted rate =

Age < 20: Adjusted rate =

(b) Was there a reversal from crude to adjusted rates?

(c) How might the results be interpreted?

Suggested Answers to Practice Exercises

1. True or false.

   (a) F Rates. The numerator and denominator of a rate should reflect a similar population; if the numerator is confined to a certain age, sex, or racial group, then the denominator should be similarly restricted.

   (b) T Rates

   (c) T Rates

   (d) T Ratios

   (e) F Ratios. The proportional mortality measures the relative importance of a specific cause of death relative to all deaths in a population.

   (f) F Ratios. All rates are ratios, but the converse is not true because in ratios the numerator is not necessarily a subset of the denominator.

   (g) T Prevalence and Incidence Rates

   (h) T Prevalence and Incidence Rates

   (i) T Prevalence and Incidence Rates

   (j) T Prevalence and Incidence Rates

   (k) F Prevalence and Incidence Rates. An underlying assumption in calculating the ID is that the risk of developing the disease of interest is constant over time and age group.

   (l) T Prevalence and Incidence Rates

   (m) F Prevalence and Incidence Rates. The denominator should include only those people considered to be *at risk* of developing the disease.

(n) F Prevalence and Incidence Rates. If women who have had a hysterectomy are included in the denominator, the rate of endometrial cancer would be underestimated because they are not part of the population *at risk* of developing endometrial cancer.

(o) T Adjusted Rates

(p) T Adjusted Rates

3. True or false.

(a) T Induced Termination of Pregnancy Measures

(b) T Induced Termination of Pregnancy Measures

(c) F Maternal Mortality Measures

(d) T Perinatal Mortality Measures

(e) F Infant Mortality Measures. Measures of infant mortality are intended to show the likelihood that live-born infants will die during the first year of life.

(f) F Live Birth Measures

(g) F Live Birth Measures

4. Answers to Example 3.10.

Example 3.11

Maternal Age and Birth Order
Exercise Answer

Birth Order	Michigan Standard Population at Risk	Maternal Age 20-24 Years p	Maternal Age 20-24 Years Expected	Maternal Age Years < 20 p	Maternal Age Years < 20 Expected
1	731,000	.00043	314.33	.00047	343.57
2	725,000	.00046	333.50	.00035	253.75
3	569,000	.00040	227.60	.00020	113.80
4	358,000	.00038	136.04	.00044	157.52
5+	443,000	.00026	115.18	.00000	000.00
Total	2,826,000	Crude Rate: .00043		Crude Rate: .00043	

(Kleinbaum and Kleinbaum, 1979)

(a) Age 20-24: Adjusted rate = $\frac{1126.95}{2{,}826{,}000}$ = .00039 or 39.9/100,000

Age < 20: Adjusted rate = $\frac{868.64}{2{,}826{,}000}$ = .000307 or 30.7/100,000

(b) The crude rates are equal. The adjusted rates are different.

(c) Because no age group has consistently higher birth-order-specific rates, our use of overall rates is questionable. Furthermore, overall rates obscure birth-order-specific differences.

References

American Academy of Pediatrics and the American College of Obstetricians and Gynecologists. Standard terminology for reporting of reproductive health statistics in the United States. Public Health Rep 1988;103:464-471.

Becerra JE, Smith JC. Maternal smoking and low birthweight in the reproductive history of women in Puerto Rico, 1982. Am J Public Health 1988;78:268-272.

Chiswick ML. Commentary on current World Health Organization definitions used in perinatal statistics. Arch Dis Child 1986;61:708-710.

Fox JP, Hall CE, Elveback LR. Epidemiology: man and disease. New York: Macmillan Publishing Company, 1970.

Health Care Financing Administration. International classification of diseases, 9th revision (1975), clinical modification. Washington, DC: Health Care Financing Administration; 1980; DHHS publication no. (PHS)80-1260.

Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.

Hubert HB, Eaker ED, Garrison RJ, Castelli WP. Lifestyle correlation of risk factor change in young adults: the eight-year study of coronary heart disease risk factors in the Framingham offspring. Am J Epidemiol 1987;125:812-831.

Kelsey L, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.

Kleinbaum DG, Kleinbaum AP. Adjusted rates: the direct rate. In: Modules and monographs in undergraduate mathematics and its applications project (Unit 330). Newton, Massachusetts: EDC/UMAP, 1979.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, California: Wadsworth Inc., 1982.

Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology. New York: Oxford University Press, 1980.

MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown and Company, 1970.

Mausner JS, Bahn AK. Epidemiology: an introductory text. Philadelphia: W.B. Saunders Company, 1974.

Nolan TF, Ory HW, Layde PM, Hughes JM, Greenspan JR. Cumulative prevalence rates and corrected incidence rates of surgical sterilization among women in the United States, 1971-1978. Am J Epidemiol 1982;116(5):776-781.

Pan American Health Organization. Mortality analysis-some new uses for old indicators. Epidemiol Bull 1989;10(2):1-16.

Pokras R. Hysterectomy: past, present, and future. Stat Bull Metro Insur Co 1989;FO:12-21.

Smith JC, Warren CW, Garcia-Nunez J. The U.S.-Mexico Border: contraceptive use and maternal health care in perspective, 1979. United States-Mexico Border Health Association: El Paso, Texas, 1983.

World Health Organization. Maternal mortality rates: tables of the available information. Geneva: World Health Organization, 1985.

World Health Organization. World Health Statistics. Geneva: World Health Organization, 1988.

World Health Organization. International Conference for the Tenth Revision of the International Classification of Diseases. Geneva: World Health Organization, In Press, 1991.

UNICEF. State of the world's children, 1987. Oxford: Oxford University Press, 1987.

Chapter 4 - Learning Objectives

After completing this chapter, you should be able to:

1. Differentiate descriptive from analytic epidemiology.
2. Define the general concepts exposure and disease.
3. Identify the following types of epidemiologic study design:

   randomized clinical trial
   cohort
   case-control

4. Identify strategies to eliminate bias.
5. Identify the following types of bias:

   selection
   information
   confounding

6. Define sensitivity and specificity and their relationship to misclassification.
7. Identify the relationship of the p value to chance.
8. Distinguish effect modification and confounding.
9. Define reliability and internal validity.
10. Recognize limitations on the generalizability of epidemiologic findings.
11. Identify factors to consider when evaluating the epidemiologic evidence of a causal relationship between an exposure and a disease or health outcome.

4 Epidemiologic Study Design

Introduction

Overview of chapter

This section provides an overview of epidemiologic study designs, introduces epidemiologic terms, and distinguishes between analytic and descriptive epidemiology. In addition to the fundamentals of study design, estimation, and testing, we must be familiar with the limitations of epidemiologic studies before we can interpret the results of studies. These limitations express themselves through biases in the selection of participants, and the collection and analysis of data. Thus, this chapter also introduces the major types of bias and other threats to the reliability and validity of epidemiologic research. Finally, the chapter enumerates the major criteria used for assessing the causality of epidemiologic findings.

Analytic and Descriptive Epidemiology

Experimental and Nonexperimental Studies

Epidemiologic studies may be categorized into two general classes, experimental and nonexperimental. The difference between these classes is based on whether the researcher has any control over the exposure—the agent that may potentially cause or control a disease—that is being evaluated. In this chapter, we use the term disease to refer to the health problem or outcome of interest.

Experiments are the first major class of epidemiologic research. Because epidemiologists typically control the exposure in experimental studies, such studies should provide stronger evidence of an association or a lack of an association between an exposure and a health problem than would nonexperimental studies. Common examples of epidemiologic experiments are

clinic-based trials of new therapeutic agents (e.g., evaluation of AZT in the treatment of AIDS patients) or communitywide interventions of health education programs (e.g., media campaigns to encourage family planning).

The second major class of epidemiologic research, nonexperimental studies, comprises the majority of all epidemiologic studies. Nonexperimental epidemiology is further divided into two major subtypes, descriptive and analytic. Descriptive studies (described in Chapter 6) are performed when relatively little is known about the risk factors or natural history of a particular disease or condition. Typically, a descriptive study focuses on patterns of disease occurrence in relation to groups of persons, geographic places, or periods in time. For example, sudden infant death syndrome (SIDS), a major cause of postneonatal mortality, has a very poorly understood etiology. Descriptive studies designed to determine if SIDS occurs more frequently among certain ethnic groups, within certain geographic areas, or at certain times of the year could be instructive.

Descriptive studies generate hypotheses

In contrast, analytic studies, sometimes referred to as etiologic studies, are performed to test specific hypotheses about a specific health problem. For example, if a particular ethnic group were to experience a relatively high frequency of SIDS deaths, the investigator might look for more specific risk factors within that ethnic group.

Analytic studies test hypotheses

Overall, the difference between analytic and descriptive epidemiology is primarily one of emphasis rather than method (Kleinbaum et al., 1982). The major differences relate to the use of comparison groups, measures of effect or association, and hypothesis generation rather than hypothesis testing. Although descriptive studies typically do not select persons into formal comparison groups, investigators often form internal comparison groups after examining the initial descriptive statistics. For example, consider data collected on pregnancy history and contraceptive practices from a random sample of women. We could summarize the data using simple descriptive statistics of family size, types of contraception, and other factors. In another analysis, the same data could be reanalyzed by making internal comparison groups based on the choice of contraceptive method. We could then compare different characteristics (e.g., family size, ethnic distribution, etc.) in relation to choice of contraceptive method.

In general, the associations observed in descriptive studies are often the basis for gathering more specific data and testing hypotheses in

additional studies. Conclusions concerning risk factors from nonexperimental research must usually be obtained in several studies before they receive widespread acceptance. Similarly, the biological plausibility of a particular finding affects its acceptability by the scientific community.

Study design is the protocol for selecting persons to study and the method in which data are collected. Study design constitutes the major difference between descriptive and analytic epidemiology. Analytic epidemiology involves the selection and comparison of two or more groups of persons, based on either their exposure or disease status, to evaluate an association between exposure and disease. Although descriptive epidemiology does not typically select persons on the basis of exposure or disease, descriptive studies may divide participants into subgroups for comparison.

Exposure defined

Exposure includes potential risk factors and interventions in the etiology of disease as well as therapies in the treatment of disease. Similarly, disease includes any health problem on the continuum from complete health to death. Thus, in the context of reproductive epidemiology, examples of exposure include maternal age, contraceptive methods, or surgical procedures; examples of disease include low birthweight, pelvic inflammatory disease, ectopic pregnancy, or surgical complications. Depending upon the question under study, a characteristic may be an exposure or a disease. For example, low birthweight is a risk factor for neonatal mortality, but low birthweight may also be a consequence of poor prenatal care. By definition, an exposure must antedate the occurrence of disease.

Measures of effect

Second, analytic epidemiology focuses on measures of effect to quantify the magnitude of the association between the exposure and the health problem under study. These measures of effect are most commonly expressed as ratios of or differences between rates, probabilities, and proportions. In addition to measuring the magnitude of the etiologic (or preventive) effect of a particular exposure, some measures of effect indicate the proportion of disease occurrences caused (or averted) within a particular subgroup or population.

Finally, descriptive epidemiology emphasizes hypothesis generation, and analytic epidemiology typically emphasizes hypothesis testing; both study designs may use statistical tests. Descriptive epidemiologic studies might suggest potential risk

factors that can be altered to reduce or prevent disease. Ultimately, those clues could generate hypotheses that would be more explicitly tested in analytic studies. By hypothesis testing, analytic epidemiology seeks either to confirm (support) or to refute (reject) previously reported associations (Hennekens and Buring, 1987). Both hypothesis generation and hypothesis testing involve writing careful research hypotheses and questions that describe the association between the outcome under study and the exposure of interest expected to be seen at the conclusion of the study. In Example 4.1, the clinical trial was designed to test the research hypothesis that infection rates following intrauterine device (IUD) insertions by nurse-practitioners differed from infection rates following IUD insertions by gynecologists.

Example 4.1

Clinical Trial
Health Care Provider and Post-IUD Insertion Infection Rates

Problem: Do post-intrauterine device (IUD) insertion infection rates differ between nurse-practitioners and gynecologists?

Research Hypothesis: The rate of infection following IUD insertions by nurse-practitioners is different from the rate of infection following IUD insertions by gynecologists.

Study Design: Clinical trial

Eligibility Criteria: Women 30 to 45 years old with at least one prior pregnancy

Treated: Women with IUDs inserted by nurse-practitioners constitute the treatment group.

Not Treated: Women with IUDs inserted by gynecologists constitute the control group.

Outcome: Each group of women is followed up on a regular basis to monitor any development of infection. The same method of assessing infection should be used for the treatment and control groups.

Data Collection Methods: Regardless of who inserted the IUD, the patients are visited twice a week for six weeks by a nurse to determine whether infection develops. The nurse does not know which women were treated by gynecologists or nurse-practitioners.

Study Design

Experimental

Study design defined

Study design refers to the manner in which groups of persons with particular characteristics are assembled and compared to evaluate the association between a risk factor or exposure and a disease or outcome.

The ideal study design from the scientific point of view is an experiment. The distinguishing feature of an experiment is that the scientist determines the exposure status of each participant and then observes the occurrence of a particular event (for example, disease). The investigator also typically establishes eligibility criteria that study subjects must meet before they can be assigned to a particular exposure. In health research, experiments are most often used to evaluate a new clinical therapy. This design, referred to as a clinical trial, involves the assignment of individual persons to one of two or more therapeutic interventions.

Clinical trials and randomized clinical trials

Clinical trial designs may also involve the element of randomization (typically called a randomized clinical trial). The epidemiologist assigns study subjects to a particular treatment at random and relies on chance to distribute an equal number of study subjects to each treatment group. Example 4.1 describes an experiment designed to evaluate complications of IUD insertion performed by different types of health care providers. If nonrandom assignment were used in this example, clinic personnel might introduce some bias when assigning a provider (e.g., by assigning patients with more favorable prognosis to the nurse).

Under simple random assignment, each eligible study subject would have an equal chance to be treated by either health care provider. Random assignment is often accomplished by clinic personnel opening a sealed envelope containing the treatment assignment for each eligible person. Randomization minimizes the opportunity for bias and also provides the theoretical basis for the use of statistical models and hypothesis testing.

Nonexperimental

From an ethical point of view, most experimental studies of risk factors are unacceptable. For example, a researcher trying to

observe the effect of passive smoking on birthweight would not randomly assign expectant mothers to environments where they either would or would not be exposed to cigarette smoke. The majority of epidemiologic studies are nonexperimental.

Although we cannot to assign exposure status, good nonexperimental studies should emulate other characteristics of a good experiment. One of these characteristics is the nonbiased collection of data among all groups included in the study. Two ways to promote nonbiased data collection are to use the same standardized data collection forms for all groups under study and observers who do not know the group to which a study subject belongs. When the observer does not know the study subject's group, the observer is considered to be blinded. When neither the observer nor the participant knows the exposure status, the study is referred to as double-blinded. Similarly, nonexperimental studies should use the same questionnaire or laboratory procedure to gather data from all study subjects. Investigators using nonexperimental designs should attempt to use blinded observers for data collection, although this may not always be practical or possible.

Blinded and double-blinded studies

Investigators who properly execute nonexperimental studies select subjects according to some predetermined, objective criteria. In this way, the study subjects should represent reproducible and comparable groups of persons.

Using nonexperimental study designs, we select persons on the basis of either exposure status or disease status. Selection based on exposure is directly analogous to the experimental design in that the epidemiologist knows the level of exposure (or particular clinical therapy) and observes the subsequent occurrence (or prevention) of a health problem. This design is called a cohort or follow-up design. In a cohort study, the epidemiologist follows up the cohort (a defined group of exposed persons) to ascertain the outcome of interest.

Cohort study defined

Most cohort studies are based on the enrollment of currently exposed and unexposed persons. A cohort study design that is based on current exposure status may be called a prospective cohort study. In contrast, historical records of exposure among a group of persons may be used to select individuals for follow-up. This design has most often been used in occupational epidemiology, where job classification may indicate exposure hazards. It is often called a historical cohort study.

Example 4.2 illustrates a prospective cohort study designed to determine the association between passive smoking and respiratory infections in children. In this example, the epidemiologist must identify families with children younger than 12 years old who live at home. Children are considered exposed to smoking if at least one smoker lives in the same household and unexposed if no smokers live in the same household.

Example 4.2

Prospective Cohort Study
Passive Smoking in Children and Respiratory Infections

Problem: Is passive exposure to tobacco smoke associated with increased respiratory infections in children?

Research Hypothesis: Passive exposure to tobacco smoke is associated with increased frequency of respiratory infections.

Study Design: Cohort study of children exposed and not exposed to tobacco smoke in their home.

Exposed: Children younger than 12 years old who live at home where at least one smoker resides.

Not Exposed: Children younger than 12 years old who live at home where no smokers reside.

Outcome: Outcome is respiratory infections in children. Families will be contacted on a monthly basis for one year to determine the frequency of respiratory infections. The method of assessing infection frequency should not vary according to smoking status.

When selection is based on current disease status rather than exposure status, the epidemiologist collects information on exposure history from diseased and nondiseased study subjects. Study subjects with disease are called cases and study subjects without disease are called controls. This design is called a case-control study, although some researchers prefer the name case-referent study. Example 4.3 illustrates a case-control study of breast cancer in relation to previous use of contraceptive hormones.

Cases-control study defined

Example 4.3

Case-Control Study
Oral Contraceptive Use and Breast Cancer

Problem: Does use of oral contraceptives influence the development of breast cancer?

Research Hypothesis: Using oral contraceptives is associated with breast cancer.

Study Design: Hospital-based case-control study of women with newly diagnosed breast cancer.

Cases: Women are selected at time of initial diagnosis of breast cancer from several hospitals during a one-year period.

Controls: Women who do not have a history of cancer are selected from the same hospitals as the cases during the same one-year period.

Exposure: Both cases and controls receive a standardized personal interview to obtain information on previous use of oral contraceptives. Both groups of women must be questioned in the same manner.

Incident and prevalent cases

When studying the etiology of a disease, investigators who use the case-control study should recruit persons with newly diagnosed disease (incident cases) rather than persons with a history of disease (prevalent cases). The study protocol must specify how persons with incident disease will be ascertained and how they will be distinguished from persons with previous disease. The causation of disease is related to factors present before the onset or incidence of disease. Persons with incident disease (incident cases) are more likely to represent the distribution of risk factors among individuals developing disease. In contrast, the prevalence (P) of disease is related to both disease incidence (I) and the average duration (D) of the disease ($P = I * D$). All persons with disease (prevalent cases) may overrepresent factors related to survival with disease rather than causation of disease. For example, women with ovarian cancer have a relatively short survival. Therefore, a case-control study that enrolled only women surviving with ovarian cancer would underrepresent

Prevalence of disease

women who died soon after diagnosis; consequently, the risk factors related to the rapid course of disease might be missed.

In summary, descriptive studies usually select persons who represent some segment of the population and do not necessarily select persons on the basis of exposure or disease status. Analytic designs select study subjects either on the basis of exposure status (persons are followed forward in time until a particular outcome occurs) or on the basis of disease status (data is obtained on exposures that occurred in the past).

Basic Formulas for 2 x 2 Tables

In the simplest terms, the goal of most epidemiologic studies is to complete a two-way table. Typically this table is a 2 x 2 table with exposure on one axis and disease (or outcome) status on the other (Figure 4.4). Data organized in this way can be used to calculate measures of association (e.g., risk ratios, rate differences, or odds ratios) and the appropriate statistical tests.

The risk difference (Rothman and Boice, 1979) is the arithmetic difference between two risks and is computed from Figure 4.4 as:

Risk difference

(4.4.1)

$$RD = \frac{a}{a+c} - \frac{b}{b+d}$$

$$= \frac{a}{n_1} - \frac{b}{n_0}$$

The confidence interval, using test-based procedures, is estimated as:

Confidence interval for risk difference

(4.4.2)

$$\text{Confidence Interval for RD} = RD * (1 \pm z / \chi)$$

where z is the normal variate associated with the desired level of precision (1.96 for 95% confidence limits; 1.65 for 90% confidence limits) and

(4.4.3)

$$\chi = \sqrt{\chi^2_{MH}} \quad \text{and} \quad \chi^2_{MH} = \frac{(t-1) * (a * d - b * c)^2}{n_1 * n_0 * m_1 * m_0}$$

Population attributable risk

Risk difference is also known as attributable risk. Using the attributable risk, it is possible to estimate the excess rate of disease in the total population attributable to the exposure. The population attributable risk (PAR) can be calculated by multiplying the attributable risk by the proportion of exposed persons in the population.

(4.4.4)

PAR = RD * proportion of exposed persons in the population

Cumulative incidence relative risk (CIR)

The cumulative incidence relative risk (CIR) for a randomized clinical trial or a cohort study is the ratio of the risk in the exposed group relative to the risk in the unexposed group; it is computed from Figure 4.4 as:

Figure 4.4
2 x 2 Table for Case-Control or
Cohort Cumulative Incidence Data

Outcome	Exposure Status: Exposed	Exposure Status: Unexposed	
Present	a	b	m_1
Absent	c	d	m_0
	n_1	n_0	t

a + b + c + d = t

(4.4.5)

$$CIR = \frac{\frac{a}{n_1}}{\frac{b}{n_0}}$$

The confidence interval, using test-based procedures, is estimated as:

Confidence interval for cumulative incidence relative risk

(4.4.6)

$$\text{Confidence Interval for CIR} = CIR^{(1 \pm z / \chi)}$$

where z is a normal variate and χ is defined in 4.4.3.

The odds ratio for a case-control study is the odds of disease in the exposed group relative to the odds of disease in the unexposed group; it is computed from Figure 4.4 as:

Odds ratio

(4.4.7)

$$OR = \frac{a * d}{b * c}$$

The approximate confidence interval for the OR, using test-based interval estimation, is estimated as:

Confidence interval for odds ratio

(4.4.8)

$$\text{Confidence Interval for OR} = OR^{(1 \pm z / \chi)}$$

where z is a normal variate and χ is defined as in 4.4.3.

Most examples in this manual are based on *count* or frequency data. With count data, all study participants are classified as exposed or unexposed and sick or well—these frequencies fill in the four cells of a 2 x 2 table. Count data from cohort studies enable us to estimate the proportion of persons who develop disease in two (or more) exposure groups after a given amount of time. This proportion is also referred to as a cumulative incidence.

Count data

Cumulative incidence

Person-time data

Incidence density

A measure of incidence that explicitly incorporates the amount of time a person is followed up until the outcome occurs is person-time incidence (also referred to as incidence density). Person-time data may be arranged in a two-way table in which the cells of the table (a and b in Figure 4.5) represent the number of cases observed among exposed and unexposed persons; the *totals* at the bottom of each exposure column (n_1 and n_0) represent the amount of person-time of follow-up for each exposure group. Person-time is accumulated by summing the amount of time (e.g., the number of days, months, or years) that an individual is exposed or unexposed until the person develops the outcome or drops out of the study, or the study is ended. (m_1 represents the total number of persons with the outcome; t represents the total amount of person-time.)

Figure 4.5
2 x 2 Table for Incidence Density Data

	Exposure Status: Exposed	Exposure Status: Unexposed	
Outcome			
Present	a	b	m_1
Person-time	n_1	n_0	t

Person-time analysis permits us to incorporate more of the information obtained in cohort studies. We may include data on persons who are lost to follow-up until the time when their outcome status was last evaluated. Person-time data also forms the basis of life table analysis, which allows us to evaluate the rate of outcome at intervals within the total follow-up period. Formulas 4.4.1 - 4.4.2 and 4.4.4 - 4.4.6 are applicable for person-time data if we let n_1, n_0, and t

represent person-time data instead of count data and let χ be defined as:

(4.4.9)

$$\chi = \frac{a - m_1 * n_1 / t}{\sqrt{m_1 * n_1 * n_0 / t^2}}$$

Bias (Systematic Error)

Bias defined

In this section we will introduce the three major classes of bias and the general approaches used to minimize their effects. (More specific information on these biases will be provided for each study design covered in this workbook.)

In epidemiology, bias is a systematic deviation between the true value of a statistic (e.g., an estimate of the relative risk) and the value estimated by the epidemiologic study. Proper study design is primarily concerned with eliminating the influence of potential biases. Although many types of bias exist, biases are typically grouped into three major categories: selection, information, and confounding. In general, bias may be minimized by ensuring that the groups of study subjects (the exposed and unexposed persons in a clinical trial or cohort design, and the cases and controls in a case-control design) are comparably selected, interviewed (or otherwise provide data), and influenced by risk factors other than the one being studied.

Extraneous risk factors may weaken the comparability of the study groups. If we can measure these factors, we can adjust the analysis for their impact. Without additional participant enrollment or data collection, however, we may have difficulty adjusting for selection and information biases that have already occurred. Nevertheless, the analysis may incorporate various *worst case* scenarios to estimate the way in which selection and information biases may have distorted the results. Sensitivity analysis assesses the effects of worst case situations and evaluates how *sensitive* the results are to different potential biases.

Sensitivity analysis

Selection Bias

Selection bias is the first major class of bias we consider. This group of biases concerns how individuals are selected for an epidemiologic study. To prevent selection bias, we must select all study subjects (cases and controls or exposed and unexposed persons) in a comparable manner from the same population.

In a clinical trial, the study subjects are assigned to a particular treatment group. Other than their treatment, all subjects should have a similar prognosis for the health problem or outcome of interest. Random assignment tends to produce similar groups with similar risks for the outcome of interest.

In a cohort study, an epidemiologist enrolls study subjects on the basis of their exposure status. The goal is to obtain two groups of persons who have similar risks for the health outcome of interest—except for the particular exposure under study. Selection bias often occurs in a cohort study when persons are lost to follow-up or when their participation is related to the outcome. Example 4.6 describes a prospective cohort study on the effect of prenatal care on Bayley's Mental Development Index (Bayley, 1969) at two years of age. Although information on prenatal care may be obtained from comparable groups of women who are giving birth, we may have difficulty maintaining contact with the women over the two-year study

Example 4.6

Prenatal Care and Bayley's Mental Development Index

Problem: Is prenatal care associated with Bayley's Mental Development Index at two years of age?

Research Hypothesis: Receipt of prenatal care is associated with increased index score.

Study Design: Cohort study; mothers are enrolled through hospital records.

Unexposed: Mothers who received no prenatal care.

Exposed: Mothers who received prenatal care.

Outcome: The outcome of interest is the index score.

period. Highly motivated parents may be more likely to maintain interest in the study; less motivated parents may tend to be lost to follow-up. At the end of two years, both groups of children may score similarly on the index. This outcome, however, could be attributed to differences in the home environment of children with *motivated* parents regardless of prenatal care.

Healthy worker effect

Thus, selection bias occurs in a cohort study when participation is related to the outcome of the study. For example, the healthy worker effect is a selection bias by which employed persons tend to be healthier than unemployed persons (particularly if the work is physically demanding) and will therefore have a lower risk of disease than the general population.

Hospital admission bias

In the case-control study, study subjects are selected on the basis of disease status. The goal is to obtain persons with incident disease (cases) and disease-free persons (controls) who represent the population from which the cases arose. Thus, other than their disease status, case and control persons would have a similar chance of having the exposure of interest. Example 4.7 shows how failure to enroll all newly diagnosed cases of cervical cancer can result in a form of selection bias called hospital admission bias. This bias occurs when the case subjects in a case-control study tend to be sicker than nonhospitalized cases, and consequently have different opportunities for previous exposures. Hospital admission bias may operate in a study of cervical cancer if only hospitalized case patients are selected because hospitalized patients are more likely than nonhospitalized patients to have advanced disease. When cervical cancer is detected early, as preinvasive disease, treatment may be given on an out-patient basis rather than on an inpatient basis. If Pap smears are associated with the detection of preinvasive cervical cancer, women with preinvasive cancer would be less available for inclusion in a study of hospitalized case patients. In Example 4.7, the table that includes all incident cases presents the hypothetical true relationship between Pap smear frequency and incident cervical cancer. Because 40% of case patients had a history of Pap smear and 60% of the controls patients had had Pap tests, we find that the odds ratio of 0.44 indicates that Pap smears are protective against invasive cancer.

If we could estimate the selection probabilities that limited the persons entering the study, we could estimate how the true odds

Example 4.7

Hospital Admission Bias
Pap Smear Screening and Cervical Cancer

Problem: Is clinically diagnosed cervical cancer associated with a history of Pap smear screening?

Research Hypothesis: History of Pap smear screening reduces the risk of clinically diagnosed cervical cancer.

Study Design: Hospital-based case-control study.

Cases: Women hospitalized for treatment of cervical cancer during the past year.

Controls: Women selected at random from all women without cervical cancer who received hospital care during the past year.

All Incident Cases	History of Pap Smear			
	Yes	No		
Cases	40	60	100	Odds Ratio = 0.44
Controls	60	40	100	
	100	100	200	

Hospitalized Cases Only	History of Pap Smear			
	Yes	No		
Cases	20	45	65	Odds Ratio = 0.29
Controls	39	26	65	
	59	71	130	

ratios may have been biased in the study. The table in Example 4.7 that includes only hospitalized cases suggests that 65% of women with incident cervical cancer were hospitalized. Furthermore, hospitalization varies by Pap smear history. For example, only 50% of women with cancer and Pap smears were hospitalized (20/40 = .5), whereas 75% of women with cancer and no previous Pap smear were hospitalized (45/60 = .75). For simplicity, we have assumed that the control patients accurately reflect the frequency of Pap smears among

women without cervical cancer. These results indicate a positive bias (an overestimation of the protective effect of Pap smears) because the data on the case patients underestimates the proportion of women with Pap smears (20/65 or 31%). Thus, the odds ratio based only on hospitalized cases is spuriously lower (0.29) than the odds ratio based on all incident cases (0.44). To reduce selection bias, the epidemiologist should obtain all newly diagnosed cases, regardless of hospitalization, for a specified period of time to ensure that women with less advanced disease are not excluded from the study. The controls should be representative of the population at risk of becoming cases.

Types of selection bias

Researchers have described several different types of selection bias (Sackett, 1979). The primary cause of selection bias varies with study design. For example, in cohort studies, selection bias occurs when factors associated with the health outcome influence study participation. In case-control studies, selection bias occurs when factors associated with exposure influence study participation (Rothman, 1986). When selection bias exists, the relationship between exposure and the health outcome is different for the study participants than for the eligible nonparticipants. If factors are identified that accurately reflect the selection bias (for example, the impact of motivated parents on mental development in Example 4.6), we can adjust for this bias in the analysis (Rothman, 1986).

Information Bias

Information bias (also known as observation, classification, or measurement bias) is the second major class of bias. This group of biases concerns the way in which information is obtained from or about study subjects. In the ideal study design, we collect information about all study subjects in a similar manner.

Types of information bias

Researchers have described several types of information bias (Sackett, 1979). The major types of information bias include ascertainment bias, recall bias, interviewer bias, missing data bias, and prevarication bias (study subject or data collection personnel intentionally distort the data). The distinguishing feature of information bias is that it occurs after the study subjects have been selected. Information bias results when the study subjects are misclassified with respect to exposure or disease or both. The

consequences of information bias depend on whether misclassification is random.

Ascertainment bias is a form of information bias that occurs if the outcome of interest is obtained by using different methods for each comparison group. In Example 4.7, ascertainment bias would occur if history of Pap smear was obtained by medical record review for case patients but by personal interview for control patients. Ascertainment bias could occur in Example 4.2 if respiratory infections among children exposed to cigarette smoke were measured by personal interviews of parents but infections among unexposed children were measured by self-administered questionnaires given to parents.

Differential misclassification

Depending on the information sought, the various forms of information bias result in differential misclassification. That is, the prevalence of a risk factor or outcome in one of the study groups is systematically underestimated or overestimated. Example 4.8 shows how the use of a different source of information for case subjects and control subjects can cause differential misclassification. Consider the possibility that control subjects were less likely to report previous diagnostic radiation than was shown in the medical records. Because the information from case subjects about previous exposure to radiation was obtained from medical records and the information from control subjects was obtained from personal interview, the use of diagnostic radiation among the control subjects might be underestimated. If radiation and breast cancer were truly not associated, the results from this study would show a positive association between radiation and breast cancer. To avoid differential misclassification, the epidemiologist should collect data from the comparison groups in the same manner. In Example 4.8, abstracting medical records for both groups would probably yield comparable data.

Nondifferential misclassification

In contrast to differential misclassification, nondifferential misclassification will occur if data are poorly collected from all study subjects. For example, if a questionnaire has not been field-tested or validated, some proportion of the study subjects in each group may be misclassified. This bias will either falsely decrease a positive association or falsely decrease a negative association. By reducing the magnitude of the true association, nondifferential misclassification increases the chance of finding no association (Kleinbaum et al., 1982).

We can illustrate nondifferential misclassification using the data in Example 4.9. Suppose the true odds ratio of the relationship between

Example 4.8

Differential Misclassification
Diagnostic Radiation and Breast Cancer

Problem: Is breast cancer associated with a history of diagnostic radiation?

Research Hypothesis: Diagnostic radiation is associated with breast cancer.

Study Design: Case-control study

Cases: Women hospitalized with newly diagnosed breast cancer. A history of previous diagnostic radiation was obtained from a review of medical records.

Controls: A random sample of women without breast cancer. Controls were selected from the communities served by the hospital that admitted women with breast cancer. A history of previous diagnostic radiation was obtained from telephone interview.

Data Analysis: Diagnostic radiation: Reported results and true results

	Reported Results			
	Yes	No	Total	
Cases	40	60	100	Odds Ratio=2.0
Controls	25	75	100	
Total	65	135	200	

	True Results			
	Yes	No	Total	
Cases	40	60	100	Odds Ratio=1.0
Controls	40	60	100	
Total	80	120	200	

Table 4.9

Nondifferential Misclassification
Pap Smear Screening and Cervical Cancer

True history of screening with Pap smear		True Results			
		Yes	No	Total	
	Cases	40	60	100	Odds Ratio=0.44
	Controls	60	40	100	
	Total	100	100	200	

Reported history screening with Pap smear		Nondifferentially (Randomly) Misclassified Results			
		Yes	No	Total	
	Cases	42	58	100	Odds Ratio=0.52
	Controls	58	42	100	
	Total	100	100	200	

screening with Pap smears and cervical cancer is 0.44. If 10% of all cases and 10% of all controls are misclassified with respect to Pap smear screening, regardless of their true screening status, we would obtain the results shown in the lower panel of Table 4.9. We obtain these results because 10% of the 40 case women who received screening are misclassified as not receiving screening, and 10% of the 60 case women who received no screening are misclassified as having screening. The misclassification probabilities result in 42 case women who had been screened (40 − (40 * 10%) + (60 * 10%)) and 58 who had not been screened (60 − (60 * 10%) + (40 * 10%)). Using similar calculations we would obtain the results shown for controls. The new odds ratio of 0.52 is closer to the null value of 1.0 than the true odds ratio 0.44; and the strength of the association has been decreased.

Nondifferential misclassification introduces a conservative bias (i.e., it biases results toward the null value—toward no association between the exposure and outcome). When estimating ratios, the bias will always adjust the estimate toward the value of 1.0—from either a positive association (e.g., from a relative risk of 2.0 to 1.7) or a negative association (e.g., from a relative risk of 0.42 to 0.76).When we estimate risk differences in a cohort study or clinical trial,

nondifferential misclassification may reduce the size of the difference toward zero (Rothman, 1986).

Assessing misclassification

Another way we can assess misclassification when collecting information is to evaluate the sensitivity and specificity of a particular question on a questionnaire. Although sensitivity and specificity are commonly used in describing the characteristics of a diagnostic screening test, these concepts may also describe data collection. The sensitivity of a screening test is the proportion of persons with disease whom the screening test classified as having the disease, and specificity is the proportion of persons whom the screening test classifies as being free from disease. For example, among women with cervical cancer, the Pap smear will correctly classify a certain proportion as being positive for cancer (i.e., that proportion is known as the Pap smear's sensitivity for cervical cancer). Among women without cervical cancer, the Pap smear will correctly classify a certain proportion as being negative for cancer (i.e., that proportion is known as the Pap smear's specificity for cervical cancer).

Sensitivity and specificity

In data collection, sensitivity and specificity refer to the correct classification of an exposure or other risk factor as well as to the correct classification of a health problem or other outcome. In Example 4.10, we provide data to calculate the sensitivity of the original question "Have you ever smoked?" Because the question correctly classified 25 of 30 smokers, its sensitivity was 25/30 or 83.3%. The original question correctly classified 55 of 70 nonsmokers, and thus its specificity was 55/70, or 78.6%.

To increase the sensitivity and specificity of the question, we must clearly define the information of interest. The question must be explicit enough to obtain the desired information. Pilot testing and validating questions will make it possible to obtain more accurate data. Also, adopting questions previously tested and validated in other studies improves the quality of data collected and permits direct comparison of the results in the current study with previous studies.

Example 4.10 shows how sensitivity and specificity relate to misclassification of a risk factor common to many diseases—smoking. In a study where two or more groups are being compared, sensitivity and specificity may vary within the groups being compared. For example, mothers of low-birthweight babies may tend to overreport smoking history, but mothers of

Example 4.10

Obtaining Information on Smoking History

Problem: Classify persons by whether they have previously smoked cigarettes.

Original Question: Have you ever smoked cigarettes? Some participants may say *yes* if they ever smoked even one cigarette but others may say *no* if they never smoked habitually, even though they smoked on a few occasions.

Information of Interest: Suppose you were only interested in finding out if people smoked a *substantial* amount. One standard way to determine this is to ask, "Have you smoked 100 or more cigarettes in your lifetime?" Because of the different interpretations of the original question, some participants who smoked less than 100 cigarettes would be misclassified as smokers and some participants who smoked more than 100 cigarettes would be misclassified as nonsmokers.

Information Obtained: Suppose we have data from 100 people, 30 of whom had smoked and 70 of whom had not smoked according to the *100 cigarette* criterion. Suppose that 25 of the smokers and 15 of the nonsmokers said they had smoked. These numbers are displayed below. Note that the title of the table refers to the *true* (100 cigarette criterion) smoking history, as reflected in the numbers 30 and 70 on the bottom row of the table (30% smokers and 70% nonsmokers).

Data Analysis: Misclassification of smoking history

		True Smoking History: Have smoked 100 or more cigarettes		
		Yes	No	Total
"Have you ever smoked?"	Yes	25	15	40
	No	5	55	60
	Total	30	70	100

normal-birthweight or high-birthweight babies may tend to underreport smoking history. When misclassification of a risk factor is the same in all groups being compared (the sensitivity and specificity are the same among study subjects with and without the disease or health outcome), nondifferential misclassification results. When misclassification varies among compared groups (the groups do not have the same sensitivity and specificity), as in this example, differential misclassification results.

If data are flawed but good estimates of the sources of misclassification can be obtained, we can make some simple adjustments to the tabular data. However, this is not practical when misclassification is contingent upon several factors (Kleinbaum et al., 1982).

Confounding Bias

Confounding bias occurs when the effects of two risk factors are mixed in the occurrence of the health problem under study. Specifically, a confounding factor must meet three criteria: 1) it must be an independent risk factor for the health problem or outcome of interest among the unexposed, 2) it must be associated with the exposure of interest in the data collected, and 3) it must not be an intermediary in the pathway between exposure and disease. Unlike selection and information biases, confounding bias may have an analytic solution—provided we can identify the presence of a particular confounder.

Regarding the relationship between prenatal care and birthweight, we know from previous studies that maternal cigarette smoking causes low birthweight. Suppose that women obtaining prenatal care are less likely to smoke cigarettes than women who do not obtain prenatal care. The cumulative incidence relative risk 3.7 (Table 4.11) indicates that low-birthweight babies are 3.7 times more likely to be born to mothers who do not receive prenatal care than to mothers who do receive prenatal care.

Table 4.11 is a *crude* table. Crude indicates that the data are combined across levels of the additional risk factor (maternal cigarette smoking) ignoring the effect of that factor. If the study does not consider maternal smoking when estimating the effect of prenatal care on birthweight, the results may be confounded by smoking status. One strategy to eliminate this potential confounding is to collect sufficient information on maternal smoking behavior and then analyze the results separately for smokers and nonsmokers.

Stratified analysis

Table 4.12 displays the same data separated according to the smoking status of the mother. Typically, this presentation is called a stratified analysis, and each smoking category forms one stratum. Clearly, relative risk for the effect of prenatal care among smokers alone and among nonsmokers alone is considerably smaller than

Table 4.11

Crude Table
Confounding of the Effect of Prenatal Care on Low-Birthweight Infants

	Exposure		
Low-birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	300	81	381
No	700	919	1,619
Total	1,000	1,000	2,000

CIR = 3.7

Table 4.12

Stratified Table
Maternal Smoking as a Confounder of the Effect of Prenatal Care on Low Birthweight

Stratum 1: Mothers Who Smoke Cigarettes

	Exposure		
Low-birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	290	21	311
No	610	79	689
Total	900	100	1,000

CIR = 1.5

Stratum 2: Mothers Who Do Not Smoke Cigarettes

	Exposure		
Low-birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	10	60	70
No	90	840	930
Total	100	900	1,000

CIR = 1.5

the crude relative risk. In fact, both strata yield the same relative risk of 1.5. If we adjust for the confounding of smoking, we would obtain an accurate summary relative risk of 1.5. The difference between the summary relative risk of 1.5 and the crude relative risk of 3.7 indicates the presence of confounding. An accurate summary relative risk, standardizing for the confounding of smoking, equals 1.5.

Mantel-Haenszel relative risk

The Mantel-Haenszel relative risk or summary odds ratio is used in case-control studies and for cumulative incidence data in cohort studies to adjust for the effects of potentially confounding variables and is computed as (Rothman and Boice, 1979):

(4.12.1)

$$RR_{MH} = \frac{\sum (a_i * d_i / t_i)}{\sum (b_i * c_i / t_i)}$$

where a, b, c, d, and t are defined as in Table 4.4 and i represents the i^{th} stratum (Mantel and Haenszel, 1959).

Confidence interval for Mantel-Haenszel relative risk

We can estimate the confidence interval for the Mantel-Haenszel relative risk (4.12.1) by using test-based procedures (Miettinen, 1976). The confidence interval is computed as:

(4.12.2)

$$\text{Confidence Interval for } RR_{MH} = RR_{MH}^{(1 \pm z / \chi_{MH})}$$

where z is a normal variate, $\chi_{MH} = \sqrt{\chi^2_{MH}}$, and

(4.12.3)

$$\chi^2_{MH} = \frac{\{\sum a_i - \sum m_{1i} * n_{1i} / t_i\}^2}{\sum (m_{1i} * m_{0i} * n_{1i} * n_{0i}) / (t_i^2 * (t_i - 1))}$$

Summary estimate of relative risk

A summary estimate of relative risk is a weighted average of the stratum-specific relative risks, whereas a crude relative risk does not properly account for the distribution of persons among the strata. The formula given in 4.12.1 uses the odds ratio to approximate the relative risk—assuming a rare outcome; formula 4.12.1 may be used for case-control studies. For incidence data

in cohort studies, a different formula (Rothman, 1986) that relies on relative risk instead of the odds ratio approximation is:

(4.12.4)

$$CIR_s = \exp\left[\frac{\sum w_i * \ln(CIR_i)}{\sum w_i}\right]$$

where $w_i = \dfrac{a_i * b_i * n_{1i} * n_{0i}}{a_i * d_i * n_{1i} + b_i * c_i * n_{0i}}$

The confidence interval for CIR_s, using a test-based procedure, is computed as:

(4.12.5)

$$\text{Confidence Interval for } CIR_s = CIR_s^{(1 \pm z / \chi)}$$

where z is a normal variate and χ is defined as in (4.12.3).

To determine how the criteria for assessing confounding apply in Example 4.11, we must first decide if smoking is a risk factor for low birthweight among mothers who do receive prenatal care (the *unexposed* group). The risk of low birthweight was 21/100, or 0.210, among women who received prenatal care and smoked and 60/900, or 0.067, among women who received prenatal care but did not smoke. Because the probability of low birthweight varies according to smoking, we find that smoking is a risk factor among women who received prenatal care. Second, we must determine if smoking is associated with prenatal care in the data we have obtained. Among the women who did not receive prenatal care (the *exposed* group), 900 out of 1,000 (90%) smoked cigarettes; among women who received prenatal care, only 100 out of 1,000 (10%) smoked. Assuming that this study obtained an accurate sample, we find that the difference between 90% and 10% is clear evidence that smoking is negatively associated with prenatal care. Finally, from substantive knowledge, we know that smoking is not an intermediate step in the pathway between prenatal care and low birthweight. That is, prenatal care is not a cause of smoking.

The Mantel-Haenszel procedure (4.12.1) and the CIR_s estimate (4.12.4) adjust relative risk estimates for the effects of confounding

variables. The formula that adjusts risk differences for potentially confounding variables for cumulative incidence data is:

Summary estimate of risk difference

(4.12.6)

$$RD_s = \frac{\sum w_i * RD_i}{\sum w_i}$$

where RD_i is the risk difference in stratum i:

$$RD_i = \frac{a_i}{a_i + c_i} - \frac{b_i}{b_i + d_i}$$

$$= \frac{a_i}{n_{1i}} - \frac{b_i}{n_{0i}}$$

and w_i is the weight for standardizing the stratum-specific estimates of risk difference:

$$w_i = \frac{1}{\frac{a_i * c_i}{n_{1i}^3} + \frac{b_i * d_i}{n_{0i}^3}}$$

The test-based confidence interval can be computed as (Rothman, 1986):

(4.12.7)

$$\text{Confidence Interval for } RD_s = RD_s * (1 \pm z / \chi)$$

Restrict participation

Formulas 4.12.4 - 4.12.7 are applicable for person-time data, if we let n_1, n_0, and t represent person-time. The w_i in (4.12.6) simplifies to $w_i = \frac{1}{\frac{a_i}{n_{1i}^2} + \frac{b_i}{n_{0i}^2}}$

Another strategy to remove confounding bias is to restrict participation in the study. In Example 4.12, the study could be restricted to women who did not smoke. Alternatively, if we are concerned about the effect of prenatal care among smokers, the study could be restricted to women who smoke. The principal issue in either scenario is that, regardless of prenatal care status,

the comparison groups should have the same smoking status to ensure that smoking does not differentially influence the occurrence of low birthweight.

Matching

The final strategy to reduce confounding is to match study subjects for the presence of another risk factor. This strategy requires that matching be considered in the analysis, or else bias may result. Matching can be costly, particularly in follow-up studies (Rothman, 1986). Also, matching has an additional implication—it precludes determining the magnitude of risk associated with the matching factor. For this reason we should first consider the strategies of restriction or increasing the size of the study groups as more efficient ways of obtaining information.

Anticipate biases in design phase of study

In summary, potential biases are best managed by anticipating them in the design phase of a study (Table 4.13). In particular, selection and information biases are not easily managed in the analysis unless additional information is collected or enrollment of participants is extended. Nevertheless, it may be helpful to consider sensitivity analysis to assess the effect of the biases or the results. Confounding has good analytic solutions if sufficient data are available for use in a tabular or regression-based analysis. Although no study is free from bias, the well-conducted randomized clinical trial is the study design that is most likely to have minimal bias.

Table 4.13

Strategies for Addressing Biases in the Design and Analysis Phases of an Epidemiologic Study

Type of Bias	Design Strategy	Analysis Strategies
Selection (study group differs from target group)	• Enroll all incident cases from geographic areas • Enroll controls from same area • Avoid self-referral • Increase frequency of contact during follow-up • Blind assignment of exposure (clinical trial)	• Estimate selection or loss probabilities • Conduct a sensitivity analysis • Obtain data on selection factor and treat as confounder

Table 4.13 (continued)

Information (error in the classification exposure or disease status)	• Standardize the measurement regardless of exposure and disease status • Use *blinded* observers • Use surrogate sources rather than dropping participants from study • Validate the measurement procedure • Obtain historical data in reference to time of disease onset	• Estimate probabilities of misclassification based on validation study • Conduct a sensitivity analysis • Look for systematic differences between observers
Confounding (effect of exposure is mixed with effect of an additional risk factor)	• Collect information on potential confounders • Restrict participation to persons with similar baseline risk of disease • Match on major potential confounder(s)	• Stratify on the basis of potential confounder in a tabular analysis • Add a term for potential confounder in regression analysis • Use adjusted rates to adjust for confounding • Stratify on matching term(s) in the analysis

Chance (Random Error)

Random error defined

Bias may introduce error at any stage of an investigation; its net effect is to produce results that differ systematically from the truth. Chance, another source of error, differs from bias in that chance represents random error. The primary way to reduce error attributable to chance is to increase the sample size of the study. However, this will be true only to the extent that more accurate information is obtained. Specifically, we must consider the distribution of exposure, disease or health status, and other individual risk factors to appreciate the amount of information present in a particular study. The more information that is available, the less will be the influence of random error (Rothman, 1986).

Reducing random error

Although increasing the sample size will reduce random error, the efficiency with which we collect the information from a given number of study participants will also affect random error. For example, a prospective cohort study of a rare disease (incidence of 1 case per 1,000 unexposed persons) that includes 1,000 exposed and 1,000 unexposed persons would probably not be as informativeas a case-control study with 100 cases and 100 controls. Finally, the cost of collecting information will have an impact on both the study design and the number of persons selected. The best study design will yield the maximum information for the lowest cost.

Increase sample size

Increasing sample size will reduce random error by reducing sampling error (lowering the p value) and by obtaining more information. Although a larger sample will lower the variance of the estimates, sample size alone will not overcome the effect of selection, information, or confounding biases.

The p value

The p value is the probability that chance alone is responsible for the differences observed between the comparison groups. For example, suppose a study was designed to evaluate the difference in the probability of infection between two groups of surgical patients treated with two different antibiotics (named *A* and *B*). If the observed difference was significant at a p value of 0.05, then we had a 5% chance of finding that difference when no difference actually existed. Note that the p value is a conditional probability; it assumes that the exposure is not associated with the outcome. For example, if all persons receiving antibiotics A and B truly have the same probability of infection, then a p value of 0.05 indicates that the study has a 5% chance of selecting comparison groups with different infection probabilities. The p value is not, however, the probability that the tested hypothesis is true (Ahlbom and Norell, 1984).

Influence of randomness

The influence of chance or randomness may be viewed from either of two philosophical perspectives. From a deterministic view of reality, the appearance of randomness is really attributable to unmeasured and unknown causal factors affecting the exposure-disease association. If those unknown factors could be perfectly measured, the occurrence of disease would be predictable. From a nondeterministic view, randomness is in itself a real but unknowable source of causation. By this philosophy, we are always faced with some unpredictable influences. When conducting a study, we cannot distinguish between these two potential sources of randomness. The

epidemiologist must instead focus on the size of the group studied and on the quality of the information obtained (Rothman, 1986).

Effect Modification

Effect modification concerns how the presence of an additional risk factor changes the magnitude of the association between a given exposure and a health outcome. We assess effect modification by comparing the magnitude of two or more stratum-specific estimates of effect with each other (e.g., the relative risks for two strata of age). Effect modification exists when these estimates are "substantially" different from each other; that is, the additional risk factor modifies the effect of the exposure on the outcome. Determining whether the difference is substantial can be assessed with a statistical test, however, some authors prefer using a criterion of clinical or practical importance (Rothman and Boice, 1979).

Risk difference example

To illustrate effect modification, we will consider the effect of prenatal care on birthweight and the use of risk difference (RD) (formula 4.4.1) as the measure of effect. In Table 4.14, women who do no receive prenatal care have 0.219 (or 21.9%) higher risk of delivering a low-birthweight child than women who receive prenatal care. However, the crude table obscures the effect of cigarette smoking. Therefore, we should examine the data stratified by smoking status to determine if effect modification (or confounding) is present.

Table 4.14
Crude Table
The Effect of Prenatal Care on Low Birthweight as Measured by the Risk Difference

	Exposure		
Low-Birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	300	81	381
No	700	919	1,619
Total	1,000	1,000	2,000

Risk Difference = (300/1,000) - (81/1,000) = 0.300 - 0.081 = 0.219

Because the RD for the effect of prenatal care is greater among mothers who smoke cigarettes (11.2%) than among mothers who do not smoke cigarettes (3.3%), we know that cigarette smoking modifies the effect of prenatal care (Table 4.15). When effect modification occurs, it is misleading to compute a single summary measure of the association (as in Table 4.14). The summary measure obscures the modifying effect of the additional factor, smoking. Also, depending upon the distribution of the modifier, the particular value of the summary measure will vary from study to study. For example, if the number of women without prenatal care in stratum 2 was increased from 100 to 200 (maintaining the same incidence), the summary RD would decrease from 0.060 to 0.050.

By comparing Tables 4.14 and 4.15, we observe that a factor that modifies the risk difference will not necessarily modify the

Table 4.15

Effect Modification and Confounding of the Effect of Prenatal Care on Low Birthweight

Mothers Who Smoke Cigarettes: Stratum 1

	Exposure		
Low-birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	290	21	311
No	610	79	689
Total	900	100	1,000

Risk Difference = (290/900) - (21/100) = 0.322 - 0.210 = 0.112

Mothers Who Do Not Smoke Cigarettes: Stratum 2

	Exposure		
Low-birthweight Infant	No Prenatal Care	Prenatal Care	Total
Yes	10	60	70
No	90	840	930
Total	100	900	1,000

Risk Difference = (10/100) - (60/900) = 0.100 - 0.067 = 0.033

measures used. Effect modification, however, represents an relative risk. Thus, effect modification depends, in part, on the interaction between the exposure and the modifier; as such, we would expect the same modifier to appear in other studies of a given exposure and disease.

In contrast to effect modification, confounding factors do not depend upon a particular measure of effect nor would they necessarily be observed in other studies. It is possible, however, for a risk factor to be both a confounder and an effect modifier in a single study. As shown in Table 4.14, the crude RD is 0.219; using formula 4.12.1, the adjusted RD from Table 4.15 is 0.060. Thus, these data also illustrate that an effect modifier may also be a confounder.

Reliability

Reliability defined

Reliability (or precision) concerns the degree to which the values obtained from repeated measurements of the same persons—under similar conditions—will yield the same result. For example, a large range in the distribution of averages implies a lack of reliability even though the average of several averages may equal the true value. We must be sure that the characteristic of the persons being measured is constant (e.g., height) and thus not a source of the variation. Reliability is distinguished from bias in that the average of several biased measurements will still differ from the true value.

Situations that may produce low reliability

Several situations are likely to produce low reliability. As described in the section on errors attributed to chance, small sample sizes may produce low reliability because of sampling error. Another source of low reliability is data collected within a nonstandardized or poorly structured manner. Examples of poorly structured data collection include poorly worded questions in a questionnaire, poorly trained interviewers, or poorly trained data abstractors. Poorly trained interviewers and abstractors probably would not obtain the same answer when collecting the data on a second attempt. Imprecision during data collection tends to produce nondifferential (random) misclassification of information, a type of information bias that reduces the chance of finding differences between compared groups.

Validity

Validity defined

The validity (or accuracy) of a study concerns the extent to which inferences may be made from the study. Epidemiologists usually refer to two types of validity— internal and external.

Target shooting may be used as a simple illustration of the relationship between reliability and validity (Figure 4.16) (Ahlbom and Norell, 1984). A good marksman, unaware that the gun used was not aligned, would hit the target with minimum scatter (high reliability) but systematically off-center (low validity). The individual shots of another marksman could scatter the shots about the center (low reliability) but do so in an evenly distributed pattern (high validity).

Internal Validity

By internal validity, epidemiologists refer to the lack of bias among the groups compared within a given study. Except for random errors, an internally valid study accurately represents the effect of the exposure on the disease outcome among the study subjects. If any of the three major categories (selection, information, confounding) of bias exist, the internal validity of that study is decreased.

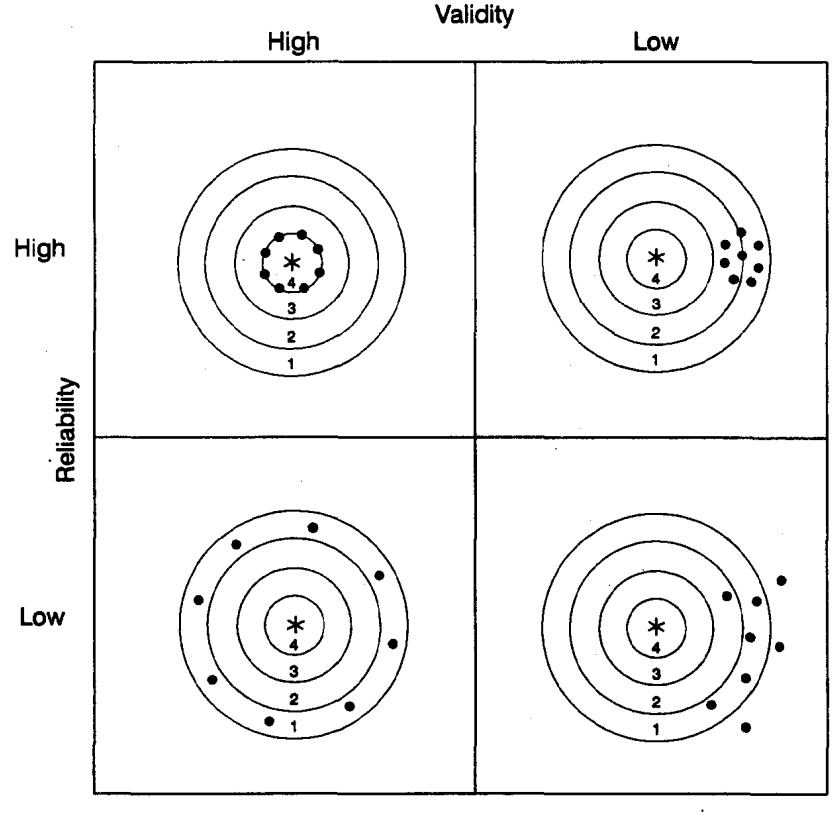
External Validity

External validity concerns the generalizability of a study's results beyond the particular groups participating in the study. As such, generalizability primarily concerns how well the study population represents other populations. Before examining external validity, however, the study comparisons must be considered internally valid. Some degree of generalizability which extends the limits of knowledge is typically the goal of scientific research. If, however, our particular study was primarily concerned with limited and unique populations, we might not be so concerned about broad generalizations.

Few studies, if any, are entirely free of bias. When evaluating the validity of a study, it is important both to identify biases and to estimate how they may have distorted the study's findings. Although biases exist, they may not be serious enough to appreciably change the conclusions of a study. Sensitivity analysis is one way of assessing the influence of potential biases on a particular study's conclusions.

Figure 4.16

Target Shooting As an Example of
High and Low Reliability and Validity



In this context, sensitivity refers to the degree to which a study's findings may be changed if certain parts of the obtained data were different.

In Example 4.17, 50% of mothers of low-birthweight infants reported a history of smoking, but only 20% of mothers with normal- or high-birthweight infants reported smoking; the resulting odds ratio was 4.0. If we believed that mothers of low-birthweight babies may have tended to overreport smoking, we could use an estimate of how much they overreported to calculate how much

Example 4.17

Maternal Smoking and Low Birthweight

Problem: Is maternal cigarette smoking associated with low birthweight?

Research Hypothesis: Maternal cigarette smoking increases the risk of low birthweight.

Study Design: Case-control study

Outcome: Low-birthweight infant

Smoking Status as Originally Reported*

	Maternal Smoking Status		
Low-Birthweight Infant	**Yes**	**No**	**Total**
Yes	50	50	100
No	20	80	100
Total	70	130	200

OR = 4.0

* Reflects overreporting of smoking among 100 mothers with low-birthweight babies.

Sensitivity Analysis after Correcting Original Results†

	Maternal Smoking Status		
Low Birthweight Infant	**Yes**	**No**	**Total**
Yes	40	60	100
No	20	80	100
Total	60	140	200

OR = 2.7

† Reflects corrected smoking status for the ten mothers who had low-birthweight infants and who incorrectly reported that they were smokers.

the odds ratio (OR) would change. For example, if we estimated that 10% of the mothers of low-birthweight babies incorrectly reported that they smoked (whereas the mothers of other babies accurately reported their smoking), the OR would change from 4.0 to 2.7. Although the results may have overestimated the effect of smoking, we would have found that low birthweight was still positively associated with smoking. By substituting *worst case* numerical estimates for the effect of potential biases, we can better understand how much the conclusions of a study may be changed by biases of a study.

Causality

Causality concerns judging whether a particular exposure causes a particular disease. This is a subjective process—particularly in epidemiology, where the epidemiologist typically has no control over the sequence of events in the natural history of a disease. In an attempt to establish a systematic way of determining causality, several authors have cited the following seven criteria for evaluating whether a particular exposure causes a particular outcome:

Criteria for causality

Proper temporal sequence. To have caused a particular outcome, an exposure must temporally precede that outcome. This tenet may be difficult to establish for chronic diseases or diseases with a long time interval between exposure and occurrence.

Experiment. Randomized assignment of an exposure results in a greater rate of the expected health outcome in the exposed group than in the unexposed group. Although this tenet provides the strongest evidence of causation, we can ethically use randomized assignment in only a few epidemiologic studies.

Dose-response. Higher levels of exposure are associated with higher rates of disease; for example, higher rates of lung cancer have been observed with an increasing number of cigarettes smoked daily.

Strength of association. The larger the measure of effect, the more likely the relationship is causal. This tenet assumes that smaller measures of effect are more likely to be attributable to bias.

Consistency of association. Similar associations between exposure and disease are observed among different studies. If similar associations are repeatedly observed, it is unlikely that they are all spurious.

Collateral evidence and biological plausibility. The likelihood that an exposure causes a particular disease is supported by evidence such as trends in vital statistics and animal or human biological models.

Specificity of association. Different studies report the same specific disease associated with a particular exposure.

Of these seven criteria, only the first is absolutely required for causation. All the remaining tenets support the attribution of causality but are not essential to causality (Kleinbaum et al., 1982; Hill, 1965). With a few exceptions, the last criterion is probably the least often fulfilled because most diseases have multifactorial etiology and because many risk factors, most notably cigarette smoking, may influence several diseases.

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F Analytic studies typically test hypotheses.

   (b) T/F The major difference in descriptive and analytic epidemiology is the way study subjects are selected and data are collected.

   (c) T/F Descriptive epidemiology focuses on measures of effect.

   (d) T/F Analytic epidemiology emphasizes hypothesis testing as opposed to hypothesis generation.

   (e) T/F Analytic epidemiology describes frequency of disease occurrence in relation to person, place, and time.

2. Distinguish descriptive from analytic studies by putting a (D) for *descriptive* or an (A) for *analytic* next to each of the following:

   ___ 2.1 Etiologic
   ___ 2.2 Tests hypotheses
   ___ 2.3 Shows patterns regarding person, place, and time
   ___ 2.4 Compares two or more groups
   ___ 2.5 Generates hypotheses
   ___ 2.6 Evaluates the association between exposure and disease
   ___ 2.7 Gives measures of effect

3. Select one response.

   3.1 Any of the following may be included as an "exposure" in health research except:

   (a) A potential risk factor
   (b) The outcome being studied
   (c) The treatment or therapy
   (d) Any potential event before the *disease*
   (e) A health problem

3.2 Any of the following may be included as a *disease* or health outcome in health research except:

(a) Illness or disability
(b) Improved health
(c) Hereditary difficulties
(d) Death
(e) Exposure

3.3 Which of the following strategies cannot be used in cohort and case-control studies to eliminate bias?

(a) Random selection process to select study subjects
(b) Same standardized instruments on both groups
(c) "Blind" observer who will do data collection
(d) Study subjects selected according to predetermined criteria
(e) Size of study groups increased

3.4 As a strategy to eliminate bias, you can make sure cases and controls are comparable with regard to all but one of the following:

(a) How they are selected
(b) How they are interviewed
(c) How they are exposed
(d) How they are influenced by other risk factors

3.5 Which of the following is not a strategy to eliminate bias?

(a) Use a validated questionnaire
(b) Use measures of effect
(c) Write a protocol for screening
(d) Restrict study subjects
(e) Use stratified analysis

4. Matching. Select A (Clinical trial), B (Cohort), or C (Case-control) to indicate the type of study design that is being described.

___ 4.1 Collects information on diseased and nondiseased study subjects
___ 4.2 Selection is based on disease status.
___ 4.3 Assesses exposure of each study subject and then observes occurrence of the health outcome
___ 4.4 Called a follow-up design because study subjects are followed to determine outcome
___ 4.5 Randomized assignment of exposure is a characteristic of this type of design.
___ 4.6 Uses selection based on exposure and then observes subsequent occurrence of the health outcome

5. Circle true (T) or false (F).

(a) T/F Misclassification may falsely increase a positive association.

(b) T/F Misclassification may increase the chance of finding no association.

(c) T/F A confounding factor must be 1) an independent risk factor, 2) associated with exposure, and 3) not in the pathway from exposure to disease.

(d) T/F Confounding may be removed by opening up restrictions on participation.

(e) T/F Matching may eliminate confounding but introduce other problems.

6. Matching. Select the appropriate letter to indicate the type of bias that is described in each example below (more than one answer may apply):

   a. Selection bias
   b. Information bias
   c. Confounding bias
   d. Effect modification
   e. Reliability

__ 6.1 Information is obtained from different study subjects by different processes.

__ 6.2 Observation, classification, and measurement bias occurs.

__ 6.3 Has an effect on odds ratios, relative risks, and risk differences.

__ 6.4 Only a subset of potential study subjects survives to become enrolled in the study (survival bias).

__ 6.5 Recall bias, interviewer bias, or bias attributable to missing data or prevarication occurs.

__ 6.6 In case-control studies, factors associated with exposure influence study participation.

__ 6.7 Shows how a third factor changes the true effect of exposure on disease.

__ 6.8 The effects of two risk factors are mixed.

__ 6.9 If you can identify its presence, you can analyze and eliminate the problem.

__ 6.10 One group provides information on questionnaires and the other from medical records.

__ 6.11 Error in classification of study subjects occurs with regard to exposure or disease or both.

___ 6.12 A difference between a crude result and a stratified result indicates its presence.

___ 6.13 Repeated administration of the same questionnaire to the same person yields different answers.

7. Circle true (T) or false (F).

(a) T/F The sensitivity is the proportion of people free of the disease who will be classified correctly by the screening test.

(b) T/F The specificity is the proportion of people who have the disease who will be classified correctly by the screening test.

(c) T/F When the sensitivity and specificity equal 1.0, misclassification results.

(d) T/F If p = 0.05, we have a 5% chance of finding a difference between the comparison groups when no difference really exists.

(e) T/F The p value is the probability that the tested hypothesis is true.

(f) T/F Increasing sample size will increase the p value.

(g) T/F Both effect modification and confounding concern the way a third variable affects the relationship between exposure and disease.

(h) T/F If you retest the same person and get the same result each time, you have reliability.

8. Matching. Select the appropriate letter to match the descriptions on the left with the terms on the right. Suppose you take repeated measurements on the same person and. . .

   a. High reliability
   b. Low reliability
   c. Bias
   d. Validity
   e. Causality

   ___ 8.1 Get the same result.
   ___ 8.2 Get the same result, but later find out the tool is inaccurate.
   ___ 8.3 Get different results, but the average measure is accurate.

9. Circle true (T) or false (F).

   (a) T/F A study with internal validity accurately reflects the effect of exposure on disease outcome among study subjects.

   (b) T/F Bias increases the internal validity of a study.

   (c) T/F External validity depends on internal validity.

   (d) T/F External validity is less of a consideration when working with unique populations.

   (e) T/F Because most epidemiologic studies contain bias, we are limited in the generalizability of the findings from a single study.

   (f) T/F Numerical estimates can be given to biases to determine their impact on the risk ratio or risk difference using sensitivity analysis.

10. Matching. Match the descriptions in 10.1 - 10.5 with the single best choice of factors (a - f) to consider in evaluating evidence of a causal relationship in an epidemiologic study.

a. Proper temporal sequence
b. Dose-response
c. Strength of association
d. Consistency of association
e. Collateral evidence and biological plausibility
f. Specificity of association

___ 10.1 Absolutely required to show causality.

___ 10.2 The larger the measure of effect, the more likely the relationship is causal.

___ 10.3 Similar associations have been observed across numerous other studies.

___ 10.4 Higher levels of exposure are associated with higher rates of disease.

___ 10.5 Least often fulfilled.

Suggested Answers To Practice Exercises

1. True or false.

   (a) T Introduction

   (b) F Study Design. No, descriptive studies generate hypotheses and analytic studies test hypotheses.

   (c) F Study Design. No, analytic epidemiology focuses on measures of effect.

   (d) T Study Design

   (e) F Study Design. No, descriptive epidemiology describes the frequency of disease occurrence in relation to person, place, and time.

2. Descriptive or analytic.

   2.1 a Study Design

   2.2 a Study Design

   2.3 d Study Design

   2.4 a, d Study Design

   2.5 d Study Design

   2.6 a Study Design

   2.7 a Study Design

3. Select one response.

3.1	b	Introduction
3.2	e	Introduction
3.3	e	Bias
3.4	c	Bias
3.5	b	Bias

4. Matching.

4.1	a/b/c	Study Design. Case-control studies select subjects by disease status and then collect exposure data; however, a and b are also correct, since scientists collect data on disease status of participants for clinical trials and cohort studies.
4.2	c	Study Design
4.3	b	Study Design
4.4	a/b	Study Design. A and B; both involve follow-up.
4.5	a	Study Design
4.6	b	Study Design

5. True or false.

(a)	T	Bias
(b)	T	Bias
(c)	T	Bias

(d) F Bias. No, confounding may be *reduced* by *making* restrictions on participation.

(e) T Bias

6. Matching.

6.1 b Bias

6.2 b Bias

6.3 a/b/c/d Bias and Effect Modification

6.4 a Bias

6.5 b Bias

6.6 a Bias

6.7 d Effect Modification

6.8 c Bias

6.9 c Bias

6.10 b Bias

6.11 b Bias

6.12 c/d Bias. Confounding is present; effect modification may or may not be present.

6.13 e Reliability

7. True or false.

(a) F Information Bias. No, the specificity is the proportion of people free of disease who will be classified correctly by the screening test.

(b) F Information Bias. No, the sensitivity is the proportion of people who have the disease who will be classified correctly by the screening test.

(c) F Information Bias. No, when the sensitivity and specificity are not equal to 1.0, misclassification results.

(d) T Chance

(e) F Chance. No, the p value is the probability of rejecting the null hypothesis when the null hypothesis is true (i.e., the probability of finding a difference when no difference exists).

(f) F Chance. No, increasing sample size will decrease the p value.

(g) T Bias and Effect Modification

(h) T Reliability

8. Matching.

8.1 a Reliability

8.2 a/c Bias

8.3 b/d Reliability

9. True or false.

9.1 T Validity

9.2 F Validity. No, bias decreases the internal validity of a study.

9.3 T Validity

9.4 T Validity

9.5 T Validity

9.6 T Validity

10. Matching.

10.1 a Causality

10.2 c Causality

10.3 d Causality

10.4 b Causality

10.5 f Causality

References

Ahlbom A, Norell S. Introduction to modern epidemiology. Chestnut Hill, Massachusetts: Epidemiology Resources, 1984.

Bayley N. Bayley Scales of infant development. New York: Psychological Corporation, 1969.

Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.

Hill, AB. The environment and disease association or causation. Proc Royal Soc Med 1965;58:295-300.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, Massachusetts: Lifetime Learning Publications, 1982.

Last JM, ed. A dictionary of epidemiology. New York: Oxford University Press, 1983.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-748.

Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976;103:226-235.

Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Bethesda, Maryland: National Institutes of Health, 1979; DHEW publication no. (NIH) 79-1649.

Rothman KJ. Modern epidemiology. Boston: Little, Brown and Company, 1986.

Sackett, DL. Bias in analytic research. J Chronic Disease 1979;32:51-63.

Chapter 5 - Learning Objectives

After completing this chapter, you should be able to:

1. Identify constraints on the sample size of a study.

2. Identify the terms:

   null hypothesis
   alternative hypothesis
   level of significance, Type I error, α
   Type II error, β
   power of test, $1-\beta$

3. Recognize the relationship between hypothesis testing and sample size and statistical power.

4. Identify purposes of sample size and power calculations.

5. Distinguish one-sided from two-sided comparisons.

6. Identify information needed for calculation of sample size for each type of study.

7. Calculate sample sizes for:

   descriptive studies
   randomized clinical trials
   cohort studies
   case-control studies

8. Calculate statistical power for:

   randomized clinical trials
   cohort studies
   case-control studies

9. Interpret power calculations.

5 Sample Size and Power

Introduction

Importance of appropriate sample size

One of the most important issues in designing a study is appropriate sample size. A sample size that is too small will limit the conclusions that can be made from the study, and a sample that is too large will result in unnecessary expenditures of time, money, and effort to reach conclusions that could have been made with fewer subjects. We can rely on statistical theory to compute the smallest sample size to properly achieve the objectives of the study. Because of practical considerations, the nature of the study, and the resources available, the sample size we ultimately choose will usually be a compromise between what is needed to satisfy the statistical requirements and what can realistically be accomplished.

When designing a study, we must apply statistical formulas to determine the required number of participants to achieve the specified statistical power or the reverse—to determine the statistical power provided by a specified sample size. Statistical hypothesis testing is the foundation for determining the sample size and for assessing statistical power.

Hypothesis Testing

The Null Hypothesis

The null hypothesis defined

To make inferences, researchers use the principles of hypothesis testing. Any study tries to prove or disprove a research hypothesis or hunch about exposures to potential risk factors or the efficacy of different treatments. To test the hypothesis, the researcher assumes that no difference exists between exposures to risk factors or treatments under study. This assumption is called the null hypothesis. Examples of null hypothesis statements for different types of studies are presented below:

- Randomized clinical trial: The new treatment has the same proportion of effectiveness as the control treatment.
- Cohort study: The proportion of the exposed individuals who develop the outcome of interest is the same as the proportion of unexposed individuals who develop the outcome of interest.
- Case-control study: The proportion of cases exposed to the potential risk factor under study is the same as the proportion of exposed controls.

The alternative hypothesis

The hypothesis test ascertains the validity of the null hypothesis against an alternative hypothesis from the evidence gathered in a study. For example, an alternative hypothesis for the cohort study cited above might state that the proportion of exposed individuals who develop the outcome of interest is higher than the proportion of unexposed persons who develop the outcome. Randomized clinical trials, case-control studies, and cohort studies are designed to seek evidence to either support or reject the null hypothesis in favor of the alternative hypothesis. The specifications of the null and alternative hypotheses form the basis for determining sample size.

Tests of Hypotheses

Decision making is an integral component of hypothesis testing. The researcher must use the data gathered from the study to determine one of two possible decisions—to accept or to reject the null hypothesis. The researcher should be aware of two types of errors that can affect this decision (Table 5.1).

Alpha level (α)

Type I error. If we reject a null hypothesis that is actually true, but the study data indicate that it is false, we have made a type I error. The chance of making this error is called the *level of significance* of the study, or the alpha (α) level.

Beta level (β)

Type II error. If we accept a null hypothesis that is actually false, but the study data indicate that it is true, we have made a type II error. The chance of making the type II error is called the beta (β) level.

Table 5.1

Hypothesis Testing and Possible Outcomes

Study Decision	Null hypothesis: True	Null hypothesis: False
Accept the null hypothesis	Correct decision	Type II error
Reject the null hypothesis	Type I error	Correct decision

Power and Confidence Level

Power defined ($1-\beta$)

The chance of correctly rejecting a null hypothesis when it is indeed false is called the statistical power ($1-\beta$) of the test of the hypothesis. Power is the probability of detecting a difference in exposure levels or treatments when a difference actually exists.

Confidence level defined ($1-\alpha$)

The probability of accepting the null hypothesis when it is true is often called the confidence level ($1-\alpha$) of the test. The confidence level is the probability of finding no difference in exposure levels or treatments when no difference exists.

When a decision is made to accept or reject the null hypothesis, the actual state of nature is not known. Regardless of how small α and β are set, we are always faced with an element of chance, and the possibility that we have made the wrong decision always exists. A decision to accept the null hypothesis does not prove that it is true; a decision to reject the null hypothesis does not prove that it is false.

Sample Size and Power

The null hypothesis is rejected when the significance level based on the study data exceeds the significance level (α) established to test the hypothesis. To guard against a type I error, the level of significance is typically set at a suitably small probability, such as $\alpha = .05$. A given statistical test is more likely to distinguish minor differences in treatments or exposure levels as the sample size increases; and hence, we are more likely to reject the null

hypothesis. We should impose some limits on sample size to ensure that only meaningful or practical differences are detected by statistical testing.

Conversely, if the null hypothesis is actually false, the likelihood that we correctly reject it (i.e., the power of the test, $1-\beta$) depends primarily on sample size. The larger the sample size, the more power a study has. A study with sufficient power enables us to detect a difference in treatments or exposure levels when a difference actually exists. If a study possesses inadequate power, a result that does not achieve significance (i.e., the null hypothesis is accepted) may just as well be attributed to the lack of power as to the fact that the null hypothesis is correct. Typically, the statistical power, $1-\beta$, is chosen between .80 and .95.

The power of a study is the key factor in interpreting results that indicate there is not a clinically important difference in two treatments or exposure levels. What constitutes a clinically important difference is typically determined by the researcher and established at the onset of a study. In summary, both the significance level (α) and the power of the study ($1-\beta$) are important in determining sample size.

Determining the Size of a Sample

One of the first questions typically asked when designing a reproductive health study is, "How many subjects should be selected for the study?" To answer this question, researchers must first answer other questions that provide information about what they expect to achieve from the study. In essence, the study must be set up in the framework of a hypothesis test. The sample size is then determined by applying statistical methodology that quantifies the relationship between the various parameters that describe the hypothesis test under consideration. The researchers must provide the parameters that will be used in the formulas.

In Example 5.2, we present the information required to calculate the sample size for a randomized clinical trial. From this information, the sample size we need for the clinical trial can be calculated with $\alpha = 0.05$, $1-\beta = 0.90$, $p_0 = 0.60$, $d = 0.10$, $f = 0$. These statistical techniques will be described later in this chapter. Similar information needs to be provided for all study designs.

Example 5.2

A Two-Sided Comparison
Required Information to Calculate the Sample Size for a Randomized Clinical Trial

Estrogen has been associated with most of the undesirable side effects of the combined oral contraceptive (OC) pill. Lower doses of estrogen should decrease the frequency and duration of side effects, such as nausea, headache, and vomiting. A randomized clinical trial was designed to compare the rate of continued use among women who took a standard-dose oral contraceptive (1.0 mg norethindrone with 50 mcg mestranol) and the rate of continued use among women who took a low-dose oral contraceptive (1.0 mg norethindrone with 35 mcg ethinyl estradiol). In designing the study, several questions were considered to determine the appropriate sample size.

What is the primary purpose of the study (Statement of hypothesis)?

To see if women who take a low-dose OC (new treatment) have a higher rate of continued use than women who take a standard-dose OC (control treatment). The reasons that women discontinue using OCs are of interest but are not the primary focus of the study. The null hypothesis states that there is no difference in the rate of continued use between low-dose users and standard-dose OC users.

What is the primary outcome measure?

Continued use of the assigned OC for the first 12 months of the study.

What level of outcome is expected with the standard-dose OC (proportion in control group, denoted p_0)?

Previous studies have shown that at 12 months the rate of continued use for the control treatment is 60% ($p_0 = .60$). The rate of discontinuation at 12 months would therefore be 40%. For this example, we have chosen to discuss continued use, although discontinuation would have been just as appropriate.

How small a difference (d) in the outcome between treatments is it important to detect and with what statistical power?

An increase in the rate of continued use is anticipated for low-dose OC users. However, the rate may be lower for standard-dose users because low-dose users may experience breakthrough bleeding. We wish to detect as statistically significant a difference in continuation rates of 10 percentage points ($d = 0.10$). That is, based on the expected rate of continued use of 60% for

Example 5.2 (continued)

women in the standard-dose group, we wish to find the continuation rate for the low-dose users to be statistically significantly different as small as 50% or as large as 70%. Further, there should be a 0.90 (or 90%) chance (i.e., power of $1-\beta = 0.90$) that a difference in continuation rates as great as 10 percentage points be detected as statistically significant.

If the two treatments really don't differ with regard to outcome, what risk are we willing to take (α) that the study will find the treatments to be significantly different?

We are willing to risk at most an $\alpha = 0.05$, or 5% chance, that a difference in the rates of continued use will be detected when a difference actually does not exist.

What proportion f of the women who are admitted to the study will drop out before the study ends for reasons other than the outcome under study?

Because continued use is the outcome under study, all reasons for leaving the study would be considered discontinuations of OC use.

(Adapted from Basnayake et al., 1983)

One-sided and two-sided comparisons

When comparing a new treatment and a control treatment, exposed and unexposed individuals, or case subjects and controls, researchers may only be interested in detecting differences in one direction. For example, we might want to determine if the new treatment is better than the control treatment, if exposure to the risk factors under study increases the chances of developing the outcome of interest, or if more cases than controls were exposed to the risk factors. These examples illustrate one-sided comparisons. In some studies, researchers may be interested in detecting whether one group differs from the other by a certain level, regardless of the direction of the difference. This type of comparison is called a two-sided comparison. Example 5.2 illustrates a two-sided comparison: the researchers want to determine if the rate of continued use of oral contraceptives (OC) among women who take a low-dose oral contraceptive is greater than or less than the rate of continued use among women who take the standard-dose oral contraceptive. The type of comparison we wish to make influences the computation of

Table 5.3

Z_α for Sample Size Formulas for Selected Values of α

	Comparison	
	One-Sided	Two-Sided
α	Z_α	Z_α
0.10	1.28	1.65
0.05	1.65	1.96
0.025	1.96	2.24
0.01	2.33	2.58

sample size through the Z statistics. Table 5.3 provides the Z statistics for both one-sided and two-sided comparisons.

Adjusting the Sample Size for Nonresponse

If information will be collected directly from the study subjects, we must consider the issue of nonresponse when estimating sample sizes. Nonresponse may be attributed to a variety of reasons, including refusal to participate or to continue in the study, inability to locate the study participants, and incorrect information on sampling lists. Sample size formulas represent the number of subjects for whom complete information must be obtained and not the number of subjects who should be selected for the study. To satisfy the confidence level $(1 - \alpha)$ and tolerance specified for the study, researchers should estimate the expected level of nonresponse and inflate the estimated sample size accordingly.

To adjust the sample size estimates for nonresponse, sample size formulas should be multiplied by a factor of:

Nonresponse adjustment formula

(5.3.1)

$$q = \frac{1}{1 - f}$$

where f = estimated nonresponse rate.

Other Considerations for Determining Sample Size

The statistical theory used to derive sample size estimates includes several assumptions. If some of these assumptions are not actually met, the formula for the sample size may not be valid. The sample size determination is most strongly influenced by two of these assumptions: 1) the population from which the sample is drawn is infinitely large, and 2) the sample is selected by a simple random sampling process. In reality, both of these assumptions are often violated. However, we can apply adjustments to correct the estimated sample size in studies for which one or both assumptions are violated.

Although the assumption of an infinitely large population from which to draw the sample will never be true, if the size of the population of interest is large relative to the sample size, we can use the standard sample size formulas to calculate accurate estimates of the required sample size. If the size of the population of interest is not large, the finite population correction (fpc) must be applied. As a general rule, the need only be applied if the sample comprises more than 5% to 10% of the population. The fpc, which yields the adjusted sample size n′, is calculated as:

Finite population correction formula (fpc)

(5.3.2)

$$\text{fpc} = n' = \frac{n}{1 + n/N}$$

where n = sample size and N = size of population of interest.

If the size of the population of interest is not precisely known, a guess about the size should be sufficient. Applying the fpc will reduce the sample size estimate. Therefore by ignoring the fpc when it is needed, researchers may be using a larger sample size than is actually required to achieve the desired accuracy. The fpc is applicable more often for descriptive studies and for the selection of controls in case-control designs.

The sample size formulas are derived under the assumption that simple random sampling is used to draw a sample from a population of interest. Because of practical concerns, such as cost, time, and the resources required to conduct the study, a simple random sample is seldom possible. Chapter 7 describes various types of sampling designs and discusses when each type should be used. If a sampling design other than simple random sampling is planned for a study, the

specific design needs to be considered when computing sample size estimates. No simple formula is available for computing the design effect adjustment.

Types of Studies and Sample Size Requirements

Descriptive Studies

In a descriptive study, we are interested in obtaining an estimate of the proportion of the population that possesses or develops a particular health outcome or attribute. Assessing the total population is frequently not possible, so we must draw a sample from the population. If the sample is constructed by applying statistical principles, we can use statistical theory to make inferences about the proportion of the entire population with respect to some attribute or outcome based on the proportion of individuals from the sample who have the attribute or outcome.

In descriptive studies, we have no hypotheses to test. The only decision that would effect sample size is how accurate we want the estimate to be. In this situation an approach based on the concept of confidence intervals is applied to determine sample sizes.

Once a proportion is obtained from measuring the study sample, we can construct a confidence interval about the sample proportion. The researcher must specify a confidence level for the interval. The confidence level is the probability that the confidence interval contains the true population proportion. The higher the confidence level, the wider the confidence interval. Again, α denotes the probability of making an error (i.e., constructing an interval that does not contain the true proportion). Once these values are specified, we can apply statistical procedures, based on the concepts of confidence intervals, to calculate the appropriate sample size. The higher the confidence level, the larger the sample size.

The size of the sample depends on the following information:

d = The distance (or tolerance) — how close to the

proportion of interest the estimate is desired to be (e.g., within 0.05).

$1-\alpha$ = The confidence level that our estimate is within distance (d) of the proportion of interest ($1-\alpha$ is expressed as a decimal, for example, 0.95).

p = Proportion or a best guess about the value of the proportion of interest. If we have no information about p, then p = 0.5 is a conservative estimate.

Sample size formula for a descriptive study

Using this information, the required sample size (n), for a descriptive study is calculated as:

(5.3.3)

$$n = p * (1 - p) * (Z_\alpha / d)^2$$

where α = (1 − confidence level).

Because the confidence interval extends for a distance (d) on both sides of the sample proportion, we may always find the value Z_α for descriptive studies in the column of two-sided comparison (Table 5.3). In Example 5.4, if the investigators had been willing to be within 0.001 of the actual proportion instead of 0.0005, the required sample size would have been $n = 0.0045 * (0.9955) * (1.96/0.001)^2 = 17{,}209.4 \simeq 17{,}210$. By allowing the estimate to be less precise, they could have lowered the number of medical records that needed to be sampled. In the actual study, the investigators sampled approximately 225,000 medical records.

Randomized Clinical Trials

In randomized clinical trials, we compare the proportion of participants with a specified outcome in a group that receives treatment A and the proportion of participants with the same outcome in a group that receives treatment B. These two treatments are referred to as the *new treatment* and the *control treatment*, which is also known as the *standard treatment* or the *placebo treatment*. The size of the sample in each group depends on the following information:

Example 5.4

Computing the Sample Size for a Descriptive Study

A study was designed to estimate and compare the 1970 and 1978 cumulative incidence of ectopic pregnancy in the United States. A sample of medical records was reviewed to abstract demographic data, final diagnoses, and surgical procedures. In 1970, the cumulative incidence of ectopic pregnancy was estimated at 4.5 per 1,000, or 0.0045, reported pregnancies. The investigators wanted to have a confidence level of 0.95 that their estimate was within 0.0005 of the actual value. The required sample size for this study was calculated using the following information:

$d = 0.0005$

$1-\alpha = 0.95$, so $\alpha = 0.05$. Therefore, $Z_\alpha = 1.96$ (see $\alpha = 0.05$ in Table 5.3 under the two-sided comparison).

$p = 0.0045$, the proportion estimated in 1970 was chosen as a guess for the 1978 proportion.

$$\text{then } n = p * (1 - p) * (Z_\alpha / d)^2$$

$$= 0.0045 * 0.9955 * (1.96 / 0.0005)^2$$

$$= 68{,}837.6 \simeq 68{,}838$$

p_0 = The proportion of the participants in the control treatment group who are expected (an educated guess) to exhibit the outcome of interest.

p_1 = The proportion of the participants in the treatment group that are expected to exhibit the outcome of interest. This proportion is usually set relative to p_0. The investigator desires to detect p_1 as being different from p_0.

For example, a commonly used spermicide had an annual failure rate of 16% (16 per 100 women who used the spermicide became pregnant, or 0.16). The researchers would like to compare this spermicide and a new spermicide to determine if the failure rate is reduced by half (i.e., to 8%).

α = The level of significance or type I error probability.

β, $1-\beta$ = β is the type II error probability and $1-\beta$ is the statistical power of the comparison. Usually β or $1-\beta$ is provided and either one may be calculated from the other.

f = The proportion of study subjects who are expected to leave the study for reasons other than the outcome under investigation. *f* is used to inflate the sample size to allow for individuals who drop out of the study before the end of the study period. The level of *f* can often be estimated from prior experience with similar treatments and in similar research settings. The sample size adjustment factor for nonresponse or dropout is $q = 1/(1-f)$.

Formula 5.4.1 determines the required number of participants for each treatment group in a randomized clinical trial.

Sample size formula for a randomized clinical trial

(5.4.1)

$$n = \frac{1}{(1-f)} * \left[\frac{2 * (Z_\alpha + Z_\beta)^2 * p * (1-p)}{(p_0 - p_1)^2}\right]$$

where

$p = (p_0 + p_1)/2$

Z_α = value from Table 5.3

Z_β = value from Table 5.5

In Example 5.6, we used formula 5.4.1 to calculate the required sample size. For this randomized clinical trial, 520 women should receive standard-dose OCs and 520 women should receive low-dose OCs.

Table 5.5

Z_β for Sample Size Formulas for Selected Values of Power $(1-\beta)$ and β

β	$1-\beta$	Z_β
0.50	0.50	0.00
0.40	0.60	0.25
0.30	0.70	0.53
0.20	0.80	0.84
0.15	0.85	1.03
0.10	0.90	1.28
0.05	0.95	1.65
0.025	0.975	1.96
0.01	0.99	2.33

Example 5.6

Computing the Sample Size of a Randomized Clinical Trial

In Example 5.2, we compared the rates of continued use among women who used standard-dose oral contraceptives (OCs) and among those who used low-dose OCs. The following information was used to calculate the required sample size:

p_0 = 0.6 of women will continue to use the standard-dose oral contraceptive for 12 months.

p_1 = 0.5 or 0.7 of women will continue to use the low-dose oral contraceptive for 12 months (two-sided comparison).

α = 0.05

β = 0.10, power = $1-\beta$ = 0.90

f = 0; all women who dropout are considered discontinuations.

With the two-sided comparison, we have two possible values for p_1. To be certain that our sample size is sufficiently large, we always use the value of p_1 that is closest to 0.5 because this value will yield the larger sample size. The required sample size for this randomized clinical trial is:

$$n = \frac{1}{(1-0)} * \left[\frac{2 * (1.96 + 1.28)^2 * 0.55 * 0.45}{(0.6 - 0.5)^2}\right] = 520 \text{ per group}$$

Example 5.6 (continued)

where

Z_α = 1.96 (see Table 5.3 under $\alpha = 0.05$ in the two-sided comparison column)

Z_β = 1.28 (see Table 5.5 under $\beta = 0.10$ or $1-\beta = 0.90$)

$p = (p_0 + p_1)/2 = (0.6 + 0.5)/2 = 0.55$

$f = 0$

Cohort Studies

In cohort studies, we compare the proportion of participants with a specified outcome who are exposed to a potential health risk (sometimes a potential health benefit) and the proportion of participants with the same outcome who are not exposed to the potential health risk. The two study groups are called the exposed and unexposed groups.

Exposed and unexposed groups

The sample size formula for cohort studies is similar to the formula used for clinical trials. The calculation depends on the following information:

p_0 = The proportion of the participants in the unexposed group who are expected to exhibit the outcome of interest.

p_1 = The proportion of the participants in the exposed group w ho are expected to exhibit the outcome of interest. This proportion is usually set relative to p_0, and the investigator desires to detect p_1 as being different from p_0.

For example, we could compare women with dysmenorrhea after sterilization surgery and nonsterilized women with dysmenorrhea who use barrier methods. Previous studies indicate that dysmenorrhea is reported by about 10% of women who use barrier methods over a period of six months. The researchers would like to detect a doubling (i.e., 20%) of the risk of dysmenorrhea, if it exists, among the women who have had sterilization surgery.

α = The level of significance, or type I error probability.

β, $1-\beta$ = The type II error probability and power, respectively.

f = The proportion of study subjects who are expected to leave the study for reasons other than the outcome under investigation q = $1/(1-f)$.

Sample size formula for a cohort study

A formula for determining the required number of participants for the exposed and for the unexposed group in a cohort study is calculated as:

(5.6.1)

$$n = \frac{1}{(1-f)} * \left[\frac{2 * (Z_\alpha + Z_\beta)^2 * p * (1-p)}{(p_0 - p_1)^2}\right]$$

where

$p = (p_0 + p_1)/2$

Z_α = See Table 5.3

Z_β = See Table 5.5.

In Example 5.7, we used formula 5.6.1 to calculate the required sample size for this cohort study. The investigator must enroll 244 women who had had sterilization surgery, and 244 nonsterilized women who had used barrier methods for contraception.

Case-Control Studies

As with cohort studies, case-control studies are designed to determine whether individuals who are exposed to a health risk are more likely to exhibit the outcome of interest than individuals who are not exposed to the health risk. In case-control studies, we compare the exposure histories of individuals who have had the outcome (cases) and individuals who have not had the outcome (controls).

Example 5.7

Computing the Sample Size of a Cohort Study

We compared dysmenorrhea among women who had had sterilization surgery and nonsterilized women who had used barrier methods. Assume that dysmenorrhea occurs in about 10% of women who have used barrier methods. Further, suppose the investigator wants to be able to detect a doubling of this level, if it exists among sterilized women with a statistical power of 0.90. The level of significance is set at 0.05; up to 10% of the participants are expected to drop out. The required sample size for each group was calculated using formula 5.6.1:

$$n \quad \frac{1}{(1 - 0.1)} * \left[\frac{2 * (1.65 + 1.28)^2 * 0.15 * 0.85}{(0.1 - 0.2)^2}\right] = 244 \text{ per group}$$

where

$p_0 = 0.10$

$p_1 = 0.20$

$p = (p_0 + p_1)/2 = (0.10+0.20)/2 = 0.15$

$\alpha = 0.05$

$\beta = 0.10$, power $= 1-\beta = 0.90$

Z_α = 1.65 (See Table 8.3 under $\alpha = 0.05$ in the one-sided comparison column)

Z_β = 1.28 (See Table 8.5 under $\beta = 0.10$ or $1-\beta = 0.90$)

$f = 0.10$

The sample size formula for case-control studies is similar to that used for clinical trials and cohort studies. To apply the sample size formula estimates, two quantities are needed: the proportion of individuals among the cases who are exposed (p_1), and the proportion of individuals among the controls who are exposed (p_0). Researchers should have some idea of the proportion of individuals in the control group who are exposed, since controls supposedly represent the general population. The two proportions are related; given an

estimate of p_0 and another value (the odds ratio [OR]), an estimate of p_1 can be obtained.

Proportion of cases exposed

Formula 5.7.1 estimates the proportion of the cases who are exposed.

(5.7.1)

$$p_1 = p_0 * OR / [1 + p_0 * (OR - 1)]$$

where

p_1 = The estimate of the proportion of individuals among the cases who were exposed.

p_0 = The proportion of individuals among the controls whom we expect have been exposed.

OR = The odds of the outcome of interest among individuals who were exposed divided by the odds of the outcome of interest among individuals who were not exposed. The odds ratio that is to be tested as being statistically significant is specified by the researcher.

In Example 5.8, we consider the relationship between the intrauterine device (IUD) use and ectopic pregnancy. In this example, p_1 is estimated to be 0.095. After p_1 is estimated as a function of p_0 and OR, the following additional information is required to calculate the sample size:

α = the level of significance or type I error probability.

β, $1-\beta$ = the type II error probability and power, respectively.

Sample size formula for a case-control study

Formula 5.7.2 may be used to calculate the required number of participants for the case group and for the control group:

(5.7.2)

$$n = \left[\frac{2 * (Z_\alpha + Z_\beta)^2 * p * (1 - p)}{(p_0 - p_1)^2} \right]$$

where

$p = (p_0 + p_1) / 2$

Z_α = value from Table 5.3

Z_β = value from Table 5.5

Example 5.8

Estimating the Proportion of Exposed Cases

The controls were nonpregnant women 18 to 44 years old who were admitted to surgical or medical services in selected hospitals. The investigator estimated that about 5% of the controls were using intrauterine devices (IUDs) at the time of their last menstrual periods. The cases were women 18 to 44 years old who were admitted to the same hospitals with a diagnosis of ectopic pregnancy. The investigator wants to be able to detect a statistically significant level of ectopic pregnancy among IUD users. This level should be twice the level for non-IUD users, if such a relative difference exists. Using formula 5.7.1, the investigator estimated the proportion of exposed cases:

$$p_1 = (0.05 * 2) / [1 + 0.05 * (2 - 1)]$$
$$= 0.095$$

where

p_0 = 0.05, estimate of proportion of controls that used IUDs

OR = 2 (i.e., investigator wants to detect twice as high an ectopic pregnancy rate in IUD users as in non-IUD users)

In Example 5.9, the investigator assumes that about 5% of the controls had been using IUDs at the time of their last menstrual period. The investigator would like to detect, with statistical power of 0.95, a doubling in the level of ectopic pregnancy among IUD users, if it exists. The level of significance desired is 0.05. Before applying sample size formula 5.7.2, the investigator must obtain an estimate of p_1 by using formula 5.7.1. The investigators must enroll 724 cases and 724 controls.

Example 5.9

Computing the Sample Size for a Case-Control Study

The required sample size for each group is calculated as:

$$n = \left[\frac{2 * (1.65 + 1.65)^2 * 0.0725 * 0.9275}{(0.05 - 0.095)^2}\right] = 724 \text{ per group}$$

where

$p_0 = 0.05$

$p_1 = 0.095$

$p = (p_0 + p_1)/2 = (0.05 + 0.095)/2 = 0.0725$

$\alpha = 0.05$

$\beta = 0.05$, power $= 1 - \beta = 0.95$

$Z_\alpha = 1.65$ (See Table 5.3 under $\alpha = 0.05$ in the one-sided comparison column)

$Z_\beta = 1.65$ (See Table 5.5 under $\beta = 0.05$ or $1 - \beta = 0.95$)

Case-Control Studies with Unequal Group Sizes

Balanced design

One final consideration that applies particularly to case-control studies is the issue of a balanced design. Frequently, a study cannot be designed with equal numbers of cases and controls. If the outcome of interest is a rare phenomenon, we may not be able to enroll a large number of cases. Usually, more individuals are eligible to be controls than cases. To optimize the total number of participants in a study, we need a one-to-one ratio of the number of controls to the number of cases, sometimes referred to as a balanced design.

The formulas given in this chapter assume equal sample sizes in each group. A slight modification to the formulas allows us to compute the sample size or power if the researcher specifies the

Adjustment factor for ratio of r controls to cases

ratio of *r* controls to cases. To compute the sample size for a given ratio of *r* controls to cases, we must modify the existing sample size formula (5.7.2) by a factor of:

(5.9.1)

$$(r + 1) / (2 * r)$$

In Example 5.10, the investigator assumes that the occurrence of ectopic pregnancies is relatively infrequent and decides to enroll three controls for every woman enrolled with a diagnosis of ectopic pregnancy. The required number of cases is 413 and the required number of controls is 3 * 413 = 1,239 (total required study sample is 1,239 + 413 = 1,652). Recall that the total study size for the balanced design (with an equal number of cases and controls) was 1,448 (724 cases + 724 controls) (See Examples 5.8 and 5.9).

The value of p in Example 5.10 is no longer a simple average of p_0 and p_1 but a weighted average of p_0 and p_1. The resulting sample size, n, is for the case group. The size of the control group is r * n. Note that if r = 1, everything simplifies to a situation in which there are equal sample sizes for cases and controls.

Example 5.10

Computing the Sample Size of a Case-Control Study With Unequal Group Sizes

The number of study participants for a case-control study with unequal group sizes is calculated by multiplying formula 5.7.2 by formula 5.9.1:

$$n = \frac{(3 + 1)}{2 * 3} \left[\frac{2 * (1.65 + 1.65)^2 * 0.06125 * .93875}{(0.05 - 0.095)^2} \right] = 413 \text{ cases}$$

where

$r = 3$

$p_0 = 0.05$

$p_1 = 0.095$

$p = (r * p_0 + p_1)/(r + 1) = (0.15 + 0.095)/4 = 0.06125$, a *weighted* average of p_0 and p_1

Example 5.10 (continued)

α = 0.05

β = 0.05, power = $1-\beta$ = 0.95

Z_α = 1.65 (See Table 5.3 under α = 0.05 in the one-sided comparison column)

Z_β = 1.65 (See Table 5.5 under β = 0.05 or $1-\beta$ = 0.95)

Assessing Statistical Power

For several reasons, assessing the statistical power that is provided by a particular sample size may be of interest. If the circumstances of the study limit the sample size, then knowledge of the statistical power available for the study is needed. Also, researchers can calculate statistical power when assessing the findings reported by others. The concept of statistical power does not apply to descriptive studies because they are not comparative (i.e., hypothesis testing is not involved).

Interpreting Power of Published Studies

Power defined

Reporting power when reporting study results in the literature constitutes good epidemiologic practice. Many published studies do not mention power. It is usually possible to compute the power of a published study from the information provided by the authors. The power is the probability that the study data will indicate a difference in treatment or exposures when a difference truly exists. Power of a study is especially important for interpreting results that indicate there is no significant difference between the treatments or exposure levels under consideration. If the power is high, there is no reason to question the study conclusions. However, if the study has insufficient power, a result of no significance (i.e., accepting the null hypothesis) may be attributed to the lack of power rather than to the possibility that the null hypothesis is correct. Published studies with inadequate power probably should not have been conducted.

A study with inadequate power does not allow the researcher to test the hypothesis of interest. Researchers designing a study should consider the statistical power provided by a specified sample size. If the calculations reveal an inadequate level of power, the sample size can be increased until adequate power is achieved. If limitations on the sample size exist because of other considerations, such as cost, time, and limited resources, it may not be feasible to conduct the study.

Randomized Clinical Trials

Power formula for a randomized clinical trial

Given the sample size for each group n, p_1, p_0, and α, formula 5.10.1 provides the power associated with a randomized clinical trial:

(5.10.1)

$$Z_\beta = \frac{\sqrt{n * (1 - f)} * |p_0 - p_1| - Z_\alpha * \sqrt{2 * p * (1 - p)}}{\sqrt{2 * p * (1 - p)}}$$

where

n = sample size per group.

p_0 = estimated or observed proportion of individuals receiving the control treatment who had the outcome of interest.

p_1 = estimated or observed proportion of individuals receiving the new treatment who had the outcome of interest.

$p = (p_0 + p_1) / 2.$

$|p_0 - p_1|$ = the absolute difference between p_0 and p_1.

f = proportion of subjects expected to drop out.

Z_α = determined from Table 5.3 based on the value of the level of significance α.

Z_β = calculated by formula 5.10.1. The statistical power of the study, $1-\beta$, may be found in Table 5.5 corresponding to this value of Z_β. Z_β value >0.0 indicates statistical power >0.50; a Z_β value <0.0 indicates a power <0.50.

In Example 5.11, Z_β is calculated as 0.14. By referring to the $1-\beta$ column of Table 5.5, we find that the statistical power $(1-\beta)$ of the study is between 0.5 and 0.6. Statistical power between 0.5 and 0.6 is considered quite poor because there is only about one chance in two that the study could distinguish a 10% difference in continuation between the two oral contraceptive dosages.

Example 5.11

Calculating the Power for a Randomized Clinical Trial

In a clinical trial that compares the rates of continued use among low-dose oral contraceptive (OC) users and standard-dose OC users in Colombo, Sri Lanka, the investigators believed they would be able to recruit a total of only 400 women for the study (200 women for each group). They wanted the level of significance to be $\alpha = 0.05$ and knew from previous studies that about 70% of women who had taken a standard-dose OC continued to use this pill for at least one year. They wanted to determine if women who took a low-dose pill would have a one-year continuation as large as 10 percentage points different from the standard-dose users (i.e., $\leq 60\%$ or $\geq 80\%$). Using formula 5.10.1, the statistical power is:

$$Z_\beta = \frac{\sqrt{200} * 0.1 - 1.96 * \sqrt{2 * 0.65 * 0.35}}{\sqrt{2 * 0.65 * 0.35}} = 0.14$$

where

$n = 200$

$p_0 = 0.7$

$p_1 = 0.6$ (because this is a two-sided comparison, choose the value of p_0 closer to 0.5.)

$p = (p_0 + p_1) / 2 = 0.65$

Example 5.11 (continued)

$| p_0 - p_1 | = 0.1$

$f = 0$

$\alpha = 0.05$

$Z_\alpha = 1.96$ (see Table 8.3 under the two-sided comparison column)

The power $(1-\beta)$ is found in Table 5.5. For $Z_\beta = 0.14$, the corresponding power is between .5 and .6.

(Adapted from Basnayake et al., 1983)

Cohort Studies

Power formula for a cohort study

The power calculation for cohort studies is similar to the calculation used for randomized clinical trials. The calculation uses the notation introduced in the section on determining the sample size for cohort studies:

(5.11.1)

$$Z_\beta = \frac{\sqrt{n * (1 - f)} * |p_0 - p_1| - Z_\alpha * \sqrt{2 * p * (1 - p)}}{\sqrt{2 * p * (1 - p)}}$$

where

n = sample size for each group.

p_0 = estimated or observed proportion of the unexposed group who developed the outcome of interest.

p_1 = estimated or observed proportion of the exposed group who developed the outcome of interest

$p = (p_0 + p_1) / 2.$

$|p_0 - p_1|$ = the absolute difference between p_0 and p_1.

f = proportion of subjects expected to drop out.

Z_α = determined from Table 5.3 based on the value of the level of significance, α, and on whether it is a one-sided or two-sided comparison.

Z_β = calculated by the formula 5.11.1; used to establish the statistical power of the study, $1-\beta$ (Table 5.5). A Z_β value > 0.0 indicates statistical power > 0.50; a Z_β value < 0.0 indicates a power < 0.50.

In Example 5.12, Z_β is calculated as 2.75. From the Z_β column in Table 5.5, the statistical power $1-\beta$ exceeds 0.99. Thus, the investigator could reduce the sample size and still have good statistical power.

Example 5.12

Calculating the Power for a Cohort Study

In a study that compared the complication rate among women who had had an interval tubal sterilization by silastic band and the complication rate among women who had had an interval tubal sterilization by electrocoagulation, an investigator anticipated enrolling 4,000 participants. She estimated that 1% of the women who had been sterilized by silastic band would exhibit a complication. She would like to be able to detect as statistically different (for $\alpha = 0.05$) a complication percentage as large as 3% among the women who had been sterilized by electrocoagulation. She expects a 5% drop-out rate. Using formula 5.10.1, she calculates the statistical power of this study as:

$$Z_\beta = \frac{\sqrt{2{,}000 * (1 - 0.05)} * 0.02 - 1.65 * \sqrt{2 * 0.02 * 0.98}}{\sqrt{2 * 0.02 * 0.98}} = 2.75$$

where

n = 2,000

p_0 = 0.01

p_1 = 0.03; one-sided comparison

Example 5.12 (continued)

$$\bar{p} = (p_0 + p_1) / 2 = .02$$

$$|p_0 - p_1| = 0.02$$

$$f = 0.05$$

α = 0.05, so $Z_\alpha = 1.65$ (See Table 5.3, one-sided comparison column)

The power $(1-\beta)$ is found in Table 5.5. For $Z_\beta = 2.75$, the corresponding power exceeds .99.

(Adapted from DeStefano et al., 1983)

Case-Control Studies

Estimating the proportion of cases exposed to the risk factor

The power calculation for the case-control study is similar to that for the cohort study, but, as with sample size determination, it requires an estimate of p_1 (the value of the proportion of the cases that were exposed to the risk factor of interest). Recall that p_1 is estimated by the relation:

$$p_1 = p_0 * OR / [1 + p_0 * (OR - 1)]$$

where

p_0 = The proportion of participants in the control group who are expected to be exposed or who are observed to be exposed.

OR = The odds of the outcome of interest among the exposed divided by the odds of the outcome of interest among the unexposed that we wish to detect as being significantly different from 1.

Once an estimate of p_1 is calculated, the power of the study with equal or unequal case and control size is calculated by:

Power formula for case-control study

(5.12.1)

$$Z_\beta = \frac{\sqrt{r * n} * |p_0 - p_1| - Z_\alpha * \sqrt{(r+1) * p * (1-p)}}{\sqrt{(r+1) * p * (1-p)}}$$

where

n = sample size of cases.

r = ratio of controls to cases; $r = 1$ for equal number of cases and controls.

$p = (r * p_0 + p_1) / (r + 1)$.

$|p_0 - p_1|$ = the absolute difference of p_0 and p_1.

Z_α = determined from Table 5.3, it is based on the level of significance (α) and on whether the study tests a one-sided or two-sided comparison.

Z_β = calculated using formula 5.12.1, it is used to determine the power of the study from values shown in Table 5.5.

In Example 5.13, Z_β is calculated as -1.52; the statistical power is less than 50% (Table 5.5). Although the results of the statistical test indicate there is no significant difference in the proportion of women with and without breast cancer who had used depo-medroxyprogesterone acetate (DMPA), the power calculation indicates a less than 50% chance of detecting a difference of the specified magnitude if such a difference truly exists. A result that indicates no significant difference exists between the study groups could be due to the lack of power rather than the possibility that no difference actually exists.

Example 5.13

Calculating the Power of a Case-Control Study

Is the use of depo-medroxyprogesterone acetate (DMPA) associated with the risk of breast cancer? The cases were women attending a family planning clinic who developed breast cancer during a 13-year period (Greenspan et al., 1980). The controls were women attending the same family planning clinic who did not develop breast cancer. The results of the study were as follows:

	Exposure	
Outcome	DMPA use	Did not use DMPA
Women who developed breast cancer	5	25
Women without breast cancer	32	147

OR = 0.9 (95% CI: 0.3 - 2.6)

Women who used DMPA were at decreased risk of developing breast cancer than were women who did not use DMPA; the difference between the groups was not statistically significant at the 0.05 level.

What is the power for these results?

$n = 30, r = 179/30 = 5.97$ (from the table above)

$\alpha = 0.05, Z_\alpha = 1.65$ (one-sided comparison)

$p_1 = 0.17$ (from the table above)

$p_0 = 0.18$ (from the table above)

$$p = (p_1 + r * p_0) / (r + 1)$$
$$= (0.17 + 5.97 * 0.18) / (5.97 + 1) = 0.179$$

$$|p_0 - p_1| = 0.01$$

$$Z_\beta = \frac{\sqrt{30 * 5.97} * 0.01 - 1.65 * \sqrt{6.97 * 0.179 * 0.821}}{\sqrt{6.97 * 0.179 * 0.821}} = -1.52$$

The power $(1 - \beta)$ is found in Table 5.5. For $Z_\beta = -1.52$, the corresponding power is less than .50.

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F Hypothesis testing is the foundation for determining sample size and assessing statistical power.

   (b) T/F If the test fails to reach significance, it may be that the study has high power rather than that the null hypothesis is true.

   (c) T/F In a general sense, the greater the sample size, the more accurate the results.

   (d) T/F Sample size for descriptive studies is based on hypothesis testing.

   (e) T/F A sample size that is too large will limit the conclusions that can be drawn from the study.

   (f) T/F If the power is sufficient, failure to achieve significance may indicate a reasonable difference in the outcomes.

   (g) T/F The proper interpretation of a test that fails to reach significance is based largely on the level of power of the study.

2. Matching. Match the definitions with the terms on the right.

   2.1 ___ The risk that the investigator is willing to make of committing a type I error; also called the significance level.

   2.2 ___ The probability of rejecting the null hypothesis when it is actually false; equal to $1 - \beta$.

   2.3 ___ The probability of committing a type II error.

   a. Type I error
   b. Type II error
   c. α
   d. β
   e. power of a test
   f. null hypothesis
   g. alternative hypothesis

2.4 ___ Affirms that there is no difference between the populations being compared with respect to the parameter under study.

2.5 ___ When, as a result of a test of significance, a null hypothesis is not rejected when it is false.

2.6 ___ When, as a result of a test of significance, a null hypothesis is rejected when it should not have been rejected because it is true.

3. Consider the problem of assessing the level of maternal mortality associated with cesarean delivery in a large network of hospitals. An estimate of the level can be obtained by reviewing a sample of medical records. Previous studies in similar hospital settings have estimated that the level of maternal mortality for all deliveries (cesarean and vaginal) was about 10 per 1,000. It is desired that the estimate be within two deaths per 1,000 of the actual level with 90% confidence.

   (a) What size sample would be required?

   d =

   $1-\alpha$ =

   α =

   Z_α =

   p =

   n =

4. In Example 5.2, we compared the rate of continued use among women who used a standard-dose OC and the rate of continued use among women who used a low-dose OC. From this Example, we know the following information:

   p_0 = 0.6 of women who use the standard-dose OC will continue to use this OC for 12 months.

   p_1 = 0.5 or 0.7 (two-sided comparison) of women who use the low-dose OC will continue use for 12 months.

   α = 0.05.

   β = 0.10, power = $1 - \beta = 0.90$.

   f = 0; all dropouts are considered discontinuations.

   With the two-sided comparison, we have two possible values for p_1. The value closer to 0.5 should be applied because it will always yield a larger sample size. In Example 5.2, $p_1 = 0.5$ yielded a sample size of 520 individuals per group.

   (a) Calculate the sample size using $p_1 = 0.7$ with all the other values the same as in the first calculation.

5. A case-control study is being designed to examine the association between IUD use and pelvic inflammatory disease (PID). It is estimated that 20% of the controls are current IUD users. Further, it is desirable to have a sample size large enough to detect a ratio as large as 2 of the proportions of IUD users who develop PID relative to the proportion of non-IUD users who develop PID. A level of significance of 0.05 and power of 0.80 are required.

   (a) Calculate the sample size for a balanced design.

   p_0 =

   OR =

   p_1 =

   p =

   α = Z_α =

   β = $1-\beta$ = Z_β =

   n = per group

   (b) Calculate the sample size for an unbalanced design where r = 2.

6. In a study that compared the complication level of two interval tubal sterilization procedures, electrocoagulation and silastic bands, the investigators suspected that electrocoagulation was not as safe as silastic bands. They preferred to set the statistical significance at the 0.05 level for detecting a doubling of risk. The acceptable power of the comparison is 0.95. They assumed that the current complication level among the women sterilized by silastic bands is 1 per 100; they do not expect any patients to drop out.

   (a) What type of study is this?

   (b) Calculate the sample size.

   p_0 =

   p_1 =

   p =

   α = Z_α =

   β = Z_β =

   f =

   n =

7. After reading the following paragraph, answer 7a - g by circling true (T) or false (F).

   The annual pregnancy rate for a commonly used spermicide A is 16%. Researchers want to compare this spermicide with a new spermicide B to see if the pregnancy rate is reduced by half (i.e., to 8%). They are willing to tolerate an α error of 0.05 and a β error of 0.10. It is assumed that 15% of the subjects will drop out of the study before its completion. Our goal at this point is to figure out the sample size.

   (a) T/F This study involves a two-sided comparison.

   (b) T/F $1-\beta = 90$

   (c) T/F $Z_\alpha = 1.65$ (from Table 5.3)

   (d) T/F $Z_\beta = 1.28$ (from Table 5.5)

   (e) T/F This is a case-control study.

   (f) T/F The formula is:

$$n = \frac{1}{(1-f)} * \left[\frac{2 * (Z_\alpha + Z_\beta)^2 * p * (1-p)}{(p_0 - p_1)^2} \right]$$

   where $p = (p_0 + p_1) / 2$

   (g) T/F $p = (0.16 + 0.08) / 2 = 0.12$

   7.1 Calculate the sample size for this example.

8. In comparing a new IUD and a standard IUD, the investigators are equally interested in the expulsion rate and the pregnancy rate accumulated over six months. Assume that $\alpha = 0.05$ and that the expected drop out rate is 20%. Further, assume that the standard IUD has a six-month expulsion rate of 10 per 100 women and a six-month pregnancy rate of four per 100 women. The investigators would like to be able to detect as different the expulsion and pregnancy rates in the new IUD as small as 5 and 2, respectively. They would like to detect a halving of rates with statistical power $(1-\beta)$ of 0.80.

   (a) Calculate the power for the expulsion rate and the pregnancy rate, assuming the study enrolled only 400 participants in each group.

$$\alpha =$$

$$Z_{\alpha} =$$

$$p_0 =$$

$$p_1 =$$

$$|p_0 - p_1| =$$

$$p =$$

$$f =$$

$$n =$$

   (b) What would these powers mean? What sample size would be needed to achieve the 0.80 power desired by the researchers?

9. Investigators planning a cohort study would like 80% power ($\beta = 0.20$) to show a difference between an incidence in the exposed of 8.0% and an incidence in the unexposed of 5.0%, when testing at a level of significance of 0.05. The risk ratio that is considered important is 1.6. They therefore calculate that 1,058 participants are necessary in each group; the total sample size should be 2,116. Compensating for a 20% loss to follow-up they calculate that they need to enroll a total of 2,645 participants. However, they anticipate that they can enroll only 600 participants in the exposed group. They would like the study to be balanced, so they will enroll a total of 1,200 volunteers in the study.

   (a) Based on the sample size limitations, calculate the power.

   (b) How would you interpret the power calculation? Is the study worth pursuing?

10. Circle true (T) or false (F).

    (a) T/F Statistical significance can be defined as $\alpha = .05$ or $\alpha = .01$ or any other small value the researcher chooses.

    (b) T/F The purpose of hypothesis testing is to distinguish *no effect* from *some effect*.

    (c) T/F If the null hypothesis is false, the likelihood that the study decision rejects the null hypothesis depends primarily on statistical power.

    (d) T/F Statistical power does not apply to descriptive studies because they are not comparative.

    (e) T/F Checking the power can help prevent the researcher from conducting a study too small to provide conclusive results.

    (f) T/F The power of a study design may confirm a nonsignificant test result, affirming the truth of the null hypothesis.

    (g) T/F Power calculations may clarify that a completed study with results indicating no difference did not have sufficient power to detect a difference.

    (h) T/F In a two-sided comparison, we are interested in detecting a difference in performance only if the new treatment is better than the control.

11. All of the following are constraints on the sample size except:

    (a) time and costs
    (b) dropout rate
    (c) interviewer skills
    (d) number of subjects available
    (e) ability to draw conclusions

12. Fill in the blanks. Complete the following decision table with one of the labels given at the right.

Study Decision	Actual State of Nature: Null Hypothesis is True	Actual State of Nature: Null Hypothesis is False
Accept null hypothesis	Correct decision	12.1____
Reject null hypothesis	12.2____	12.3____

a. Correct decision
b. Type I error, α
c. Type II error, β
d. Confidence level

13. If you were conducting a study and would like to detect an incidence ratio of 3.0 with a significance level of 0.05, you would compute a sample size. Assume that you expect to find an incidence of about 7% in the unexposed groups and that you want a power of 80% (β = 0.20).

(a) What type of study is this?

(b) What is the formula for n?

(c) Is this a one- or two-sided test?

(d) Calculate the sample size, using Tables 5.3 and 5.5.

(e) If you could enroll 150 participants in the exposed group, how would that affect the power of the test?

(f) If you have not already done so in Question 13.5 above, calculate the power of the test with 150 participants in each group.

(g) What would you do in this situation?

14. A household survey is being considered in country Y to estimate the extent of sexual activity among female adolescents 13 to 19 years old. Studies in other countries with similar populations have found that 33% of females aged 13 to 19 years have reported being sexually active. It is desired that the estimate be within 1% of the actual level with 95% confidence. The response rate is estimated to be 85%.

    (a) What size sample would be required?

    d =

    $1-\alpha$ =

    α =

    Z_α =

    p =

    q =

    n =

    (b) Suppose a recent census indicated that Country Y has approximately 80,000 female adolescents aged 13 to 19 years in the population. Is it necessary to apply the finite population correction (fpc) to estimate the sample size? If so, what is the corrected sample size?

Suggested Answers to Practice Exercises

1. True or false.

   (a) T Introduction

   (b) F Sample Size and Power; no, because it has low power.

   (c) T Sample Size and Power

   (d) F Descriptive Studies; no, it is based on confidence intervals.

   (e) F Sample Size and Power; no, but too small a sample size will.

   (f) F Sample Size and Power; no, it would indicate a negligible difference.

   (g) T Sample Size and Power

2. Matching.

   2.1 c Tests of Hypotheses

   2.2 e Tests of Hypotheses

   2.3 d Tests of Hypotheses

   2.4 f Tests of Hypotheses

   2.5 b Tests of Hypotheses

   2.6 a Tests of Hypotheses

3. Assess the level of maternal mortality.

   (a) Using the estimate of maternal mortality from all deliveries in previous studies as a guess at the level of maternal mortality for cesarean deliveries in the proposed study.

$$d = 2/1{,}000 = 0.002$$

$$1-\alpha = 0.90$$

$$\alpha = 0.10$$

$$Z_\alpha = 1.65 \text{ (two-sided comparison)}$$

$$p = 10/1{,}000 = 0.01$$

$$n = p * (1-p) * (Z_\alpha/d)^2 = 0.01 * 0.99 * (1.65/0.002)^2 = 6{,}738.2 \simeq 6{,}738$$

4. Calculate a sample size.

   (a) In two-sided comparisons, two values are available for p_1. We should choose the value closer to 0.5 so we do not underestimate the appropriate sample size. The two values for p_1 were 0.5 and 0.7. We used 0.5 in our calculation to get a sample size of 520 per group. If we had used 0.7, the sample size would be:

$$n = \frac{1}{(1-0)} * \left[\frac{2 * (1.96 + 1.28)^2 * 0.65 * 0.35}{(0.6 - 0.7)^2}\right] = 478 \text{ per group}$$

   If p_1 later actually turned out to be 0.5, then our sample size (calculated with $p_1 = 0.7$) would not be large enough to meet our statistical requirements.

5. Calculate sample size for a balanced and unbalanced design.

(a) $p_0 = 0.2$

OR $= 2$

then,

$p_1 = 0.2 * 2 / [1+0.2 * (2-1)]$

$= 0.33$

$p = (0.2 + 0.33)/2 = 0.265$

$\alpha = 0.05 \quad Z_\alpha = 1.65$ (one-sided comparison)

$\beta = 0.20 \quad 1-\beta = 0.80 \quad Z_\beta = 0.84$

$$n = \left[\frac{2 * (1.65 + 0.84)^2 * 0.265 * 0.735}{(0.2 - 0.33)^2}\right] = 143 \text{ per group}$$

(b) $r = 2$

$p = (2 * 0.2 + 0.33) / 3 = 0.243$

so,

$$n = \frac{3}{4} * \left[\frac{2 * (1.65 + 0.84)^2 * 0.243 * 0.757}{(0.2 - 0.33)^2}\right] = 102 \text{ cases}$$

The number of controls would be $2 * 102 = 204$.

6. Determine the type of study and calculate sample size.

(a) Cohort study

(b) $p_0 = 0.01$; we treat the silastic band participants as the unexposed group.

p_1 = 0.02; we want to detect a doubling of the risk of complication among the women who were sterilized by the electrocoagulation procedure.

p = (0.01 + 0.02) / 2 = 0.015

α = 0.05 Z_α = 1.65 (one-sided comparison)

β = 0.05 power = $1 - \beta = 0.95$ Z_β = 1.65

f = 0

$$n = \frac{1}{(1 - 0)} * \left[\frac{2 * (1.65 + 1.65)^2 * 0.015 * 0.085}{(0.01 - 0.02)^2}\right] = 3{,}218 \text{ per group}$$

7. True or false.

(a) F No, it is one-sided; we are interested only in detecting a reduced pregnancy rate.

(b) F No, $1 - \beta = 1 - 0.10 = 0.90$ (or 90 percent).

(c) T Table 5.3, under one-sided comparison for $\alpha = 0.05$.

(d) T Table 5.5, for $\beta = 0.10$.

(e) F This is a randomized clinical trial.

(f) T See sample size formula for randomized clinical trials.

(g) T See sample size formula for randomized clinical trials.

7.1 Sample size for the example:

$$n = \frac{1}{(1 - 0.15)} * \left[\frac{2 * (1.65 + 1.28)^2 * 0.12 * 0.88}{(0.16 - 0.08)^2}\right] = 333 \text{ per group.}$$

8. Compare new and standard IUDs.

(a) Pregnancy rate comparison:

$$\alpha = 0.05$$

$$Z_\alpha = 1.65, \text{ a one-sided comparison}$$

$$p_0 = 0.04$$

$$p_1 = 0.02$$

$$|p_0 - p_1| = 0.02$$

$$p = (0.04 + 0.02)/2 = 0.03$$

$$f = 0.2$$

$$n = 400$$

$$Z_\beta = \frac{\sqrt{400 * (1 - 0.2)} * 0.02 - 1.65 * \sqrt{2 * 0.03 * 0.97}}{\sqrt{2 * 0.03 * 0.97}} = -0.17$$

$Z_\beta = -0.17$; from Table 5.5 the power $1 - \beta$ is < 0.50.

For expulsion rate comparison all values are the same except:

$$p_0 = 0.1$$

$$p_1 = 0.05$$

$$|p_0 - p_1| = 0.05$$

$$p = (0.1 + 0.05)/2 = 0.075$$

$$Z_\beta = \frac{\sqrt{400 * (1 - 0.2)} * 0.05 - 1.65 * \sqrt{2 * 0.075 * 0.925}}{\sqrt{2 * 0.075 * 0.925}} = 0.75$$

Z_β = 0.75; from Table 5.5, the power $1 - \beta$ is between 0.70 and 0.80.

(b) The power for the expulsion rate is almost sufficient but the power for the pregnancy rate comparison is not. A power $(1 - \beta)$ less than 0.50 is very poor because there is less than one chance in two that the study could distinguish a difference in pregnancy rates of the order that the researcher specified. To achieve the 80% power (Z_β = 0.84) that the researcher desires for both comparisons, use values for the pregnancy rate comparison because there is more power for an expulsion rate comparison. The sample size will have to be increased substantially from 400 per group to about 1,130 per group.

$$n = \frac{1}{(1 - 0.2)} * \left[\frac{2 * (1.65 + 0.84)^2 * 0.03 * 0.97}{(0.04 - 0.02)^2}\right] = 1{,}128 \text{ per group}$$

9. Calculating power.

(a) $\alpha = 0.05$, $Z_\alpha = 1.65$, (one-sided comparison)

$p_0 = 0.05$

$p_1 = 0.08$

$p = (p_0 + p_1) / 2 = 0.065$

$|p_0 - p_1| = 0.03$

$n = 600$

$f = 0.2$

then,

$$Z_\beta = \frac{\sqrt{600 * (1 - 0.2)} * 0.03 - 1.65 * \sqrt{2 * 0.065 * 0.935}}{\sqrt{2 * 0.065 * 0.935}} = 0.24$$

From Table 5.5, the power is 60%.

(b) A power $(1 - \beta)$ of 0.60 is not very good because there is about a three-in-five chance that the study could distinguish a difference in incidence rates between the exposed and unexposed groups of the order that the researcher specified. Having only 600 subjects available for each group puts serious limitations on the power of the study and thus limits the conclusions that may be drawn from the study. It is probably not worth pursuing the study.

10. True or false.

(a) T Sample Size and Power

(b) T Tests of Hypotheses

(c) F Decision making and Sample Size and Power; no, it depends on sample size.

(d) T Assessing Statistical Power

(e) T Sample Size and Power

(f) F Decision making and Sample Size and Power; you only know you can't reject the null hypothesis.

(g) T Sample Size and Power

(h) F Determining the Sample Size; this describes a one-sided comparison.

11. c Introduction

12. Fill in the blanks.

12.1 c
12.2 b
12.3 a

13. Computing sample size.

(a) Cohort study

(b) $$n = \frac{1}{(1 - f)} * \left[\frac{2 * (Z_\alpha + Z_\beta)^2 * p * (1 - p)}{(p_0 - p_1)^2}\right]$$

(c) One-sided test

(d) $$n = \left[\frac{2 * (1.65 + 0.84)^2 * 0.14 * 0.86}{(0.07 - 0.21)^2}\right] = 76 \text{ per group.}$$

(e) Because 150 persons per group is more than 76, the power of the test would be improved to a level higher than the 80% specified by the researcher.

(f) $$Z_\beta = \frac{\sqrt{150} * 0.14 - 1.65 * \sqrt{2 * 0.14 * 0.86}}{\sqrt{2 * 0.14 * 0.86}}$$

Z_β = 1.84 (from Table 5.5, the power is between 0.95 and 0.975).

(g) More participants are available. Therefore, if the cost or effort to do the study with more participants is not appreciably increased, the improved power might make using 150 participants per group more appealing. If an 80% power level is acceptable to the researcher, then there is no need to use more than 103 participants per group.

14. Estimate the extent of sexual activity among female adolescents.

(a) Using the level of sexual activity in a neighboring country as a guess for the level of sexual activity in country Y, the sample size is calculated as follows:

$d = 0.01$

$1-\alpha = 0.95$

$\alpha = .05$

$Z_\alpha = 1.96$ (two-sided comparison)

$p = 0.33$

$f = 0.15$

$$n = \frac{1}{(1-f)} * p * (1-p) * (Z_\alpha/d)^2 = \left[\frac{1}{1-0.15}\right] * (0.33) * (0.67) * \left[\frac{1.96}{0.01}\right]^2 = 9992.68 \approx 10{,}000$$

(b) Yes, because a sample of 10,000 comprises 12.5% of a population of 80,000, the finite population correction is needed.

$$fpc = \frac{n}{1 + n/N} = \frac{10{,}000}{1 + \frac{10{,}000}{80{,}000}} = 8{,}888.8 \approx 8{,}889$$

References

Basnayake S, de Silva SV, Miller P, Rogers R. A comparison of Norinyl and Brevicon in three sites in Sri Lanka. Contraception 1983;27:453-464.

Cochran WG. Sampling techniques. New York: John Wiley & Sons, 1977.

DeStefano F, Greenspan JR, Dicker RC, Peterson HB, Strauss LT, Rubin GL. Complications of interval laporoscopic sterilization. Obstet Gynecol 1983;61:153-158.

Fleiss J. Statistical methods for rates and proportions. New York: John Wiley & Sons, 1981.

Greenspan AR, Hatcher RA, Moore M, Rosenberg MJ, Ory HW. The association of depo-medroxyprogesterone acetate and breast cancer. Contraception 1980;21:563-569.

Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.

Schlesselman JJ. Case-control studies. New York: Oxford University Press, 1982.

Chapter 6 - Learning Objectives

After completing this chapter, you should be able to:

1. Recognize appropriate uses of descriptive studies.

2. Describe three types of comparisons often made by descriptive studies:

   population subgroup comparisons
   geographic comparisons
   temporal comparisons

3. Specify descriptive characteristics of person data.

4. Describe different types of geographic comparisons.

5. List various reasons for a geographic difference in exposure or disease occurrence.

6. Describe three types of temporal changes identified in descriptive studies:

   short-term changes
   cyclical changes
   secular (long-term) changes

7. Describe sources of data and methods of collecting data for descriptive studies.

8. Recognize advantages and disadvantages of descriptive studies.

9. Recognize how to organize descriptive study data for presentation and interpretation.

6 Descriptive Studies

Introduction

Descriptive epidemiology defined

Descriptive epidemiologic studies describe patterns of disease occurrence or exposure to risk factors in relation to person, place, and time. Person refers to *who* is affected, place refers to *where* the health problem is more or less common, and time refers to *when* the health problem is occurring. This information is valuable to public health personnel: administrators, clinicians, epidemiologists, and researchers. Knowing which population subgroups are most likely to be affected allows targeting of particular population segments for education, preventive programs, and resources. Identifying descriptive characteristics can be the first step in identifying risk factors that can be altered or eliminated to reduce or prevent disease and other health problems. Descriptive studies can also provide valuable clues for generating specific hypotheses that can be tested subsequently in analytic studies (Hennekens and Buring, 1987).

Descriptive studies serve many purposes. For example, a descriptive study of the demographic characteristics and socioeconomic factors of women who died of pregnancy-related causes can determine which population subgroups are at highest risk. A study describing women who develop pelvic inflammatory disease (PID) among women who use intrauterine devices (IUD) can identify potential risk factors and provide information for hypothesis testing in future analytic studies. Descriptive studies are also used to monitor long-term time trends in the occurrence of health outcomes or exposures (e.g., the prevalence of breast-feeding over the last ten years).

Although descriptive studies do not, by definition, include a formal comparison group, valid comparisons are possible. For example, a descriptive study of an entire population or geographic area will include people with and without the risk factor or health outcome of interest. Therefore, it is possible to compare the characteristics of those with and without a particular exposure or

outcome. For example, in a descriptive study of maternal mortality in hospitals in Chile, rates and causes of pregnancy-related deaths in rural hospitals could be compared with urban hospitals. A descriptive study of births in a single hospital could compare the characteristics of women who have had stillbirths with women who have had live births. Again, information from these comparisons can provide important leads for subsequent analytic studies. Example 6.1 illustrates the type of comparisons that can be made by using information from the World Fertility Study (WFS), a multinational descriptive study.

Example 6.1

Birth Spacing and Infant Mortality

Research Question: Does birth spacing have an important effect on child survival?

Study Design: The World Fertility Survey (WFS) was an international research project that conducted surveys in 41 countries in Asia, Africa, the Middle East, and Latin America to provide high-quality data on fertility and family planning.

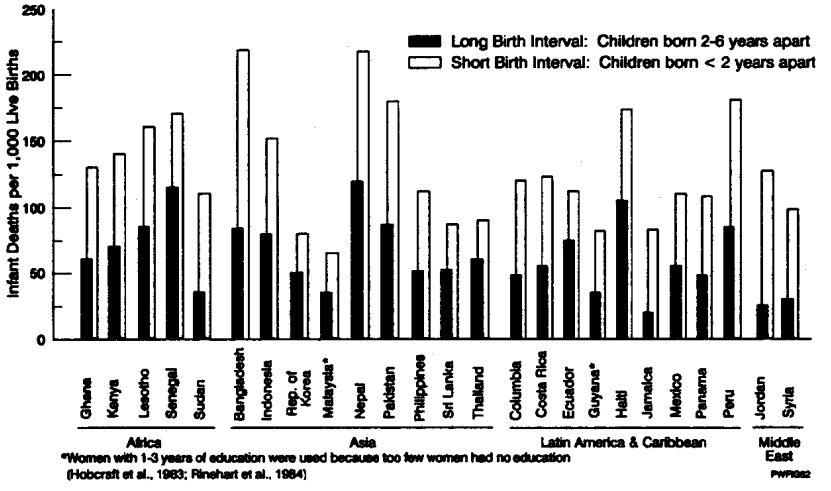
Data Collection Methods: The WFS questionnaire included a core of basic questions that were asked in every survey so that comparable data would be available for all countries. Optional modules were used by many countries to tailor the survey to the need and social setting of the particular country. The WFS core questionnaire contained two sections: the household schedule and the individual questionnaire. The household schedule asked the age, sex, and marital status of all members of the household. The individual questionnaire collected data on the background of the respondent (age, level of education, literacy, work history), marital and maternity history, and knowledge and use of contraception. As a result, comparisons of knowledge, use, and availability of family planning services among countries were possible.

Results: Several multicountry findings revealed highly relevant information about the determinants of fertility and mortality. For example, Figure 6.2 demonstrates the clear and consistent relationship between birth spacing and infant mortality.

(Rinehart et al., 1984)

Figure 6.2

Death Rates for Infants Born After Short or Long Birth Intervals to Uneducated Mothers* in 25 Countries



In the 25 countries surveyed for the WFS (Figure 6.2), analysis indicated that infants born soon after another birth were more likely to die than infants born after a longer interval. To minimize socioeconomic differences that might affect birth spacing and mortality, these data cover only the children of mothers with no formal education. This study highlights the relevance of family planning to maternal and child health programs, whether or not population control is a motive. Closely spaced births is a high-risk situation that affects infant mortality and causes other maternal health problems. Spacing of less than two years can mean lower birthweights, poorer nutrition, shorter periods of breast-feeding, and more competition for family resources and care. Other comparisons of fertility by age, education, and urban/rural residence were possible from the WFS data and provided important information for individual countries in planning their contraceptive service programs.

Types of Descriptive Studies

Descriptive studies address three basic types of comparisons: population, geographic, and temporal. Public health personnel need to formulate the research questions to be answered by the descriptive study in terms of person, place, and time measures.

Population comparisons are the who or person measure. Descriptive data related to *person* address questions such as *Who is developing the disease or health problem?* or *Who is at higher risk of exposure?* People affected or exposed are characterized by basic demographic and social characteristics, such as age, sex, marital status, race or ethnic origin, religion, and measures of socioeconomic status.

Geographic comparisons are the *where* or place measure. Descriptive characteristics related to *place* consider *Where are the rates of disease highest and lowest?*, (i.e., *Are cases equally distributed with respect to country, state or district within country, urban-rural residence, or area within an affected local community?*)

Temporal trends are the *when* or time measure. Descriptive data on time address such questions as *When does the disease occur commonly or rarely?* and *Is the frequency of the disease at present different from the corresponding frequency in the past?* For example, is there any unusual feature of the distribution of cases by year, season, month, or day of occurrence?

Note that descriptive studies can cover a wide range of time spans, from a study conducted at a single point in time to a surveillance study that monitors long-term trends in the occurrence of health outcomes or exposures.

Population Subgroup Comparisons

Population comparisons

Comparisons of disease rates by who is most and least affected involve collecting descriptive data on the persons affected. Who is affected is an important determinant of outcome, and a careful description of the types of people affected can be helpful in delineating causes of disease as well as in targeting population subgroups for health programs and resources.

Important descriptive characteristics

The characteristics of person must include age and sex. Marital status, race, ethnicity, and measures of socioeconomic factors such as occupation and literacy are other important factors that relate to disease occurrence. In less developed areas, variables that indicate

the availability of running water, electricity, and sanitation may be used to measure socioeconomic status. Important characteristics that affect reproductive outcomes are parity, family size, birth order, child spacing, prenatal care, contraceptive usage, and maternal age.

Age

Age should always be considered in an epidemiologic study. With most health outcomes, the variation in occurrence by age is greater than that found with any other variable. Information about the relationship between age and outcome occurrence is important for two reasons: first, study of variation in the frequency of an outcome with age may assist in understanding the factors responsible for its development; second, associations between age and the frequency of a health outcome are usually so strong that age may produce indirect effects that must be considered in interpreting differences in outcome patterns among different populations (MacMahon, 1970).

The association of disease frequency with age is usually measured by relating the number of cases of the disease in each age group to the population in the same age group and deriving a succession of age-specific incidence or prevalence rates (MacMahon and Pugh, 1990). These rates are useful for delineating specific age groups that are most affected or at highest risk. Rates for reproductive outcomes for developed countries are generally restricted to ages 15 to 44 years. The age may extend to 49 years for developing countries.

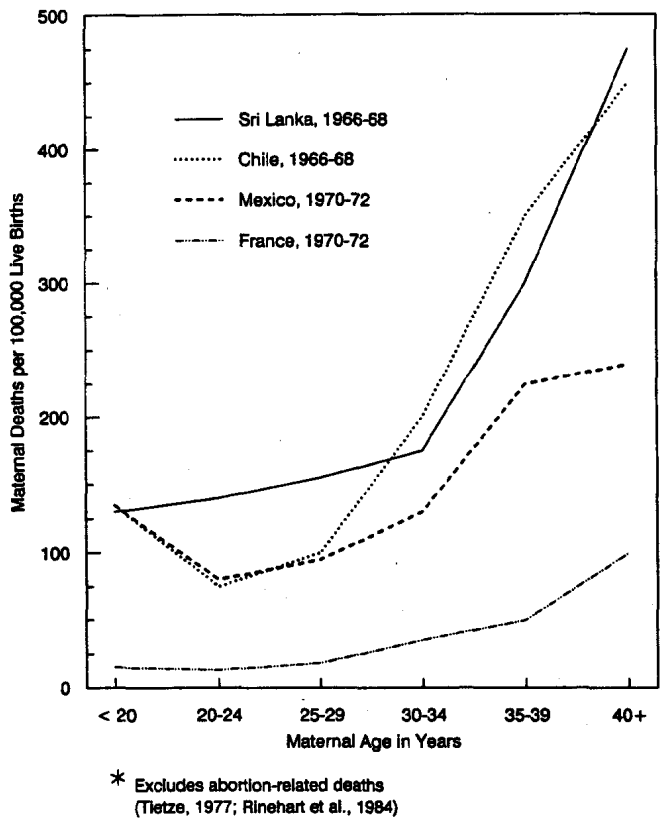
Maternal age

Figure 6.3 shows the strong effect of maternal age on the risk of childbearing in Chile, Sri Lanka, Mexico, and France. When plotted against age, maternal mortality rates form a J-shaped curve. This figure shows that among these four countries, childbearing is safest for women in their 20s. Maternal mortality rates are highest for older women and young adolescents. This same pattern has been observed in countries with greatly differing maternal mortality rates and socioeconomic conditions (Rinehart et al., 1984; National Research Council, 1989).

Sex differences

Since some outcomes occur more frequently in males and others occur more frequently in females, sex should be considered in an epidemiologic study. For example, the greater occurrence of coronary heart disease in young men than in young women is partly explained by sex differences in known risk factors—such as blood lipids, blood pressure, cigarette smoking, diabetes, and

Figure 6.3
Maternal Death Rates* by Maternal Age



obesity—and partly due to possible protection of the female by estrogens before menopause (Freidman, 1980). Differences by sex can have an important impact on reproductive health outcomes. Descriptive studies of reproductive health issues can provide valuable information on the differences between men's and women's attitudes toward contraception, childbearing, and accessibility to health care services. Example 6.4 describes a survey that provided valuable insights about husbands' attitudes and behaviors concerning issues related to fertility and family planning in Jordan. Any population policy related to family planning in Jordan should take these attitudes into consideration (e.g., policymakers may want to concentrate

Example 6.4

Jordan Husbands' Fertility Survey

Research Question: What are husband's attitudes and behaviors toward family planning and fertility-related issues? How do these attitudes and behaviors differ from their wives?

Study Design: The 1985 Jordan Husbands' Fertility Survey interviewed the husbands of women who were currently married when interviewed in the 1983 Jordan Fertility and Family Health Survey.

Data Collection Methods: The questionnaire consisted of a short household screening form and a detailed individual form. The individual husband's questionnaire asked information on marriage background, respondent's background, contraceptive knowledge and attitudes, the wife's recent fertility and birth planning, and the husband's and wife's use of contraception. Three hundred and thirty interviews were completed from a sample of husbands of women interviewed in the 1983 Jordan Fertility Survey.

Results: Nearly 40% of husbands did not believe in using contraception and over 50% viewed family size as "up to God." Husbands gave very different reasons for not using fertility control (religion) than their wives (want to get pregnant). Many wives did not appear to be in control of their fertility behavior. Nearly one-third of women who did not want an additional child in 1983 had a birth between the two surveys.

(Warren et al., 1987)

their efforts on younger males, among whom attitudes were less rigid).

Ethnic or racial groups

Ethnic or racial groups often differ in disease experience. In some cases, the differences are genetically determined, such as the black-white differences in sickle cell anemia. With other outcomes, the explanation may not be so simple, especially when racial or ethnic differences are accompanied by differences in socioeconomic or environmental factors. A difference in disease occurrence between cultural groups may reflect a difference in risk factors, such as common exposure to infectious agents, diet, prenatal care, or personal hygiene. Also, access to or acceptability of medical care can explain the patterns observed. Nevertheless, descriptive studies of disease rates among ethnic, racial, or tribal

subgroups of a single population may be useful for applying specific preventive measures and may be particularly helpful in providing leads for further epidemiologic study (Fox et al., 1970; Hennekens and Buring, 1987).

Overall, in Example 6.5 and Table 6.6, black infants had twice the risk of dying in their first year of life as white infants. This gap was more pronounced for infants weighing 2,500 grams or more. Part of the black-white gap is related to the relatively disadvantaged status of blacks in the United States. However, the widespread differences between black and white infant mortality across all maternal characteristics suggest that there may be other problems, including lack of access to effective health care for black infants and pregnant black women. Public health strategies should aim to decrease the incidence of low birthweight to increase neonatal survival for black infants weighing 3,000 grams or more and to increase postneonatal survival for all black infants.

Example 6.5

Black-White Differences in Infant Mortality

Research Question: Is there a black-white difference in infant mortality in the United States, and if so, does this difference vary by birthweight?

Study Design: The National Infant Mortality Surveillance (NIMS) project for the 1980 U.S. birth cohort provided neonatal, postneonatal, and infant mortality risks for blacks, whites, and all races in 12 categories of birthweights.

Data Collection Methods: Fifty states, New York City, and the District of Columbia participated in this national surveillance project. All reporting areas linked birth and death certificates for infants who were born alive in 1980 and who died within the first year of life in 1980 or 1981. Information on all resident births for each state serves as the denominator for infant mortality risks. States also provided the number of infant deaths by birthweight, age at death, race, and other infant and maternal characteristics.

Results: Neonatal mortality decreased sharply with increasing birthweight up to 4,000 grams for both blacks and whites; for all races, neonatal deaths per 1,000 live births decreased from 647.6 for infants weighing 500 to 999 grams to 1.4 for infants weighing 3,500 to 3,999 grams (Table 6.6). Black infants weighing less than 3,000 grams had a lower birthweight-specific neonatal

Example 6.5 (continued)

mortality than whites, whereas heavier black infants had a much higher birthweight-specific neonatal mortality. Postneonatal mortality decreased with increasing birthweight up to 4,000 grams for all races, although the slope was not as steep as for neonatal mortality; postneonatal deaths per 1,000 neonatal survivors decreased from 135.2 for infants weighing 500 to 999 grams to 1.9 for infants weighing 4,000 to 4,499 grams. Blacks had higher postneonatal mortality than whites within all birthweight categories. Infant deaths included neonatal deaths and postnatal deaths. The risk of infant mortality for blacks (18.9) was 2.0 times the risk for whites (9.3).

(Hogue et al., 1987)

Marital status

Marital status appears on medical and civil records almost as regularly as age and sex. Marital status (e.g., never married, married, widowed) is particularly useful in epidemiologic studies of reproductive outcome. For example, studies of breast and cervical cancers have shown substantial differences by marital status. Breast cancer is more apt to develop in single women or women who marry late in life, whereas cervical cancer is associated with early marriage. Analytic studies based on these observations have suggested that cervical cancer is associated with sexual activity at an early age and that having a first pregnancy at an early age is protective against breast cancer (Freidman, 1980).

Socio-economic status

Socioeconomic status is not in itself a measurable characteristic, but a theoretical concept. It is usually measured indirectly by occupation, educational level, income, and other factors. For example, the standard socioeconomic variables collected in the WFS were the educational attainment of the respondent and her partner, their present and childhood places of residence, the partner's occupation and work status, and the employment of the respondent both before and after her first marriage (her occupation, work status, and place of work) (Singh and Casterline, 1985). Because social class is often a reliable predictor of health experience, classification of persons into groups according to socioeconomic characteristics is particularly useful for planning health care resource allocations. For example, a marked socioeconomic gradient in infant mortality has long been noted. In fact, infant mortality rates have often been used as an index

Table 6.6
Infant Mortality Risk, by Birthweight,* Age at Death, and Race of Single-Delivery Infants Born During 1980, United States

Race	Less than 500g	500 to 999g	1,000 to 1,499g	1,500 to 1,999g	2,000 to 2,499g	2,500 to 2,999g	3,000 to 3,499g	3,500 to 3,999g	4,000 to 4,499g	4,500g or more	Total
					Neonatal Deaths† per 1,000 Live Births						
Blacks	1,000.0	615.6	131.3	36.1	10.6	3.6	2.4	2.5	2.8	8.7	12.5
Whites	1,000.0	660.8	212.1	61.6	18.3	4.2	1.8	1.3	1.4	3.0	6.2
All races‡	1,000.0	647.6	186.5	53.9	16.0	4.0	1.9	1.4	1.5	3.5	7.3
					Postneonatal Deaths§ per 1,000 Neonatal Survivors						
Blacks		157.1	49.8	24.2	11.6	6.5	4.4	3.2	3.3	4.1	6.5
Whites		115.0	43.7	18.9	9.4	4.4	2.5	1.8	1.7	2.0	3.1
All races‡		135.2	45.8	20.7	10.2	4.9	2.9	2.0	1.9	2.2	3.7
					Infant Deaths‖ per 1,000 Live Births						
Blacks	1,000.0	676.0	174.6	59.4	22.1	10.0	6.8	5.7	6.1	12.8	18.9
Whites	1,000.0	699.8	246.5	79.3	27.5	8.5	4.3	3.1	3.1	5.1	9.3
All races‡	1,000.0	695.2	223.7	73.5	26.0	8.9	4.8	3.5	3.4	5.7	11.0

*Number of infants with unknown birthweight were redistributed according to percentage of infants with known birthweight.
†Includes the number of deaths < 28 days per 1,000 live births.
‡Includes unknown race and infants of other races.
§Includes the number of deaths from 28 days to under one year per 1,000 neonatal survivors.
‖Includes the number of deaths < 1 year per 1,000 live births.
Source: National Infant Mortality Surveillance Report.

(Hogue et al., 1987)

both of living standards and of availability of medical services for comparisons between countries and comparisons of areas within countries (Freidman, 1980).

Socioeconomic indicators should be culture-specific

Other information can be used to indicate socioeconomic status, such as the possession of various household durables (e.g., refrigerator, television) and the type of residence (e.g., number of rooms, toilet, source of drinking water, electricity). Specific socioeconomic indicators must be considered within the individual setting, since each society has its own type of economy, cultural traditions, and social structure (Singh and Casterline, 1985).

Occupation and work status

Socioeconomic status is often measured by occupation and work status. This information is easy to collect and often appears on routinely collected records. Occupation is often categorized by type, such as professional or technical, clerical, skilled or unskilled, agricultural, and domestic. In developing countries, work status is sometimes categorized as working or not working outside the home. It may be important to separate collected data according to urban or rural status before stratifying by occupation because occupation distributions often differ substantially in urban and rural areas.

Income, level of education, and literacy

Income may seem a more direct measure of socioeconomic status than occupation. However, income data are often not readily available or are irrelevant in cultures that barter for required goods. Educational level and literacy status are also indicators of socioeconomic standing and, like occupation, often differ significantly between urban and rural areas. Literacy status can be assessed by asking those with minimal schooling whether they can read a letter or newspaper (Zimbabwe National Family Planning Council, 1985). Information on the occupation and educational level of both partners among married couples should be collected, since in many societies the occupation or education of the husband has a greater effect on the socioeconomic status of the family than that of the wife.

Geographic Comparisons

Descriptive characteristics related to place can serve two important purposes: they can provide major insights into disease etiology (Hennekens and Buring, 1987), and they can provide essential information for health policy and program decision

making (e.g., planning the location and size of medical-care facilities). Geographic comparisons of health outcomes can be made at many levels—between countries, between regions in a country, between regions in a city, or between states, provinces, or counties. For example, maternal mortality rates vary substantially among different countries (Table 6.7).

Interpret international differences cautiously

International differences should always be interpreted cautiously. Countries differ substantially in the reliability of diagnosis, completeness of reporting, and validity of population census data. Any or all of these issues may affect reported estimates, mask real differences in the health problem under study, or even show a difference where none exists.

Table 6.7

Maternal Mortality and Maternal Deaths as a Proportion of all Deaths of Women of Reproductive Age, Various Countries

Country	Percentage of Deaths from Maternal Causes	Maternal Mortality Rate per 100,000 Live Births	Year
Bangladesh (Rural Jamalpur)	46	623	1982-1983
India (Rural Andhra Pradesh)	45	874	1984-1985
Bangladesh (Rural Tangail)	33	566	1982-1983
India (Urban Andhra Pradesh)	28	545	1984-1985
Paraguay	27	275	1984
Bangladesh (Matlab)	26	510	1983
Indonesia (Bali)	23	718	1980-1982
Egypt (Menoufia)	23	190	1981-1983
Egypt (South)	21	300	1984-1985
Ecuador	15	190	1980
Romania	10	149	1984
Mexico	10	88	1984
El Salvador	8	70	1984
Mauritius-Maurice	6	103	1985
Costa Rica	5	26	1983
Cuba	3	45	1983
Japan	1	16	1985
United States	1	8	1983
Hong Kong	1	5	1985
Sweden	0	2	1984

(Royston and Lopez, 1987)

Urban and rural differences

Example 6.8 provides information about infant mortality and maternal risk factors from a survey conducted in northeastern Brazil during 1980. As can be seen in the last column of Table 6.9, children who had never been breast-fed were more likely to die in infancy than breast-fed children. However, the association between breast-feeding and infant mortality was much stronger in rural areas than in urban areas. When the data were stratified according to urban and rural status, certain important relationships were discovered (Table 6.10). For example, in urban areas, the association between infant mortality and breast-feeding was weaker.

In rural areas however, not only was infant mortality higher, but the association between infant mortality and failure to breast-feed was much stronger. In rural areas, the use of both maternal and child health services further adjusted the effect. This information suggested that for the greatest public health impact in

Example 6.8

Infant Mortality and Breast-Feeding in Northeastern Brazil

Research Question: What is the prevalence of breast-feeding in northeastern Brazil, and how does breast-feeding affect infant mortality in this area?

Study Design: The data for this analysis came from the Northeast Brazil Family Planning/Maternal Child Health Survey of 1980. Northeastern Brazil is the poorest region in Brazil.

Data Collection Methods: This survey covered four of the nine states in the northeastern region of Brazil. These four states contain over 20 million inhabitants, roughly one-sixth of Brazil's population. This survey consisted of interviews with 7,852 women of all marital statuses between the ages of 15 and 44. Each respondent was asked about the breast-feeding experience of her most recently born child; such information was obtained for 5,190 children. For children no longer alive at the interview date, women were asked the age (in months) at which the child had died. Other social, economic, and demographic information was also collected by interview.

Results: Children who had never been breast-fed were almost twice as likely to die in infancy as breast-fed children (Table 6.9).

(Goldberg et al., 1984)

Table 6.9

Percentage of Children Dying During First Year of Life, by Breast-Feeding Status and Selected Variables, Family Planning/Maternal and Child Health Survey, Northeastern Brazil, 1980

	Total	Breast-Fed	Not Breast-Fed	Not Breast-Fed ÷ Breast-Fed
Total	4.9 (3,457)	4.4 (2,674)	6.9 (783)	1.6
Residence				
State capitals	1.0 (814)	1.6 (621)	3.1 (193)	1.9
Other urban	5.2 (1,386)	5.1 (1,051)	5.5 (335)	1.1
Rural	6.1 (1,257)	4.9 (1,002)	11.4 (225)	2.3
Maternal-child health services				
No service	6.6 (793)	5.6 (652)	12.2 (141)	2.2
1 service	6.4 (1,127)	5.7 (886)	9.3 (241)	1.6
2 services	2.5 (1,524)	2.4 (1,127)	2.8 (397)	1.2

Note: Figures in parentheses are unweighted numbers of cases.

(Goldberg et al., 1984)

northeastern Brazil, educational efforts to increase breast-feeding should be concentrated in rural areas.

When comparing rates in smaller geographic areas, such as between villages or townships, it becomes more important to have an accurate measure of the population at risk of the health problem. When the population at risk (the denominator) is small, morbidity and mortality rates become less stable. For example, in a town with 1,000 births per year, there is a substantial difference in the maternal mortality rate if there is one maternal death per year or three maternal deaths per year (Hogue et al., 1985). In an area with a large population at risk, three maternal deaths would not significantly alter rates.

Geographic differences

A geographic association can be explained in many ways. Different geographic areas can vary by climate, geology, socioeconomic conditions, nutrition, and culture. Geographic areas can also differ in the quality and availability of medical services, the attitude of the

Table 6.10

Percentage of Children Dying During First Year of Life, by Breast-Feeding Status, Residence, and Selected Variables, Family Planning/Maternal and Child Health Survey, Northeastern Brazil, 1980

	Urban			Rural		
	Breast-Fed	Not Breast-Fed	Not Breast-Fed ÷ Breast-Fed	Breast-Fed	Not Breast-Fed	Not Breast-Fed ÷ Breast-Fed
Total	4.0 (1,665)	4.8 (526)	1.2	4.9 (1,009)	11.3 (257)	2.3
Mother's education						
No school	5.5 (397)	6.0 (113)	1.1	5.6 (483)	17.4 (107)	3.1
< Completed primary	5.8 (493)	5.8 (132)	1.0	5.2 (483)	7.9 (89)	1.5
≥ Completed primary	1.9 (281)	3.7 (281)	1.9	1.5 (158)	2.9 (61)	1.9
Mother's work status						
Currently employed	4.4 (467)	7.0 (140)	1.6	8.3 (303)	23.6 (76)	2.8
Not currently employed	3.8 (1,195)	4.0 (383)	1.1	3.5 (702)	6.5 (180)	1.9
Maternal Health Services						
No service	7.8 (235)	9.6 (57)	1.2	4.4 (417)	13.7 (84)	3.1
1 service	5.8 (500)	7.1 (140)	1.2	5.6 (386)	13.1 (101)	2.3
2 services	1.8 (928)	2.7 (328)	1.5	5.4 (199)	3.5 (69)	0.6

Note: Figures in parentheses are unweighted numbers of cases.

(Goldberg et al., 1984)

different segments of the population toward these services, and the quality of the medical and vital statistics (Lilienfeld et al., 1967). If there is a demonstrated association between the health problem and place, the following reasons should be considered: 1) the inhabitants of the particular place possess characteristics of etiologic importance to the development of the outcome that differ from such characteristics of inhabitants of other places (e.g., age distribution, ethnic differences, cultural practices, diet); 2) there are etiologic factors present in the biologic, chemical, and physical environment of the people inhabiting the affected places (e.g., climate, geology, nature and quantity of animal and vegetable

species); 3) there are differences in the availability of medical care; and 4) there are differences in the reporting of medical and vital statistics (MacMahon and Pugh, 1970).

Temporal Comparisons

Changes over time

Descriptive studies of disease occurrence by time are useful for suggesting possible etiologic hypotheses and for planning future trends in medical needs. These changes over time, or temporal trends, have been useful to reproductive epidemiology in the study of such diverse outcomes as infant mortality, ectopic pregnancy, and teenage pregnancy. Three major kinds of temporal change can be identified: short-term fluctuations (e.g., weeks in an epidemic), cyclic changes (e.g., seasonal changes), and secular trends (e.g., changes over many years).

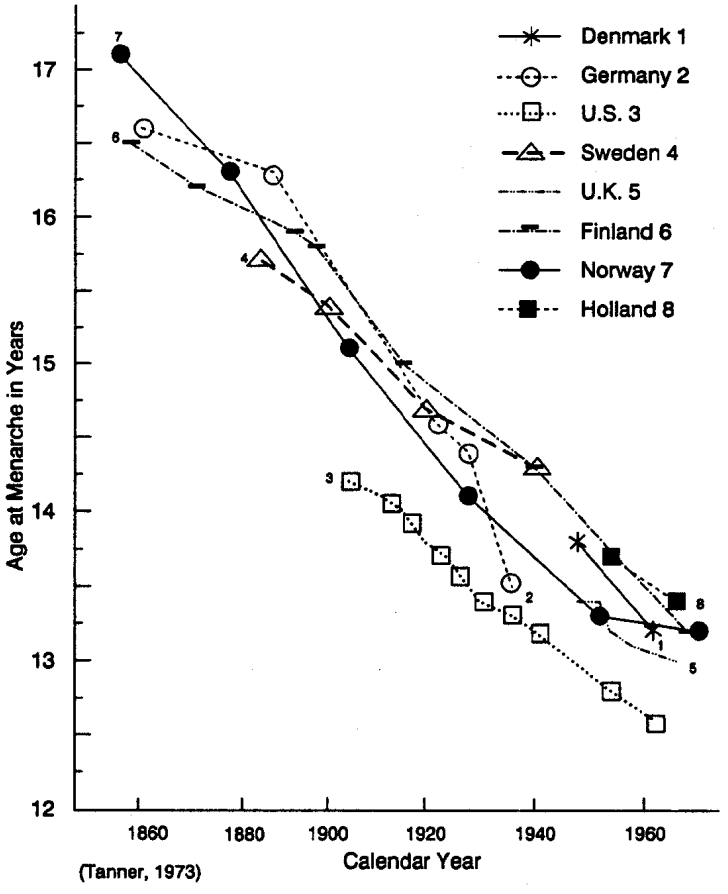
Short-term changes

Short-term changes or fluctuations in disease rates are measured in hours, days, weeks, or months. These are most often observed in the study of infectious disease epidemics or foodborne outbreaks. Examples include brief bacterial contaminations of water supplies, acute air pollution episodes resulting in epidemics of asthmatic attacks and increases in the daily number of deaths, or foodborne outbreaks due to bacterial or chemical contamination (Fox et al., 1970).

Cyclical changes

The incidence of some diseases shows cyclical or regular recurring increases and decreases. The pattern may exhibit cycles that last several years or occur annually and represent seasonal variation in disease occurrence (Freidman, 1980). Although seasonal fluctuations are most often associated with infectious diseases, reproductive outcomes also show cyclic changes. For example, there is a persistent seasonal pattern in the United States in the average daily number of live births, and presumably also in conception rates (National Center for Health Statistics, 1966). For diseases of early life, variation in risk by season of birth may suggest environmental factors operating during intrauterine or very early postnatal life (MacMahon and Pugh, 1970). In reproductive epidemiology, examining secular or long-term trends, or changes in frequency that take place over years or decades, can be particularly useful. For example, an interesting secular trend is the change in age at menarche between 1840 and 1970 (Figure 6.11). In Figure 6.11, changes in nutrition probably account for the striking decline in menarcheal age over time (Tanner, 1973).

Figure 6.11
Secular Trend in Age at Menarche
Northwest Europe, 1840-1970



Secular trends

A major source of information on secular trends has been death certificate data. Variations in death rates over time are most informative if they can be shown to reflect real changes in incidence rates. However, several other factors can be responsible for a change in death rates over time. These include:

- Improved medical techniques resulting in improved diagnoses and recognition of disease or health problems.

- Medical care becoming available to larger segments of populations and to outlying areas and increasing recognition of disease or health problems.
- Changes in rules and procedures for recording causes of deaths.
- Changes in the collection of vital statistics data (e.g., greater coverage in remote areas and greater accuracy in recording).
- Changes in the availability and accuracy of population size estimates.
- Changes in populations at risk (e.g., a population is composed of persons at different ages, and the age distribution is likely to change as time passes).

Birth and death registration

However, only one-third of the world's population live in countries where the registration of births and deaths is reasonably complete and the certification of causes of death is reasonably reliable (Royston and Lopez, 1987). In many areas, periodic surveys may be the only practical method to measure secular changes. For example, the potential of surveys for studying reproductive health has been demonstrated on an international scale with the WFS and the Contraceptive Prevalence Surveys. Both promote a series of surveys in a large number of countries and use comparable survey methodology and content. These programs have provided substantial information on comparative contraceptive-use patterns over time (Hermalin and Entwisle, 1982). Example 6.12 illustrates the use of repeated surveys to measure long-term trends.

Total fertility rates (TFR)

Table 6.13 presents estimates of total fertility rates (TFR) and age-specific fertility rates (ASFR) in Jordan. These data indicate that total fertility in Jordan fell from 7.7 to 6.6 lifetime births per women between 1976 and 1983. The most pronounced decrease in fertility occurred among younger women; for example, the fertility rate for 20- to 24-year-olds decreased by 34 percent, from 344 births per 1,000 women during the period 1971 to 1975 to 228 per 1,000 by 1983. Among women who were at least 40 years old, there were no declines in fertility.

Fertility rates in Jordan between 1976 and 1983 declined somewhat (Table 6.13). However, the approximately seven lifetime births per

Example 6.12

Fertility Trends in Jordan

Research Question: What are the fertility trends and determinants in Jordan?

Study Design: Household surveys

Data Collection Methods: Fertility estimates were based on several surveys: the 1976 Jordan National Fertility Survey, the 1979 Jordan Fertility Survey, the 1981 Jordan Demographic Survey, the 1982 Jordan Manpower Survey, and the 1983 Jordan Fertility and Family Health Survey. The 1983 survey sampled 6,068 housing units from the East Bank area of Jordan. The questionnaire consisted of a household form and an individual questionnaire administered to all ever-married women living in the household. The individual questionnaire included questions on the health and immunization status of all children in the household, on infant feeding practices, on the use of maternal and infant health services and family planning.

Results:

Table 6.13

Total Fertility Rate and Age-Specific Fertility Rates per 1,000 Women

Rate	1976	1979	1981	1982	1983
TFR	7.7	7.5	7.1	6.6	6.6
ASFR (years)					
15-19	133	66	87	40	49
20-24	344	280	252	219	228
25-29	358	361	340	332	335
30-34	336	331	316	321	305
35-39	245	262	239	245	233
40-44	104	141	134	117	127
45-49	11	50	49	45	40
Period covered	1971-1975	1978-1979	1979-1981	1981-1982	1980-1983

(Anderson et al., 1985)

woman still represents a very high level of fertility. Further analyses suggested that Jordan's fertility decline was mainly due to later marriage, not to increased contraceptive use. Because there is probably a limit to the effect that changes in age at marriage can have on fertility decline, improved contraceptive practice is the factor most likely to bring about greater fertility reduction. Besides reducing the high rate of population growth, higher levels of contraceptive use will also benefit women's and infants' health by increasing the duration of intervals between births. Additionally, Jordanian women, on average, gave birth approximately two years (26.7 months) after a previous birth (Table 6.14). The average interval increased with age, and women living in urban areas had longer intervals between births than women living in rural areas.

Table 6.14
Mean Interval Between Live Births in the Five Years Preceding Interview and Length of Interval in Which 25%, 50%, and 75% of Jordanian Women Had Their Next Birth, by Age at Time of Interview and Residence, 1983

Age and Residence	Average Interval* (months)	25%	50%	75%	No. of Intervals
Total	26.7	18.1	24.9	38.9	5,755
Age (years)					
15-19	21.4	15.3	22.4	25.3	144
20-24	21.6	15.5	21.7	27.5	1,003
25-29	24.0	16.2	23.7	32.3	1,467
30-34	25.9	18.5	24.4	36.3	1,154
35-39	31.3	21.8	29.1	45.5	1,129
40-49	39.7	23.9	36.2	†	858
Residence					
Urban	28.4	18.1	25.8	43.7	3,839
Rural	25.2	18.2	24.2	34.1	1,916

*Weighted average of the length of the interval based on the 25th, 50th, and 75th percentiles.

†By the time of interview, fewer than 75% of 40- to 49-year-olds who gave birth in the preceding five years had another child. An estimated value for the 75th percentile was used to calculate the trimean for 40- to 49-year-olds.

(Anderson et al., 1985)

Surveillance

Surveillance system

Epidemiologic surveillance is the ongoing systematic collection, analysis, and interpretation of health data essential to planning, implementing and evaluating public health practice (Centers for Disease Control, 1986). A surveillance system is a series of descriptive studies over time that provides useful long-term descriptive information for planning and evaluating health services for a given population. A surveillance system can also provide information to evaluate the feasibility of conducting a particular analytic study. For example, in a case-control study of oral contraception and liver cancer, one needs to know whether there is a sufficient number of histologically diagnosed liver cancer cases in age groups that are likely to have used oral contraceptives. Accurate surveillance systems are expensive and complicated and have been instituted in only a few countries. Efforts to establish surveillance systems in less developed areas will inevitably be accompanied by changes in the coverage and quality of diagnostic services. When surveillance systems are established, it will be important to consider whether observed changes in rates simply reflect changes in the availability of diagnostic services (Lilienfeld et al., 1967).

Sources and Methods of Collecting Data

Sources

Data for descriptive studies can be derived from existing systems, such as the vital statistics records systems, or can be collected from health surveys. As mentioned previously, most countries do not have vital record systems that are complete and accurate (Royston and Lopez, 1987). Consequently, other sources of information may have to be used, including health interview surveys, medical records from hospital or clinic files, medical ledgers or logs, and direct observations or measurements (e.g., physical examinations, anthropometric measurements).

Health interview surveys

Health interview surveys are especially valuable for countries that do not have established systems for recording morbidity and mortality and that have a high prevalence of the health problem under study. Health interview surveys also permit the collection of demographic, risk factor, and cause of death data for describing

the health problem under study in relation to person, place, and time. The types of data available for study and the methods of collection usually depend on the available resources and the extent of the problem.

Descriptive studies of maternal or infant mortality, for example, use various sources of data to estimate the extent of the problem and to describe relevant associations. Example 6.15 illustrates the use of a survey to collect information about maternal mortality in rural Bangladesh. The overall maternal mortality rate was 62.3 per 10,000 live births and accounted for 46.0 percent of all deaths to women of reproductive age (Table 6.16). Maternal mortality rates were lowest among women aged 20 to 24 years and increased substantially with increasing age (Table 6.16).

Example 6.15

Maternal Mortality in Rural Bangladesh

Research Question: What is the rate and pattern of maternal mortality in rural Bangladesh?

Study Design: Household survey

Data Collection Methods: The study was conducted in two rural areas of the Jamalpur district in Bangladesh. The two areas comprised 240 villages with a population of about 267,000. The study employed 109 combined traditional birth attendants on a part-time basis as the primary reporters of births and maternal deaths in their study area; each attendant covered an average of two villages with a total population of about 2,500. The birth attendants were required to visit their areas once every 2 weeks to collect information on maternal deaths, live births, and neonatal deaths. Because the birth attendants were generally illiterate or semiliterate, the collected mortality information was reported verbally to literate supervisors, who recorded their findings. In a second step, the supervisors routinely visited all houses where the birth attendants had reported a maternal death. During the follow-up visit, the supervisors interviewed the family or close relatives of the deceased woman to collect additional information on the woman's health and the circumstances surrounding her death. To ensure complete reporting, all deaths among women of reproductive ages were reviewed to identify the cause of death. Menstrual status immediately prior to death was also recorded to identify possible cases of induced abortion performed early in pregnancy. A single one-page form recorded the death, symptoms, circumstances leading to death, time of death in

Example 6.15 (continued)

relation to pregnancy termination or childbirth, and medical care received by the woman before death. Data were collected for a period of 12 months from September 1982 to August 1983.

Data Analysis: The maternal mortality rate was calculated in this example as the number of deaths of any woman of any cause while pregnant or within 42 days of termination of pregnancy, irrespective of the duration or site of pregnancy, per 10,000 live births (Table 6.16).

(Khan et al., 1986)

Table 6.16

Number of Total Births, Live Births, and Deaths Among Women of Reproductive Age (WRA) and Maternal Deaths, by Mother's Age, 1982-1983

Maternal Age (Years)	Total Births	Live Births	Deaths to WRA	Maternal Deaths	Maternal Deaths as a % of WRA Deaths	Maternal Deaths per 10,000 Live Births
<20	1,827	1,744	27	10	37.0	57.3
20-24	3,118	3,006	17	8	47.1	26.6
25-29	2,435	2,337	22	11	50.0	47.1
30-34	1,255	1,221	14	9	64.3	73.7
35-39	767	729	16	13	81.3	178.3
40+	309	280	30	7	23.3	250.0
Total	9,711	9,317	126	58	46.0	62.3

(Khan et al., 1986)

Hospital records

Data on maternal deaths are often obtained from vital records or from hospital record abstractions (Rochat, 1987). However, not only are vital record systems in many countries incomplete, but certification of the cause of death is rarely reliable (Royston and Lopez, 1987). Hospital data are commonly used as an alternative to death registration data for estimating the maternal mortality rate. However, hospital rates may not be representative of the general rates because many maternal deaths occur outside the

hospital. Data for women who deliver in a hospital may include a higher proportion of high-risk women (e.g., the rate will be inflated if women who intend to deliver at home are transported to the hospital for emergency care). Conversely, if the hospital is fee-paying, it will attract economically advantaged women; the measured maternal mortality rate may thus be lower than that prevailing in the community. However, hospital-based studies can give valuable cause-of-death information that can suggest specific interventions (Royston and Lopez, 1987).

Consider Example 6.17. Of the 36 maternal deaths following cesarean sections, only 27 had case records available for review. The probable causes of death for these 27 cases are presented in Table 6.18. Maternal sepsis was responsible for 81.5% of the maternal deaths after cesarean section. Examination of the data according to time of death revealed that the majority of patients dying from sepsis (40.8%) did so between the second and fourth day after cesarean section.

Example 6.17

Characteristics of Maternal Deaths Following Cesarean Section

Research Question: What are the causes of maternal mortality after cesarean section in the University of Ilorin Teaching Hospital, Nigeria?

Study Design: Hospital record abstractions

Data Collection Methods: The University of Ilorin Teaching Hospital serves as a referral center for Kwara State, in the middle belt of Nigeria. Data were obtained from hospital records covering a five-year period from 1982 to 1986.

Results: During the five-year study period, of 48,974 deliveries, 1,992 were by cesarean section. The total number of maternal deaths was 125, with 36 deaths following cesarean section (Table 6.18). The fatality rate for cesarean section was thus 18.1 per 1,000 deliveries. The total number of vaginal deliveries was 46,982 with 89 maternal deaths—a fatality rate of 1.89 per 1,000 deliveries.

(Ojo et al., 1988)

Table 6.18

Probable Cause of Maternal Death After Cesarean Section, by Time of Death

Cause	Time of Death (Days After Delivery)	Number (%)
Sepsis	0-1	0 (0.0)
	2-4	11(40.8)
	5-7	5(18.5)
	8+	6(22.2)
	Total	22(81.5)
Eclampsia	0-1	0 (0.0)
	2-4	1 (3.7)
	5-7	0 (0.0)
	8+	0 (0.0)
	Total	1 (3.7)
Postpartum hemorrhage	0-1	1 (3.7)
	2-4	0 (0.0)
	5-7	2 (7.4)
	8+	1 (3.7)
	Total	4(14.8)
Total	0-1	1(3.7)
	2-4	12(44.5)
	5-7	7(25.9)
	8+	7(25.9)
	Total	27(100.0)

(Ojo et al., 1988)

The high maternal mortality after cesarean section reported in this study reflects the general pattern seen in developing countries. Reasons cited for the higher mortality rates included delays in admission (i.e., doctors live at a long distance from the hospital, and there are no telephone facilities for immediate summons of doctors; also, patients seek medical care late in labor) and the failure to use prophylactic antibiotics (some patients were not on antibiotics until the third postoperative day). Recommendations to reduce the cesarean section mortality rate included developing a

department policy for the management of high-risk patients (especially in the early stages of labor), appropriately using prophylactic antibiotics, and educating traditional birth attendants to recognize the need for early referral of high-risk pregnancies.

Investigation of deaths

Finally, other approaches to data collection for studying maternal and infant mortality include careful investigation of causes of death among all women of reproductive age and prospective surveillance of pregnant women and pregnancy outcomes. These options may be feasible only where the prevalence of maternal or infant mortality is high.

Advantages and Disadvantages of Descriptive Studies

Descriptive studies may be the most frequently conducted type of epidemiologic study (Hennekens and Buring, 1987). The advantages for conducting descriptive studies include the following:

- They are generally easy to conduct.
- They are less expensive to conduct than other epidemiologic study designs.
- They can reveal patterns of disease occurrence and trends over time.
- They permit the collection of information on important potential risk factors, such as age, race, sex, and geographic location. These data may be used to compare prevalence and to develop hypotheses for future analytic studies.
- They provide a basis for planning, providing, and evaluating health services for a given population.
- They raise few ethical problems.

The disadvantages of descriptive studies include the following:

- They do not test etiologic hypotheses. There is no formal comparison group and therefore no valid method to assess whether the study prevalence is greater or less than would be expected.

- The temporal relationship between the health problem under study and potential exposures cannot be easily determined. This fact plus the lack of a formal comparison group prohibits the assessment of causality. The evaluation of causality requires analytic study designs, such as experimental studies or randomized clinical trials, cohort studies, and case-control studies.

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F Descriptive studies provide information to administrators on patterns of disease occurrence.

   (b) T/F Descriptive studies are designed to test epidemiologic hypotheses.

   (c) T/F Descriptive studies may tell epidemiologists who is affected.

   (d) T/F Information from descriptive studies provides clues for generating hypotheses.

   (e) T/F Descriptive studies provide useful comparison data through use of control groups.

   (f) T/F Descriptive studies help target specific population subgroups most likely to be affected by disease.

   (g) T/F Descriptive studies are easy to conduct but relatively expensive.

   (h) T/F Descriptive studies provide a basis for planning, providing, and evaluating health services to a population.

   (i) T/F Descriptive studies provide a way of determining whether prevalence is greater or less than expected.

   (j) T/F Descriptive studies can be used to evaluate trends over time.

   (k) T/F Causality cannot be determined by descriptive studies because there is no formal comparison group.

2. Multiple Choice. Select one response.

2.1 All the following are common types of comparisons made by descriptive studies except:

(a) Etiologic
(b) Temporal
(c) Geographic
(d) Population subgroup

2.2 All of the following are important person descriptors that affect disease occurrence except:

(a) Age
(b) Race
(c) Sex
(d) Marital status
(e) Geographic location
(f) Religion

2.3 Of the choices given in Question 2.2, which almost always accounts for more variation than any other characteristic?_____

2.4 All of the following are types of geographic comparisons except:

(a) Between countries
(b) Between states or provinces
(c) Among socioeconomic groups
(d) Among regions in a city
(e) Urban versus rural areas

2.5 All of the following are reasons for finding an association of disease with place except:

(a) People characteristics
(b) Time differences
(c) Environmental factors
(d) Availability of care
(e) Reporting of statistics

2.6 Changes in all the following may explain changes over time in death rates except:

(a) Quality of diagnosis
(b) Accuracy of records
(c) Availability of care
(d) Geography
(e) Population
(f) Survival

3. Matching. Match the study with the type of temporal change by filling in a letter next to each number.

a. Short-term change
b. Cyclical change
c. Secular change

3.1 ___ Asthma attacks caused by air pollution
3.2 ___ Bacterial contamination of water
3.3 ___ High risk of retardation among children born in February
3.4 ___ Decrease in menarche age over the past 30 years in Europe
3.5 ___ Foodborne outbreaks due to bacterial contamination
3.6 ___ Seasonal pattern of average daily number of births

4. Consider the data in Table 6.19.

Table 6.19

Percentage of Maternal Deaths, by Category and Specific Cause of Death, 1982-1983

Category and Cause of Death	Percentage (and Number) of Maternal Deaths	
Direct obstetric	86.2	(50)
Eclampsia	20.7	(12)
Antepartum	10.3	(6)
Postpartum	10.3	(6)

Table 6.19 (continued)

Category and Cause of Death	Percentage (and Number) of Maternal Deaths	
Hemorrhage	10.3	(6)
Antepartum	5.2	(3)
Postpartum	5.2	(3)
Tetanus	6.9	(4)
Sepsis	31.0	(18)
Septic abortion	20.7	(12)
Postpartum	10.3	(6)
Difficult labor	17.2	(10)
Obstructed	10.3	(6)
Retained placenta	6.9	(4)
Indirect obstetric	13.8	(8)
Cardiovascular	8.6	(5)
Suicide	1.7	(1)
Not diagnosed	3.5	(2)
All causes	100.0	(58)

Note: Percentage subtotals and total do not agree because of rounding.

(Khan et al., 1986)

(a) What were the most common causes of death?

(b) How many and what percentage of deaths were due to difficult labor?

(c) What specific actions can be taken to reduce the levels of maternal mortality?

5. Consider Table 6.20 from a study of contraceptive procedures performed in the People's Republic of China during 1971 to 1977.

Table 6.20

Birth-Planning Operations in the People's Republic of China, 1971-1977*

Year	IUD Insertions	Sterilizations		Induced Abortions	Total
		Female	Male		
1971	6,173	1,745	1,224	3,910	13,052
1972	9,220	2,087	1,716	4,814	17,837
1973	13,950	2,956	1,933	5,110	23,949
1974	12,580	2,276	1,445	4,985	21,286
1975	16,744	3,260	2,653	5,084	27,741
1976	11,620	2,700	1,490	6,570	22,380
1977	12,974	2,776	2,616	5,229	23,595

*In thousands

(Chen and Kols, 1982; Zhang, 1980)

(a) Describe the comparison groups.

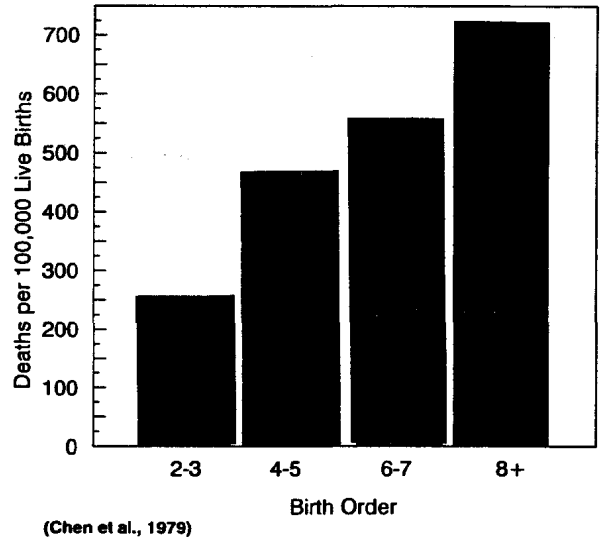
(b) What trends do you see in the data? How would you interpret these trends?

(c) What are some reasons for the trends in contraception that are exhibited in the data?

6. Consider the data summarized in Figure 6.21, from a study of maternal mortality conducted in Bangladesh during 1968 through 1970.

Figure 6.21

Maternal Deaths, by Birth Order, Matlab, Bangladesh, 1968-1970.



(Chen et al., 1979)

PWFIG621

(a) Describe the comparison groups.

(b) Summarize the information provided in Figure 6.21.

(c) What is the most important variable to consider when interpreting these data?

7. Consider the contraceptive use data summarized by survey year and urban-rural residence in Table 6.22.

Table 6.22

Percentage of Currently Married Women Aged 15-49, by Current Contraceptive Use Status and Residence

Contraceptive Use Status	1976			1983		
	Total	Urban	Rural	Total	Urban	Rural
Currently using	22.8	29.5	7.0	26.0	31.7	12.3
Pill	11.9	15.4	3.8	7.8	9.2	4.4
IUD	2.0	2.7	0.3	8.3	0.9	2.1
Sterilization	1.9	2.2	1.2	3.8	4.2	2.7
Rhythm	2.1	2.8	0.5	2.9	3.4	1.6
Withdrawal	3.3	4.3	1.0	2.4	2.8	1.4
Condom	1.4	1.9	0.2	0.6	0.8	0.2
Injectable	NA*	NA*	NA*	0.2	0.3	0.0
Other	0.1	0.2	0.0	0.1	0.1	0.0
Not using	77.2	70.5	93.0	74.0	68.4	87.7
Total	100.0	100.0	100.0	100.0	100.0	100.0

*Not applicable

(Anderson et al., 1985)

(a) Describe the comparison groups.

(b) Summarize the information provided in Table 6.22.

(c) What are some possible explanations for the observed geographical differences in contraceptive use?

Suggested Answers to Practice Exercises

1. True or false.

   (a) T Introduction

   (b) F Introduction. No, analytic studies such as experimental studies, cohort studies, and case-control studies are designed to test epidemiologic hypotheses.

   (c) T Introduction

   (d) T Introduction

   (e) F Introduction. No, descriptive studies do not have a formal comparison group.

   (f) T Introduction

   (g) F Advantages #1 and #2

   (h) T Introduction

   (i) F Introduction and Disadvantage #1. Descriptive studies cannot assess whether the study prevalences is greater or less than would be expected. However, descriptive studies permit comparisons between subgroup populations.

   (j) T Advantage #3

   (k) T Disadvantage #1

2. Multiple choice.

   2.1 a Types of Descriptive Studies

   2.2 e Types of Descriptive Studies

2.3 Age Population Subgroup Comparisons

2.4 c Geographic Comparisons

2.5 b Geographic Comparisons

2.6 d Secular Trends

3. Matching.

3.1 a

3.2 a

3.3 b

3.4 c

3.5 a

3.6 b

4. Consider data in Table 6.19.

(a) The most common causes of death were sepsis and eclampsia.

(b) Ten deaths (17.2%) were due to difficult labor.

(c) More than one-fifth of all maternal deaths in this study were due to septic abortion. In most cases the abortion was induced by indigenous practitioners using unsanitary and unscientific procedures leading to infection and death. Wider availability of safe fertility regulation services can substantially reduce the level of maternal mortality. Deaths from postpartum sepsis can be prevented with the provision of aseptic birth care services and antibiotics. Deaths caused by difficult labor, hemorrhage, and toxemia can be prevented by timely management and referral to the nearest health care facility. Only 4 out of 51 mothers were admitted to an institutional facility before death. Since more than

95% of all deliveries in rural areas take place at home and are mostly managed by traditional birth attendants, appropriate training of the attendants to recognize high-risk cases for timely referral to clinical facilities would be a practical preventive measure. Moreover, providing women with available contraceptive methods and educational programs on the advantages of limiting and spacing births through the practice of contraception could help reduce the number of women at risk of maternal mortality.

5. Consider data from Table 6.20.

   (a) The numbers and types of birth planning procedures performed in China during 1971 to 1977 are compared.

   (b) In general, the number of birth-planning procedures performed in China during 1971 to 1977 increased from the lowest counts in 1971 to the highest counts during 1975. The IUD was the most frequently used birth-planning operation in China, outnumbering sterilizations and induced abortions two to one.

   (c) In general, three factors affect the availability and use of various contraceptive procedures: political/legal, religious, and cultural. The Chinese government has actively supported contraceptive services and emphasized low-cost, highly effective methods such as the IUD and sterilization. Their comprehensive policy reduced the birthrate from 34 per 1,000 in 1970 to 18 in 1979. Although difficult to quantify, cultural factors and religious prohibitions both have important effects on contraceptive use. For example, the acceptance of vasectomy is low in areas of the world where women are presumed to have greater responsibility for contraception than men.

6. Consider the data in Figure 6.21.

   (a) Maternal death rates are examined according to birth order.

   (b) In Matlab, Bangladesh, during 1968 through 1970, there was an increasing risk of maternal death with increasing birth order. The number of maternal deaths per 100,000 live births was lowest for mothers

delivering their second or third infant; the maternal death rate increased to approximately 750 deaths per 100,000 live births among women delivering at least their eighth infant.

(c) One should always consider the effects of age when interpreting disease or reproductive health outcomes. Women who are delivering their eighth infant will, as a group, be older than women delivering their second or third infant. Because maternal mortality increases with age, some of the association between maternal deaths and birth order is probably age related.

7. Consider data in Table 6.22.

(a) Current methods of contraception among currently married women aged 15 to 49 years are described according to year of survey and urban or rural residence status.

(b) For most methods of contraception, usage is higher in urban areas than in rural areas. In 1976, contraceptive use in urban areas was more than 4 times greater than in rural areas; by 1983, the difference was reduced, and contraceptive use was only 2.5 times higher. In 1976, women used the pill more than 4 times more frequently than other methods. By 1983, the percentage of women using the IUD was slightly greater than the percentage using the pill.

(c) One possible explanation for the observed differences between contraceptive use in urban versus rural settings is access to medical care. If health services that dispense contraceptive services are unavailable in specific areas, or if women have to travel long distances to receive contraceptive services, levels of use may be lower. Other explanations might include socioeconomic reasons (pill prescriptions and IUD insertions are expensive), local culture (some subgroups of the population may desire more children), and education (some subgroups of the population may not know about specific methods of contraception or how to use them effectively).

References

Anderson JE, Abdul-Aziz A, Morris L, Shreydeh B. Fertility trends and determinants in Jordan. Int Fam Plann Persp 1985;11:47-51.

Centers for Disease Control. Comprehensive plan for epidemiologic surveillance. Epidemiology Program Office, Atlanta, 1986.

Chen LC, Gesche MC, Ahmed S, Chowdhury AI, Mosley WH. Maternal mortality in rural Bangladesh. Stud In Fam Plann 1979;5:334-341.

Chen P, Kols A. Population and birth planning in the People's Republic of China. Popul Rep 1982;10(1).

Fox JP, Hall CE, Elveback LR. Epidemiology: man and disease. New York: MacMillan Publishing Company, 1970.

Freidman GD. Primer of epidemiology. 2nd ed. New York: McGraw Hill, 1980.

Goldberg HI, Rodrigues W, Thome AMT, et al. Infant mortality and breast-feeding in north-eastern Brazil. Popul Stud 1984;38:105-115.

Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company; 1987.

Hermalin A, Entwisle B, eds. The role of surveys in the analysis of family planning programs. Liege, Belgium: Ordina Editions, 1982.

Hobcraft J, McDonald J, Rutstein S. Child-spacing effects on infant and early child mortality. Popul Index 1983;49:585-618.

Hogue CJR, Rubin GL, Schulz KF. [Unpublished report]. Reproductive epidemiology in a nutshell. [1985]. Located at: Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, Georgia.

Hogue CJR, Buehler JW, Strauss LT, et al. Overview of the national infant mortality surveillance (NIMS) project. Public Health Rep 1987;102:126-137.

Khan AR, Jahan FA, Begum SF. Maternal mortality in rural Bangladesh: the Jamalpur district. Stud in Fam Plann 1986;17:7-12.

Lilienfeld AM, Pedersen E, Dowd JE. Cancer epidemiology: methods of study. Baltimore: John Hopkins Press, 1967.

London KA, Cushing J, Rutstein SO, et al. Fertility and family planning surveys: an update. Popul Rep 1985;13(4).

MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown and Company, 1970.

National Center for Health Statistics. Seasonal variation of births, United States, 1933-1963. Rockville, Maryland: National Center for Health Statistics, 1966; DHEW publication no. (PHS)1000, (Vital and health statistics; series 21; no 34, p. 41).

National Research Council. Contraception and reproduction: health consequences for women and children in the developing world. Washington, DC: National Academy Press, 1989.

Ojo VA, Adetoro OO, Okwerekwu FEO. Characteristics of maternal deaths following cesarean section in a developing country. Int J Gynecol Obstet 1988, 27:171-176.

Rinehart W, Kols A, Moore SH. Healthier mothers and children through family planning. Popul Rep 1984;12(3).

Rochat RW, Jabeen S, Rosenberg MJ, et al. Maternal and abortion related deaths in Bangladesh, 1978-1979. Int J Gynaecol Obstet 1981;19:155-164.

Rochat RW. The magnitude of maternal mortality: definitions and methods of measurement. In: Landy U, Ratnam SS, eds. Prevention and treatment of contraceptive failure. New York: Plenum Press, 1987:201-214.

Royston E, Lopez AD. On the assessment of maternal mortality. World Health Stat Q 1987;40:214-224.

Singh S, Casterline J. The socioeconomic determinants of fertility. In: Cleland J, Hobcraft J, eds. Reproductive change in developing countries: insights from the world fertility survey. New York: Oxford University Press, 1985:199-222.

Tanner JM. Growing up. Sci Am 1973 Sep;229(3):34-43.

Tietze C. Maternal mortality excluding abortion mortality. World Health Stat Rep 1977;30:312-339.

Warren C, Morris L, Hiyari F, et al. [Unpublished report]. Jordan husbands' fertility survey. [1987]. Located at: Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, Georgia.

Zhang L. Birth control, late marriage, and decline in the natural increase rate of population. Popul and the Economy 1980;1:35-39.

Zimbabwe National Family Planning Council. Zimbabwe reproductive health survey, 1984. Columbia, Maryland: Westinghouse Public Applied Systems, 1985.

Chapter 7 - Learning Objectives

After completing this chapter you should be able to:

1. Define the following terms:
   probability sample
   self weighting
   sampling frame
   primary sampling unit (PSU)
   probability proportional to size (PPS)
   stratification
   cluster sampling
   design effect
   segmentation
   sampling interval
   equal probability of selection method (EPSEM)
   weighting
   design weights
   coverage error
   nonresponse
   target household
2. Recognize general principles guiding sample design.
3. Recognize when the principle of self-weighting may be ignored.
4. Identify factors influencing sample size.
5. Order the steps in carrying out sampling operations.
6. Identify advantages and disadvantages of using a preexisting sample.
7. Identify factors influencing size of clusters.
8. Identify advantages and disadvantages of listing and segmentation and their alternatives.
9. Identify safeguards to prevent biased selection of sample segments and dwellings.
10. Specify how to estimate the number of households needed in a survey.
11. Recognize the importance of documenting the sample.
12. Recognize how to introduce weights into an analysis.
13. Carry out a systematic sampling procedure.
14. Recognize how to use weights to:
    compensate for unequal selection probabilities
    compensate for nonresponse
    compensate for a difference between the distribution in the sample and the population distribution of a characteristic.
15. Recognize the impact of disproportionate sampling.
16. Recognize when weighting is necessary.
17. Carry out PPS interval-based selection.
18. Identify ways to reduce coverage error.
19. Recognize what may be done when coverage must be restricted.
20. Recognize problems in estimating variances with standard computer programs.

7 Survey Sampling

Introduction

Chapter overview

This chapter presents guidelines and general recommendations for designing samples, using sampling frames, weighting, and implementing sampling plans. When possible, a survey statistician, or someone experienced in survey design, should be involved in planning and executing a sample survey.

This chapter specifically addresses sampling issues in surveys, but much of the information can be generalized to other types of data collection activities, such as abstracting medical records or sampling hospital discharge and physician office records. Its focus is primarily on multistage household-based surveys in which households or residences are selected, and in which respondents from these residences are chosen for the survey. The methodology is also applicable, with some modifications, to surveys based on the selection of individuals from other populations (e.g., schools or clinics, instead of households). The methodology can be applied to innumerable types of populations (i.e., infants, women of reproductive age, agricultural workers), as long as the researcher has adequate information on the population of interest to select a representative sample of that population.

General Principles of Sampling Design

Probability and nonprobability sampling

Sample design for cross-sectional and descriptive surveys should be guided by several general principles, although some modifications of these principles may be required in specific situations. First, scientific probability sampling must be used. A probability sample is a sample in which all members of the population of interest have some known probability of selection and in which this probability is greater than zero. A nonprobability sample, on the other hand, is one in which probabilities of selection are unknown or, for some population members, are zero. In a nonprobability sample, for example,

information intended to be representative of an entire population would in fact be collected only from clinic attendees or people visiting the local marketplace. Clearly such samples may differ in important ways from the population as a whole. Nonprobability methods are often less expensive and easier to use than probability samples, but they should be avoided because the quality of the estimates obtained from a nonprobability sample is usually impossible to determine. Certainly, nonprobability methods cannot provide the confidence necessary to adequately explain unexpected findings. The use of nonprobability methods may also lead to criticism of the survey design when unexpected findings occur.

Self-weighting sample design

Second, samples should be self-weighting unless there is good reason to depart from this approach. A self-weighting sample is one in which no component of the population intentionally makes up a larger or smaller proportion of the total sample than its proportion of the population. Samples that are not self-weighting require the computation of weights to make the sample representative of the population. Computing weights can be an appreciable burden. The investigator must carry them as part of the data base, assess when and how they should be applied to statistical tests on weighted data, and correctly report their use. Alternatively, these considerations may be counterbalanced by certain advantages of adopting unequal sample weights in some circumstances.

Preexisting sampling frames

Third, if an adequate preexisting sampling frame is available, it should be used. Integrating study activities with other survey activities is economical and helps in the coordination of survey activities.

Fourth, the sample design should be as simple and straightforward as possible to facilitate planning and data collection and to simplify analyses.

In the following sections, general recommendations are presented for a number of specific aspects of sample design. In the first part of the chapter, technical issues are treated briefly. The second part presents these issues in greater detail and includes a number of examples.

Sample Size

Determining sample size

The determination of sample size involves both technical and practical decisions. The larger the sample the more elaborate and detailed are the analyses that can be sustained. The choice of

sample size involves balancing the demands of analysis against the capability for data collection, funding, and resource constraints.

Larger samples are more difficult to manage and supervise. If the fieldwork period cannot be extended, and if survey resources are limited, inflated sample sizes should be avoided.

Factors affecting sample size determination

As the number of subpopulations or strata for which stable estimates are required increases, so does the overall required sample size. If any of the objectives of the survey include the study of rare events, the sample size may have to be increased. For instance, the size of a sample may have to be increased for an analysis of mortality of a population with low mortality rates or for a characterization of individuals who exhibit an unusual condition. The size requirements of a sample also increase if a primary goal of the survey is to detect whether a statistically significant change in a parameter has occurred over time.

Various factors may affect sample size. First, the key variables must be considered. If the survey is a multipurpose one, as most cross-sectional surveys are, no simple formula can be used to determine the optimum sample size.

One suggestion for designing multipurpose surveys is to base sample size calculations on the key variables that require the largest sample and to use a preliminary estimate for the expected level of that variable in the population to be sampled. When sample size is based on the expected occurrence of a particular variable, standard sample size formulas can be used.

Basic Sample Design

Multistage designs

Most population-based surveys employ a multistage design. In a typical survey, geographic areas are selected first, then households are chosen within the selected geographic areas, and finally individual respondents are chosen from the selected households. At each stage the selection procedures can differ greatly from each other, and each stage has different potential problems and considerations.

Sampling Frames

To design a sample that is representative of a population,

Enumeration or census districts (EDs)

survey personnel must have a valid sampling frame—a recent compilation of population sizes, numbers of households, or other appropriate units for all subareas constituting the geographic area(s) to be included in the study. Most countries possess convenient area sampling frames in the form of the enumeration or census districts (EDs) from the most recent population census. Sketch maps and population size estimates are often available for the EDs. In most countries, EDs do not vary greatly in population size. However, many developing countries have no satisfactory lists of dwellings, households, or individuals, and no address system outside the more affluent parts of cities. In general, survey personnel have to make their own lists or maps. Sometimes these lists or maps can be obtained from other surveys or by selecting a subsample from a master sample. If EDs do not exist or if they are not suitable as a sampling frame, other frames, such as populations for administrative districts or records from government agencies and service providers, may be available.

Primary sampling units (PSUs)

Probability proportional to size (PPS)

Once an adequate sampling frame has been obtained, sampling may begin. In the first stage, primary sampling units (PSUs), which are usually EDs or other suitably small geographic or administrative units, are selected. Selection is usually done with probability proportional to size (PPS)—the probability of selecting a given unit is determined by its proportion of the entire population under study. For example, a unit that is twice the size of another unit has twice the probability of being selected as the smaller unit. PPS selection ensures that every individual unit (people or households) within the population has an equal probability of selection. Examples of how PPS sampling is implemented are presented in the second part of this chapter.

Segmentation

PSUs (most often census EDs) are often too large (typically 1,000-2,000 persons) to be economically feasible for survey personnel to list all households in the sample of PSUs. The survey personnel, therefore, must often segment PSUs into smaller areas before households are listed and selected. In many cases, the census maps are not accurate enough for segmentation to be done in the office. A field operation may be needed to select EDs for mapping and segmentation.

Preexisting Samples

A preexisting sample can be used only when the investigators are confident that it was well designed and that it meets the needs of the new survey. The investigators must also be aware of the weighting and stratification procedures used in the preexisting sample or of any unconventional sampling procedures that may result in a different number of interviews than are needed for the survey. The survey statistician must design a subsampling procedure (i.e., sampling from the preexisting sample) that produces a suitable sample for the survey. Although this will not always be possible, investigators are afforded more flexibility in developing a subsample design when the parent sample is larger than the desired subsample.

The decision to use a preexisting sample depends on whether the listing of dwellings, households, or other appropriate units is relatively current or whether it needs to be updated. If updating is required, the use of a preexisting sample may not be worthwhile.

Using a preexisting sample has two potential advantages: economy and increased analytic power through cross-analysis of two or more surveys. The disadvantages are adapting the sample to survey requirements and repeating interviews with the same households or individuals, which can result in respondent fatigue and contamination—answers to follow-up interviews are influenced by answers given in the first survey.

Often, investigators can use all but the final stage of sampling (selection of households or individuals) from a preexisting sample. Then a new cluster of households can be selected from the PSUs selected in the first survey. This eliminates any problems that might result from selecting the same households in both surveys.

Stratification

Strata and stratification variables

Stratification consists of dividing the population into mutually exclusive groups or *strata* according to some key variable or characteristic and then selecting separate samples for each stratum. Samples are commonly stratified according to factors such as sex, place of residence (urban or rural), or regions of a country. In Table 7.1, to study the relationship between maternal behaviors during pregnancy and pregnancy outcomes, investigators

used birth certificate data (birthweight and race) to stratify infants who were born in State A in 1990. A simple random sample (n = 400) was selected within each stratum, except for the stratum of birth certificates with unknown birthweights. Because the frequency of certificates with unknown birthweight was low, the investigators included in the sample all certificates with unknown birthweight. The overall sample size was 1,765 birth certificates.

Table 7.1

Stratification by Birthweight and Race

Birthweight by Race	State A Births	(%)	Sample Size
< 2,500 g Black	3,363	(2.5)	400
≥ 2,500 g	20,648	(15.3)	400
< 2,500 g Nonblack	6,150	(4.6)	400
≥ 2,500 g	104,500	(77.5)	400
Unknown	165	(.1)	165
Total	134,826		1,765

Stratification is generally performed as a part of the first phase of selection. Before this selection, all PSUs are divided into strata. A sample of PSUs is then selected from each stratum to obtain the proportionate mix of the desired strata.

Stratification helps ensure that groups of particular interest have large enough samples for analytic purposes and increases efficiency. When stratified samples are used, appropriate weights must be applied to the data for tabulations that include the total population, since probabilities of selection differ for units in different strata. In Table 7.1, the probability of selection varied from 0.0038 (400/104,500) among nonblack infants weighing at least 2,500 g to 0.1189 (400/3,363) among black infants weighing less than 2,500 g to 1.0 (165/165) for infants with unknown birthweights.

Cluster Sampling

In the ideal situation, survey sampling consists of simple random selection of households or respondents. For practical reasons related to cost, time, and management, however, a simple random selection is rarely possible. The types of surveys discussed in this chapter usually use cluster sampling. This type of sampling involves selecting groups of households within chosen sampling units, such as EDs. Although clustering reduces the cost and effort involved in collecting data, the technique has its own costs. Because individuals who live near each other tend to be more similar than individuals who live far apart, clustering increases the standard error and sample variance and functionally reduces sample sizes. The use of cluster sampling also requires that certain questions be addressed before the sampling takes place. For example, How large should the clusters be? How is the cluster starting point selected? How does one choose succeeding sample points after starting? How many interviews should be expected from a cluster that contains a given number of households?

Design effect

The design effect is the ratio of the variance or standard error of a sample (for a given variable) to the variance for a randomly selected sample. This measure is commonly used to quantify the effect of using a clustered sample. The actual sample size divided by the design effect gives the effective sample size—the size of a simple random sample required to yield the same variance or standard error (depending on which was used to estimate the design effect).

Effective sample size

The optimum number of potential respondents to be selected per cluster depends on the variables under consideration. Geographic clustering refers to the tendency for persons who live near each other to be more similar than randomly selected individuals with regard to the variables of interest. As a general rule, when clustering of the primary variable of interest is extensive, the cluster size should be small. For highly clustered variables, such as immunization status or use of health care services, calculations from past surveys suggest an optimum average number of completed interviews, or *take*, of 15 to 20 respondents per cluster. For variables that tend to be less clustered, larger clusters can be used.

The optimum take

A larger number of interviews per cluster reduces the costs of

travel, staff time, and logistics. However, extremely large cluster takes should be avoided in multipurpose surveys, or in those surveys in which many different variables are being measured, but no one variable has greater importance than the other.

For multipurpose surveys, a cluster take of about 40 respondents is recommended in the rural sector and of approximately 20 respondents in the urban sector. The cost advantage of large *takes* is generally smaller in the urban sector than in the rural sector. Costs can be further reduced when a preexisting, recent household list is available because a major factor favoring a large *take* is savings in listing operations. If the overall sample size is small (fewer than 2,000 interviews), the size of the clusters should also be small in order to increase the number of clusters, thereby ensuring that the variation within a population is adequately represented.

Listing and Segmentation

Household listings and map construction

Listing households or constructing maps of neighborhoods before selecting a sample is an appreciable but essential field cost. Investigators should avoid having the interviewers combine listing, sampling, and interviewing into a single operation while they work through the sample area. Even less acceptable is avoiding listing altogether by having the interviewers create clusters as they go along or by having them select a sample at fixed intervals during a random walk up to a predetermined quota of households or to individuals who have completed interviews. Although designed to eliminate conscious choice in selection, these methods fail to meet the requirement that the sample be selected to give a known and nonzero probability to every potential respondent. They also permit bias through intentional or unintentional avoidance of households. For example, interviewers may avoid houses with poor accessibility, houses that are not seen from the road, or houses that look uninviting. An independent listing can serve to determine the completeness of the interviewers' work, and the investigators can determine if the interviewers covered the whole sampling segment or if they omitted or substituted households.

Segmentation

The cost of listing can be reduced by using segmentation—breaking sectors into two or more roughly equal-sized pieces to decrease the size of the area that has to be listed. One of the segments is then randomly chosen. However, segmentation generates its own costs and requires skills in map making and in map interpretation. Because

it is usually difficult to delineate the natural boundaries of very small segments, segmentation becomes progressively harder as segments become smaller. A point is reached where additional segmentation is not useful. In planned urban areas, listing is much easier. Usually, an address system is in effect and segmentation is not needed.

Sometimes listing can be avoided by making segments so small that they are equal to the required cluster size. One could then use a *take-all* rule at the last stage of sampling. In urban areas, the difficulty of delineating small segments can be reduced by defining blocks or single buildings as segments.

The Sampling Operations

Selection of dwellings

After the primary sampling units have been selected (EDs or other areas) and the dwellings have been listed, the survey personnel must determine which dwellings will be included in the sample. During this step, investigators should provide safeguards to prevent biased selection. Preferably, the selection process should be performed in the office or in the field by a senior staff member. The selection of dwellings from the list can be done by either clustering consecutive dwellings with a random start or by systematically selecting from the list.

The first option is preferable, especially in rural areas, in terms of cost, time, and logistics. The second option, however, reduces the clustering effect by incorporating more variation within clusters. Systematic selection involves numbering the dwellings from 1 to N, randomly selecting one sample unit from the first k units, and subsequently selecting every kth sample unit to obtain a sample of predetermined size n (Cochran, 1963).

Household interview

Finally, the interviewing team visits the area and an interviewer is assigned to each selected household. The interviewer generally begins with a brief household interview, lists the household members, and identifies all persons who usually live in the household. The purpose of this listing is twofold. First, it yields a roster of household residents who are eligible for an interview, and second, it permits comparisons of basic characteristics (such as age) of the individuals who are interviewed and those who are not.

Respondent eligibility

Opinions differ as to whether only one person or all people within a household who meet the predetermined eligibility

requirements (e.g., all males, all married females, all teenagers, all users of a given health program) should be targeted for interview. From a statistical and a cost perspective, all potential respondents should be interviewed. However, *contamination*—interviewing more than one person in a household—may bias the responses on sensitive issues. If only one respondent is chosen per household, this person must be randomly selected from the household list of individuals eligible to be respondents. Household members who are home when the interviewer first visits or who are willing to be interviewed may differ substantially from other household members (e.g., housewives or the unemployed are most likely to be home) and thus may bias the results.

Contamination

If the members of a household or an eligible person are not available, the interviewer must make callbacks, when feasible, on different days and at different times of the day before the interview is abandoned. When and how many callbacks are made depends on a number of factors, including time and budget constraints and the ease with which the household can be visited.

Interviewer callbacks

Estimating the Number of Dwellings to Visit

The survey design must also include an estimate of the number of dwellings[1] or households that must be visited in order to achieve the desired sample size. Estimates must take into account the number of vacant dwellings that are likely to be encountered, the proportion of dwellings without eligible respondents, refusal rates, noncontact rates (i.e., when no one is home or when the selected individual is not home), and the number of eligible respondents per dwelling (if all eligible persons are to be interviewed in the dwellings). By starting with the desired number of interviews, we can multiply by each of the relevant rates to estimate the number of dwellings to visit. In Example 7.2, we illustrate this procedure by using a hypothetical survey.

Vacant dwellings, refusal rates, noncontact rates

[1] The terms *household* and *dwelling* can have different meanings in different societies. Dwellings generally refer to physical entities that would normally contain a household or family, although one dwelling could conceivably contain more than one household. Households can be defined in many ways but are generally viewed as one or more people sharing a dwelling and functioning as a unit (e.g., sharing the same cooking fire or sleeping under the same roof). In practice, in countries where extended families are the norm, households and dwellings are not always easy to identify. Survey researchers must determine the most appropriate units for sample selection based on their knowledge of the population and of the principles of survey sampling.

Example 7.2

Estimating the Number of Dwellings to Visit
All Eligible Women or One Eligible Woman per Dwelling

In Country X, public health officials decided to interview approximately 1,000 married women 15 to 49 years old. Five hundred interviews would be conducted in urban areas and 500 in rural areas. How many dwellings must be visited to obtain the required number of interviews if all eligible women in each selected dwelling are to be interviewed? How many if one eligible woman in each dwelling is to be interviewed?

	Selected Dwellings			
	All Eligible Women		One Eligible Woman	
	Urban	Rural	Urban	Rural
Required number of completed interviews	500	500	500	500
Percentage of Dwellings occupied*	.98	.98	.98	.98
Mean number of women or percentage of occupied dwellings*	.90†	1.10†	.65‡	.65‡
Percentage of eligible women contacted*	.90	.95	.75	.75
Percentage of contacted women interviewed*	.98	.99	.98	.99
Estimated dwellings to visit	643	493	890	723

*Percentages listed are hypothetical but are similar to those typically observed in developing countries.

†Mean number of eligible women per occupied dwelling. From census, earlier survey, or other data source.

‡Percentage of dwellings occupied by eligible women.

To obtain the desired number of completed interviews, we estimate that about 643 urban dwellings (500 ÷ .98 ÷ .90 ÷ .90 ÷ .98 = 643) and 493 rural dwellings (500 ÷ .98 ÷ 1.10 ÷ .95 ÷ .99 = 493) must be visited under the given set of assumptions (Example 7.2). These sample size estimates are based on the estimates of the percentage of occupied dwellings in each setting, an estimate of the mean number of women per occupied dwelling, the expected percentage of women who can be contacted, and the expected interview completion rate.

As expected, if only one woman is selected from each household, more dwellings would have to be visited to obtain the desired number of interviews. A significant increase in cost may result because of the additional time needed to visit more dwellings.

Documenting the Sample

The task of the survey sampler is not complete until every step and detail of the sampling operation has been carefully documented. Documentation is required to compute sampling errors, to link other data sources, and to perform various checks and supplementary studies. Experience has shown that special efforts are needed during this operation to ensure that the work is carried out effectively. Documentation is particularly important at the time of sample design, at the end of the field work, and at the completion of the data file. If documentation is delayed, considerable effort will be required to reconstruct the missing information when it is needed.

Selected Sampling Techniques and Issues

Disproportionate Sampling

EPSEM

It is usually preferable to design surveys so that each final unit of selection in the population (e.g., every patient, every child, every household) has an equal probability of selection. This sample design is known as the equal probability of selection method (EPSEM), and results in a self-weighting sample. The results of a self-weighting sample can be treated as directly representative of the entire population under study, and weights are not needed in the analysis.

This section deals with departures from this simple model and, in particular, with the deliberate introduction of different probabilities

of selection in different strata of the sample. Even in these departures, however, EPSEM should be used whenever possible within strata.

Reasons for disproportionate sampling

There are two main motives for disproportionate sampling between strata or domains:

- Cost efficiency is increased if a larger sampling fraction (f) (see 7.4.1) is used when the population variance is greater and the unit costs are lower. Thus, sampling fractions may be manipulated to reach an optimum design. For instance, if stratum A is very homogeneous relative to stratum B with regard to a key variable, stratum A may be undersampled relative to stratum B, since researchers might assume that a larger sample from A would yield little additional information.

- The survey planner may wish to report findings for a population subgroup that constitutes only a small percentage of the whole population. If a fixed sampling fraction (i.e., EPSEM) is used everywhere, this small group will be allocated a correspondingly small sample (possibly too small) on which to base valid conclusions. The problem can be resolved by oversampling the small subpopulation, thus reducing its sampling error. When considering the *whole* population, introduce a weight to compensate for the unbalanced sampling.

Example 7.3 illustrates disproportionate sampling. The basic idea is first to allocate the sample proportionately to the strata and then to adjust these allocations by some factor based on the amount of oversampling done.

To determine the size sample necessary to maintain the same sampling precision for the total-country estimate as existed with the proportionate sample, we use the disproportionate sampling rates in column 3. In general, disproportionate sampling would increase the standard error for the total-country estimate. Furthermore, an increase in sample size would be needed to compensate for the increase in standard error that results from disproportionate sampling.

Example 7.3

Disproportionate Sampling
A Hypothetical Survey

A country has an urban population of 1 million and a rural population of 4 million. We wish to select a nationwide sample of 5,000 people. To ensure more precision in the estimates for the relatively small urban sector, we decide to double the urban sampling fraction but maintain the same total sample size. What do we gain? What do we lose? How do we proceed?

	1	2	3
Stratum	Population Size (N_h)	Proportionate Sample Size (n_h)	Disproportionate Sample Size (n'_h)
Urban	1,000,000	1,000	2,000 k = 1,667
Rural	4,000,000	4,000	4,000 k = 3,333
Total	5,000,000	5,000	6,000 k = 5,000

Column 1 gives the population estimates, N_h.

Column 2 allocates the given total sample, n, of 5,000 between urban n_1 and rural n_2 in proportion to the population in column 1.

In Column 3, to ensure more precision for estimates for the urban domain, we begin by doubling the sample size for the urban domain (1,000 * 2 = 2,000) and leave the rural domain unchanged (4,000). This would yield a total of 6,000 interviews, 1,000 interviews more than the predetermined sample size of 5,000. To adjust the sample size back to 5,000, which we consider a fixed number, we introduce a factor k. k is computed as the ratio of the total sample size (in this example 5,000) to the sum of n'_h (in this example, 6,000). Each n'_h is then multiplied by the factor k (5,000/6,000) to obtain the disproportionate sample sizes (1,667 for the urban domain and 3,333 for the rural domain).

Neither the advantages nor the disadvantages of disproportionate sampling are very substantial unless the sampling fractions depart considerably from equality. In Table 7.4, the two-to-one oversampling reduced the urban sampling error substantially but increased the rural and total sampling errors slightly. In general, if strata vary by population size, the effects of using disproportionate samples instead of equal samples in all regions are still minor. The relatively few benefits of a disproportionate sample need to be balanced against the disadvantages of a weighted sample. In many cases, the benefits of proportional sampling will not justify the inconvenience.

Sample Weighting

Probability of selection

If a sample of size n is selected from a population of size N, then the probability (PR) of selection for any sample unit is:

(7.4.1)

$$PR\ \{\ \text{selection for any sample unit}\ \} = \frac{n}{N} = f$$

The probability of selection or the sampling fraction (f) is the proportion of the population that is included in the sample. Any population or stratum total can be estimated by multiplying the corresponding sample total by the inverse of the selection probability N/n. The multiplier N/n is called a weight:

Weight

(7.4.2)

$$\text{Weight} = \frac{N}{n}$$

For example, if one stratum is oversampled by a factor of 2 (i.e., twice the number of interviews are conducted as would have been in a proportional sample), the weight for that stratum would be the inverse of 2 (i.e., 1/2 or 0.5). Every interview in that stratum would only be counted 0.5 times when results for the entire population are calculated. In Table 7.4, we demonstrate the use of weights in making population estimates. Weights may be introduced at all stages of the sampling process. They are most often used in first-stage sampling to compensate for undersampling or oversampling of the strata.

In Table 7.4, column 1 displays the sampling fraction for each

Table 7.4

Comparison of Weighted and Unweighted Results for a Hypothetical Population

	1	2	3	4	5	6	7	8
			Sample Data			Population Estimates		
Stratum h	Sampling Fraction f_h	Weight $w_h=1/f_h$	Total Women	Married Women	Women (%) Married (Unweighted)	Total Women	Married Women	Women (%) Married (Weighted)
East	1/1,000	1,000	900	810	90	900,000	810,000	90
West	1/2,000	2,000	1,000	800	80	2,000,000	1,600,000	80
Coast	1/2,500	2,500	1,200	900	75	3,000,000	2,250,000	75
Total			3,100	2,510	81	5,900,000	4,660,000	79

stratum. Column 2 lists the weights that are used to inflate the sample data of columns 3 and 4 to population estimates in columns 6 and 7. Comparing columns 5 and 8, we see that the weights do not affect the estimates within strata because the weight within any stratum is constant. However, the total in column 8, which is obtained by dividing the total of column 7 by the total of column 6, shows 79 percent instead of the 81 percent obtained from the unweighted sample data. The differential weighting has lowered the estimate slightly. Because an area of high prevalence (the East) was oversampled, unweighted estimate (81%) exceeds the weighted estimate (79%).

Because the weights appear in the numerator and the denominator of the final estimates, we can remove any common factor from the weights. For example, all the weights in column 2 can be divided by 1000 and become 1, 2, and 2.5. Thus, only the relative values of the weights must be considered when estimating means, rates, proportions, percentages, or ratios. However, weights should not be reduced if absolute numbers of individuals with a characteristic in the population are of interest. Weights are often used to estimate the total number of individuals in the population who have some characteristic or condition, who are at risk of some condition, or who use specified services.

Weighting when one person per household is interviewed

If only one eligible person per household is interviewed, rather than all eligible people, a second weight must be included. This weight is equal to the number of eligible respondents living in the household and is necessary because the interviewed individual represents all eligible individuals in the household. Furthermore, this weight compensates for the reduced selection probability of individuals in households with more than one eligible person.

Weighting for Nonresponse. Weights can also be used to compensate for nonresponse or the failure to obtain data for some of the sampled units. In Table 7.5, suppose that column 1 represents only the individuals actually interviewed and column 2 represents individuals selected for interview (see column 3 from Table 7.4). The completion rates—respondents as a percentage of these selected—are presented in column 3 of Table 7.5 and can be factored into the weighting system.

To compensate for nonresponse, the inflating factors could be increased from 1,000 to 1,000/0.9 = 1,111 for the East, from 2,000 to 2,000/0.8 = 2,500 for the West, and from 2,500 to 2,500/0.85

Table 7.5

Weighting for Nonresponse

Geographic Area	1 Individuals Interviewed	2 Individuals Selected	3 Completion Rate	4 Design Weight	5 Combined Weight (Unadjusted)	6 Combined Weight (Adjusted)
East	900	1000	.90	1.00	1.11	1.00
West	1000	1250	.80	2.00	2.50	2.25
Coast	1200	1412	.85	2.50	2.94	2.65

=2,941 for the Coast. The combined design and nonresponse weights, equal to (1/Completion Rate) * Design Weight, are shown in column 5 of Table 7.5. If these weights are adjusted and East is set to 1.0, the combined weights are: 1 for the East; 2.5/1.111 = 2.25 for the West; and 2.941/1.111 = 2.65 for the Coast. The combined weights differ from the sampling weights (1, 2, and 2.5) alone because they also compensate for unequal rates across the three regions. The West and Coast have lower completion rates than the East, so their weights have been increased relative to the rates for the East.

Completion rates

Noncompletion or nonresponse adjustments are an attempt to counteract potential biases by increasing the weights of the respondents to represent the nonrespondents. These adjustments, however, do not compensate for any bias that may result from systematic differences between respondents and nonrespondents within the strata.

Post-stratification weighting

Another form of weighting, known as poststratification weighting, compensates for differences between the sample distribution and the known population distribution for some characteristic(s). For example, even with a perfectly implemented equal probability sample, the age distribution in the sample will differ somewhat from the population age distribution because of sampling fluctuations. If the population age distribution is known (from the census, for example) the sample can be reweighted, age group by age group, to approximate the population distribution. The same type of weights N_h / n_h are used, where h designates the age group.

When Is Weighting Necessary? The overall effect of weighting on results is usually small. The main purpose is to provide the best possible estimates of a wide variety of population characteristics from the sample. If varying sampling fractions have been used in the sample design, these should be reflected in the estimates; in these circumstances, weighting should be considered obligatory even if the effects are small. Such weights are termed design weights. In an equal probability sample, no design weights are required, which is why such a sample is termed self-weighting. For nonresponse weights, the weighting adjustment corrects only a part of the nonresponse bias. Corrections will nearly always be trivial in developing countries, where response rates tend to be high. It is reasonable to omit nonresponse weighting if the sample is otherwise self-weighting. Alternatively, if design weights are to be used for strata, little complexity is added in modifying these weights to account for variations in nonresponse rates between strata. The combined weight for stratum h will be the design weight times the inverse of the completion rate.

Design weights

Finally, poststratification weighting or weighting to compensate for differences between the sample and the population is best avoided in most circumstances. This type of weighting is recommended only when the investigator is confident that the census or other data used as a standard are accurate or if the investigator strongly believes that severe undercoverage of people or areas with certain characteristics in the survey sample occurred.

In general, self-weighting samples are recommended. When the sample is not self-weighting, design weights, combined with nonresponse weights for the same strata, should be entered in the data record and used in all analyses. The standard method for introducing weights into the analyses is to include one or more weight variables on each individual record. In Table 7.5, for instance, the data records for each woman sampled in the East would have a value of 1 for the design weight variable, those in the West would have a value of 2, and those in the Coastal domain a value of 2.5. If there were weights for eligible household members, they would be included as a second weight on the data record. Finally a combined, or overall, weight should be included. If a weight variable is included as a separate variable on the data record, it is easy to include this variable in the calculations.

Adding weights to the data file

Systematic Sampling

Systematic sampling is the selection of units (e.g., PSUs, households, individuals, clinics) at fixed intervals from a list. Typically, the starting point for selection is randomly determined. Systematic sampling can be used at any stage of sampling. Compared with random selection, systematic sampling has three advantages:

- It is easier to perform.
- It allows the selection to be easily checked.
- It provides a degree of stratification with respect to the variable on which the list is based, if the list is in some order. That is, it helps ensure that no large segments of the population are omitted from the sample.

Because of these advantages, systematic selection is used much more often than random selection for choosing PSUs.

Systematic sampling is implemented in the following manner. First, the sampling interval I is calculated. If the sample design specifies the number of units to be selected (e.g., 100 enumeration districts), then

(7.6.1)

$$I = \frac{N}{n} = \frac{\text{Number of units of population}}{\text{Number of units of sample}}$$

Alternatively, if the sample design specifies a sampling fraction (f) or probability (e.g., 1/10 or .1 of the enumeration districts will be selected), then

(7.6.2)

$$I = \frac{1}{f} = \frac{n}{N}$$

Random start

After the sampling interval is calculated, a *random start* is selected. This start consists of selecting a random number, R, between 1 and I. On the complete list of units in the population, the Rth unit is the first one selected. I is then added to R to determine the second unit to be selected. The third unit is R + 2 I. This process continues until the list is exhausted, at which point n units will have been

chosen. In systematic sampling with probability proportional to size, any unit whose size equals or exceeds I is certain to be selected. In some cases, units larger than the interval may be selected two or more times. Table 7.6 shows an example of systematic selection of enumeration districts for a hypothetical survey.

Table 7.6

Systematic Selection of Sampling Units
(I = 4 and R = 2)

Enumeration District	Selection*	Enumeration District	Selection
1		11	
2	X	12	
3		13	
4		14	X
5		15	
6	X	16	
7		17	
8		18	X
9		19	
10	X	20	

The area to be surveyed consists of 20 enumeration districts. It has been previously decided that five of them will be sampled. Therefore, the sampling interval, I, is 20/5=4. R has been selected at random from numbers 1 through 4; R = 2 has been selected. Thus, the following districts are selected:

R = 2
R+I = 6
R+2I = 10
R+3I = 14
R+4I = 18

After the fifth selection, number 18, the list is exhausted.

* X denotes the enumeration district selected.

Sampling with Probability Proportional to Size

Systematic selection of sampling units has a serious deficiency when the populations of sampling units vary substantially. Systematic selection gives every unit an equal probability of selection, regardless of its size. Thus, systematic sampling without regard to size overrepresents smaller units relative to larger ones. Therefore, smaller units have larger selection probabilities relative to their population.

A common way to avoid this problem is to select sampling units with a probability proportional to their population size. Thus, if unit A is estimated to be ten times as large as unit B, it is given a ten times greater chance of being selected. This method is called sampling with probability proportional to size, or PPS sampling. After selected areas have been sampled with PPS, we can use clusters of the same size within each area. Selecting areas with PPS and using a fixed number of households within these areas approximates the sampling objective that all individuals in the population have the same probability of selection.

When the same number of households or individuals to visit in each area have been selected, the fieldwork can be organized so that each interviewing area requires about the same amount of time, personnel, and resources. Equalizing the work load in each area unit constitutes the main attraction of PPS sampling. The fieldwork is easier to organize if work loads do not vary substantially from area to area. This advantage is particularly significant when sampling ordinary administrative units (e.g., villages, neighborhoods, political jurisdictions), which may vary in population size.

PPS sampling

Selection using PPS may be performed as follows:

1. List the area units, with the estimated size M'_i shown for each.

2. Cumulate the values of M'_i and record this cumulative figure—cumulative M'_i —for each unit. Cumulative M'_i for the last entry should equal the population of the entire survey area.

3. Compute the sampling interval $I = M'/a$, where a is the number of area units to be selected.

4. Select a random number R between 1 and I.

5. Compute the sampling numbers: R; R + I; R + 2*I; R + 3*I; and so on.

6. For the first sampling number, find the first cumulative M'_i that equals or exceeds it. The corresponding unit is the one selected.

7. For each succeeding sampling number, follow the same procedure as number 6.

Selecting area units and households

Table 7.7 and Example 7.8 describe systematic sampling with probability proportional to the values M_i', the populations of the PSUs. Table 7.7 describes the first stage of selection (area units) for the urban part of the sample. For the second stage of the selection (households), selection is to be interval-based, rather than determined by choosing adjacent households. For each selected area unit, compute the household sampling interval (I) as follows:

(7.7.1)

$$I_i = 1/p_i = a * M'_i / f * M'$$

where

p_i = probability of selection in area i

a = 70 (number of PSUs)

f = 1/30.55 = .0327 (sampling fraction)

M′ = 38,500 (population of eligibles in survey area)

M'_i = estimated PSU sizes from Table 7.7.

I_i = 0.0555 * M'_i.

For the three units selected in Table 7.7 this gives:

Unit 001 I_i = 0.0555 * 150 = 8.3
Unit 007 I_i = 0.0555 * 110 = 6.1
Unit 011 I_i = 0.0555 * 140 = 7.8

Table 7.7

Selection of Area Units Using Systematic Sampling With Probability Proportional to Size

Area Unit	Size M'_i	Cumulative M'_i	Sampling Numbers	Area Units
001	150	150	90	Area units in survey area:
002	60	210		A = 350.
				Planned sample size:
003	80	290		a = 70 units. Sum of unit
004	70	360		sizes: M′ = 38,500.
005	130	490		Interval: I = M′/a = 550.
006	90	580		
007	110	690	640	Random number selected
008	140	830		between 001 and 550:
009	150	980		R = 090.
010	70	1,050		
011	140	1,190	1,190	Units selected into the
012	110	1,300		sample: 001, 007,
				011, etc.
.	.	.		
.	.	.		
.	.	.		
350	90	38,500		

Thus, in unit 001 every eighth household is chosen (after rounding 8.3 to 8), in unit 007 every sixth is chosen, and in unit 011 every eighth is chosen. The final steps involve listing all households in the selected areas, choosing a random starting point, and selecting a systematic sample of households using the above intervals.

If clusters of adjacent households are selected, the procedure is much simpler. A number between 1 and M'_i is randomly selected. The household on the area map or list corresponding to this number becomes the starting point. Each succeeding household on the map or list is visited until the designated number of households in each PSU have been visited.

Example 7.8

Estimating the Desired Number of Clusters

Given:	
Desired sample size	5,000 women aged 15 to 49 years
Number of women aged 15 to 49 years in entire country (extrapolated from census or survey data)	168,000 (N)
Number of women aged 15 to 49 years in urban sector	47,100
Sample selected to achieve desired sample size (allowing approximately 10% nonresponse)	5,500 women aged 15 to 49 years (n)
Desired cluster take (urban)	20 women interviewed/cluster
Computed:	Sampling fraction f 5,500/168,000 = 1/30.55 = .0327
Women selected in urban sector	47,100/30.55 = 1,542
Number of clusters selected (average of 22 women per cluster)	1,542/22 = 70

Errors of Coverage and Nonresponse

Coverage error refers to any lack of correspondence between the sample design and attempted interviews. Nonresponse refers to interviews attempted but not achieved. Thus, undercoverage occurs if the interview is erroneously not even attempted, and nonresponse occurs if the interview is unsuccessfully attempted.

Undercoverage. There are two main sources of undercoverage. The first occurs at the listing stage. In some surveys, listings do not completely cover the designated sample area for many possible reasons. As a result, large geographic areas and individual households within selected areas may be excluded. Second,

Listing stage

eligible persons can be excluded from the sample if the eligibility criteria for the interview, such as an age interval or other characteristic, is inaccurately reported. Both types of undercoverage errors can be caused by field-workers who seek to reduce their work load. Intentional errors can only be controlled by intensive training and close supervision. However, errors that result from inaccurately reporting an eligibility characteristic are not necessarily conscious or deliberate. For example, in many developing countries field-workers often have to estimate the ages of older respondents because their date of birth is unknown. Training interviewers to guess objectively, without bias, is difficult. In such instances, bias cannot be eliminated entirely.

Inaccurate reporting of eligibility criteria

An outdated sampling frame can also produce errors in coverage. These errors can be reduced by taking special steps to update the frame, especially in areas of known new or expanded settlement, such as new housing developments, squatter areas, and refugee camps (if they are to be included in the survey).

Deliberate Restriction of Coverage. In many surveys, especially nationwide ones, investigators deliberately exclude territories that are difficult to access or that would be excessively costly to survey. Two distinct situations exist:

Excluding specific areas from the sampling frame

- Exclusion of clearly identified areas from the sampling frame. The coverage limitations should be stated in the survey report, which then becomes a report on the remainder of the country. Such exclusions are not regarded as coverage or response error but simply as part of the definition of the survey domain. For example, areas where military conflict is ongoing, where the safety of survey workers may be jeopardized, or where sparsely or widely dispersed populations require an undue amount of time and resources to cover may be deliberately excluded from the sample. If areas must be excluded, they should constitute a coherent domain. For example, if some areas of a country are to be excluded because of inaccessibility, the exclusions should be done as complete provinces, enumeration districts, or other entities for which population estimates are available. A survey from which a number of scattered zones have been excluded is difficult to interpret and to use.

- Ad Hoc exclusions that are decided during or just before fieldwork. In developing countries, survey organizations often abandon fieldwork in certain sample clusters for various reasons such as, floods, civil disturbance, or practical constraints. These exclusions usually occur after sample selection. If the excluded areas form a meaningful domain, the exclusion problem may be managed by redefining the survey domain. More commonly, however, the excluded areas will not *make sense* and will have to be accepted as constituting error. This type of error should be classified as nonresponse rather than coverage error. Substituting new areas for those excluded is not recommended.

Ad Hoc exclusions

Nonresponse. The concept of nonresponse seems simple and clear. Specifically, nonresponse refers to the percentage of persons who should have been interviewed but were not. However, the seemingly straightforward distinction between nonresponse and undercoverage becomes somewhat unclear in practice. For example, the use of multistage samples and the replacement policy for nonrespondents complicate this simple issue.

First, in most household surveys the final interviews are identified through a multistage process. For example, in the typical survey, the researcher selects PSUs, lists and selects dwellings, collects information on the household members who live in the dwellings, and then interviews any eligible respondents in the households. If failure occurs at the second or third step, the information that permits accurate classification of nonresponse at the final level would not be available. For example, if the interviewer cannot find the selected dwelling, then whether it contains anyone eligible for interview is unknown. If the dwelling does not contain eligible individuals, the failure has no effect on the interview response rate. If the head of the household refuses to provide information on household members, the interviewer does not know if the household contains eligible persons for interview. Therefore, calculating nonresponse becomes problematic. There is no standard way of managing this and determining a single nonresponse rate.

Sources of nonresponse

Replacement of nonrespondents

Some surveys propose a policy on replacing nonrespondents. However, replacement of nonrespondents is generally not

Replacing nonrespondents

recommended because replacement tends to bias the survey results toward the characteristics of those individuals who are most easily located or most cooperative. Preferably, procedures should be setup to improve efforts to locate hard-to-find individuals. If families, instead of dwellings, are selected, and if the selected family has moved away between the listing and the interview, interviewing the family (if any) that moves in as a replacement sample unit is acceptable. Such situations do not require any special treatment since the target family sample is defined as the set of families found at the time of interviewing in the dwellings selected from the dwelling list.

Quota sampling. Interviewing until the target number of interviews has been achieved, called quota sampling, should be avoided. Often, surveys will be designed to ensure that the exact number of interviews needed for each PSU and for the entire survey are obtained. However, this procedure can introduce substantial biases into survey results. The problem stems from the fact that the survey quota is likely to have a disproportionately large number of easily found and easily interviewed respondents, who may tend to differ in relevant ways from the overall population. Although some uncertainty is introduced, interviewing only in preselected households is preferable.

Sampling Errors

After data collection is complete, the sample design should be kept in mind during data analysis. The sample designs for most surveys involve features such as stratification, clustering, and weighting. All these features affect the sampling errors of the survey estimates. The standard formulas for sampling errors found in statistical texts and used in most computer programs do not take these features into account; consequently, (most of) these programs are often inappropriate. Formulas for the estimated standard errors of proportions and means are usually intended for application only to simple random samples. These formulas tend to understate the sampling errors of the survey estimates, sometimes to an appreciable degree.

Use special software to estimate standard errors

The computations required to estimate sampling errors of survey estimates based on complex sample designs are more burdensome than those based on simple random samples. In recent years, a number of computer programs have been developed to perform these

computations (e.g., the RTI SESUDAAN procedure calculates standard errors for weighted, stratified, and clustered samples). The correct computation of sampling errors with complex sample designs requires knowledge of weights and of the stratum and primary sampling unit to which each sampled individual belongs. Therefore, this information must be recorded on each respondent's computer data record if survey estimates will be derived using survey software.

Nonsampling Errors

Typically, survey researchers focus on sampling errors during the analysis of data to estimate the likely range of the rates, proportions, and means within the population. As noted above, sampling errors, when calculated correctly, take into account the design of the survey—clustering, weighting, and stratification. However, sampling errors describe the range of overall population values if the sample and the population's only difference is that the entire population is not being interviewed. In fact, a variety of other potential sources of survey error, referred to as nonsampling errors, can influence the results. Nonsampling errors can originate from any source other than the design of the sample. Some common nonsampling errors include:

Sources of nonsampling errors

- Intentionally faulty information from respondents
- Unintentionally faulty information from respondents
- Poor questionnaire design resulting in lack of clarity (regarding exactly what information is being sought) by interviewers or respondents
- Intentional or unintentional interviewer errors
- Coverage errors
- Nonresponse
- Data entry and data processing errors

Minimizing nonsampling errors

Without going into a detailed discussion of each of these, we can make some general recommendations for minimizing nonsampling errors. Thorough training of interviewers and other field staff is an essential component. Effective supervision of interviewers by well-motivated and capable supervisors is also vital. Questionnaires should be designed to minimize refusals to answer

questions and to reduce anxiety about providing information on highly personal topics. Coding and data entry procedures should be uncomplicated. This includes designing forms to make these tasks as easy as possible. Unfortunately, the extent of nonsampling errors cannot be measured. The best solution is to take actions that minimize these errors.

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F A probability sample involves the selection of units with known probability.

   (b) T/F In a probability sample, it is generally acceptable for some area units to have a zero probability of selection.

   (c) T/F Using preexisting samples is usually more expensive than designing a new sample.

   (d) T/F Sampling frames are theoretical constructs that list all units of the population under study.

   (e) T/F If one enumeration district is twice as large as another and has twice the chance of being selected, this is probability proportional to size.

   (f) T/F Stratification is dividing and sampling a population according to a characteristic of interest.

   (g) T/F Cluster sampling requires a random starting point and saves money and time.

   (h) T/F It is generally agreed that selecting all adults at home when the interviewer arrives does not introduce potential bias.

   (i) T/F Weights can be used to inflate sample data to population estimates.

   (j) T/F Design weights are required for analyzing data from a self-weighting sample.

   (k) T/F If an interview is erroneously not attempted, coverage error occurs.

   (l) T/F Nonresponse relates to interviews attempted but not achieved.

   (m) T/F If available time and resources were unlimited, simple random samples would usually be the best survey design.

(n) T/F In a multipurpose study, no simple formula can determine sample size.

(o) T/F In a multipurpose study, sample size calculations should be based on the variable requiring the smallest sample.

(p) T/F Continuing from house to house until a predetermined number of interviews have been obtained is a recommended strategy for conducting a household survey.

2. Multiple choice. Select one response.

2.1 Which of the following is generally not an acceptable reason for exclusion from a survey?

a. Excluding a sparsely populated desert province that would require extensive resources to cover
b. Excluding villages that have been inaccessible due to flooding
c. Allowing an interviewer to omit a household because of safety concerns
d. Excluding villages that are difficult to get to because of a lack of public transportation

2.2 Which general principle is probably violated when the researcher selects a sample from among clinic workers when he is seeking information for the entire population?

a. Scientific probability sampling
b. Self-weighting sample
c. Using preexisting sampling frame
d. Using simple, straightforward design

2.3 The principle of self-weighting may be overridden when a population of interest

a. is a small proportion of the population.
b. comes from a previous study.
c. uses predetermined segments of the population.
d. is known to have high refusal rates.

2.4 All of the following are likely to require a larger than usual sample size except:

a. More strata
b. Study of rare event
c. Self-weighting sample
d. Multi-purpose study

3. Circle true (T) or false (F).

(a) T/F Use of a preexisting sample of a population is almost always advisable.

(b) T/F If the characteristic of major interest is extremely variable within clusters, very small clusters must be used.

(c) T/F The larger the number of interviews per cluster, the less expensive the design.

(d) T/F Combining listing, sampling, and interviewing into a single operation is an appropriate way to reduce costs.

(e) T/F Interviewers should be trained to create clusters while in the field.

(f) T/F Segmentation meets the requirement that the sample be selected to give a known nonzero probability to every potential respondent.

(g) T/F Household selection procedures should be designed to prevent interviewers from avoiding uninviting households.

(h) T/F Nonsampling errors can derive from sources beyond the control of survey researchers.

(i) T/F Segmentation reduces listing costs.

(j) T/F Listing can be avoided by using small clusters and a take-all rule.

(k) T/F Most survey researchers agree that interviewing all potential respondents in a household is always preferable to interviewing only some.

(l) T/F Those not available to be interviewed should be dropped and replaced by others who are available.

(m) T/F Respondents may be selected from a household based on who is at home at the time the interviewer comes.

(n) T/F The number of households to be visited can be calculated by multiplying the required number of interviews by the expected rates of vacancy, refusal, and other key factors.

4. Multiple choice. Select one response.

4.1 All of the following are reasons to preserve sampling documentation except:

a. Computation of sampling errors
b. Linkage with other data sources
c. Various checks and supplementary studies
d. Estimating sample sizes

4.2 All of the following are needed to introduce weights into the analysis except:

a. Include weight variable on each individual record.
b. Use computer program capable of handling weights.
c. Weight respondent according to eligible respondents in house.
d. Allocate weights according to a predetermined random scheme.

4.3 All of the following will reduce coverage error except:

a. Intensive training
b. Close supervision
c. Excluding parts of selected areas that are hard to reach
d. Updating area frames

4.4 Exclusions of clearly defined areas from the sample frame are:

a. Generally not problematic when analyzing data
b. Considered to be coverage errors
c. Classified nonresponse
d. Replaced by new areas

5. Circle true (T) or false (F).

(a) T/F Weighting can compensate for nonresponse by multiplying respondents in a stratum by the inverse of the response rate.

(b) T/F Each age group can be weighted by its proportional distribution in the population.

(c) T/F Gains and losses resulting from disproportionate sampling may be significant when changes in sampling fractions are small.

(d) T/F Because the overall effect of weighting is usually small, the design weights can be ignored.

(e) T/F Nonresponse weights can sometimes be overlooked.

(f) T/F Poststratification weighting is best avoided in many situations.

(g) T/F RTI SESUDAAN is recommended for computing standard errors of survey data that are not self-weighting.

(h) T/F Standard errors for complex sample designs are easy to derive using standard computer programs.

6. Arrange in order from 1 to 7 the steps in designing a sample:

   (a) ___ Select area units (for example, EDs).

   (b) ___ Interview eligible respondents in a household.

   (c) ___ Segment areas into smaller areas.

   (d) ___ Get a sampling frame.

   (e) ___ Select segments for listing.

   (f) ___ List and select dwellings.

   (g) ___ Interview households in a dwelling.

7. For the selection of a systematic sample from a list of enumeration districts, describe the steps used in the selection. Assume that there are 24 enumeration districts (EDs) that are approximately the same size and that there are eight EDs in the sample.

8. Select a systematic sample using PPS from the list of villages below. The sample should include five EDs.

Table 7.9

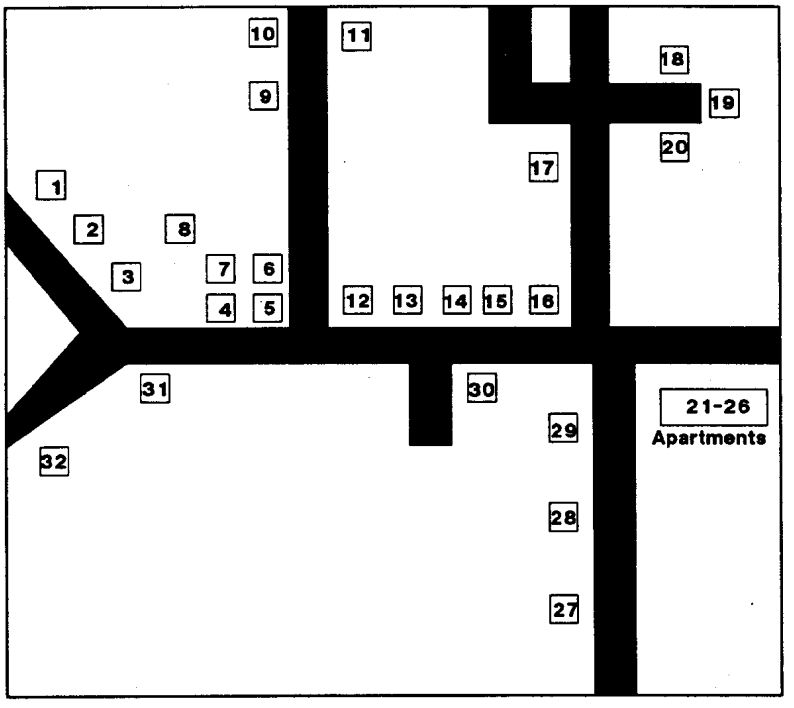
Systematic Sampling

Village	Population	Village	Population
1	90	11	30
2	20	12	120
3	240	13	20
4	60	14	580
5	400	15	90
6	10	16	160
7	190	17	100
8	180	18	220
9	70	19	50
10	340	20	30
Total			3,000

9. Assume that the map in Figure 7.10 is available for your use and that all households are numbered from 1 to 32 in a village that was selected for a survey. Select a sample of eight households in two different ways. Describe how you performed each selection.

Figure 7.10

Household Sample Selection



Suggested Answers to Practice Exercises

1. True or false.

   (a) T Introduction

   (b) F Introduction. No, in a probability sample, area units have a known nonzero probability of selection.

   (c) F Use of a Preexisting Sample. No, use of preexisting samples is less expensive if it can meet the needs of the survey being planned.

   (d) F Basic Sample Design. No, sampling frames are recent compilations of population sizes, numbers of households, or other appropriate units for all subareas composing the geographic area to be included in the study.

   (e) T Basic Sample Design

   (f) T Stratification

   (g) T Cluster Sampling and Sampling Operations

   (h) F Sampling Operations. No, opinions differ regarding whether all persons having the predetermined interview eligibility requirements within a household should be interviewed. Interviewing all potential respondents in a households may bias responses.

   (i) T Sample Weighting

   (j) F Disproportionate Sampling Between Strata. No, design weights are not required for analyzing data from a self-weighting sample.

   (k) T Errors of Coverage and Nonresponse

   (l) T Errors of Coverage and Nonresponse

   (m) T Cluster Sampling

   (n) T Sample Size

(o) F Sample Size. No, sample size determinations should be based on both technical and practical considerations.

(p) F Quota Sampling. No, quota sampling should be avoided since this procedure can introduce substantial biases into survey results.

2. Multiple choice.

2.1 d Deliberate Restriction of Coverage

2.2 a Introduction

2.3 a Stratification

2.4 c Sample Size

3. True or false.

(a) F Use of a Preexisting Sample. No, updating an out-of-date preexisting sample may not be economical.

(b) F Cluster Sampling. No, when clustering of the primary variable of interest is extensive, the cluster size should be small.

(c) T Cluster Sampling

(d) F Listing and Segmentation. No, listing should not be performed by the interviewer.

(e) F Sampling Operations. No, clusters should be formed in the office or by senior survey personnel in the field.

(f) T Listing and Segmentation

(g) T Listing and Segmentation.

(h) T Sampling Errors

(i) T Listing and Segmentation

(j) T Sample Weighting

(k) F Sampling Operations. No, see 1. (h).

(l) F Sampling Operations. No, study subjects who are not available for interview should not be replaced by others who are available. Callbacks must be made at different times and on different days.

(m) F Sampling Operations. No, respondents must be selected from all eligible members of the household.

(n) T Sampling Operations

4. Multiple choice.

4.1 d Documenting the Sample

4.2 d Sample Weighting

4.3 c Errors of Coverage and Nonresponse

4.4 a Deliberate Restriction of Coverage

5. True or false.

(a) T Sample Weighting

(b) T Sampling Operations

(c) F Sample Size

(d) F Sample Weighting. No, usually design weights are considered obligatory even if the overall effect of the weights is small.

(e) T Sample Weighting

(f) T Sample Weighting

(g) T Sampling Errors

(h) F Sampling Errors. No, standard errors for complex sample designs are difficult to derive and require using complex computer programs.

6. Arrange in order.

(a) 2

(b) 7

(c) 3

(d) 1

(e) 4

(f) 5

(g) 6

7. The selection steps follow:

a. Calculate sampling interval (24/8 = 3).
b. Choose a random start (1, 2, 3).
c. Beginning with the random start, cumulate by 3s to select 8 EDs.

8. The systemic sample may be selected as follows:

a. Calculate sampling interval (3,000/5=600).
b. Choose a random start between 1 and 600.
c. Cumulate populations and select the village when cumulative total is greater than RS, RS+600, RS+1200, RS+1800, RS+2400.

9. You would select the sample by:

   1. Selecting a cluster of 8 contiguous households by making a random start between 1 and 32.

   2. Design an interval sample:

      a. Calculate a sampling interval (32/8=4).
      b. Select a random start.
      c. Select every 4th household from the random start. (When you have a selection greater than 32, begin again with 1.)

References

Cochran WG. Sampling techniques. New York: John Wiley & Sons, Inc., 1963.

Kish L. Survey sampling. New York: John Wiley & Sons, Inc., 1965.

Shah BV. SESUDAAN: standard errors program for computing standardized rates from sample survey data. Research Triangle Park, North Carolina:1963.

Chapter 8 - Learning Objectives

After completing this chapter you should be able to:

1. Distinguish between parallel and successive treatment designs.
2. For clinical trials recognize how to:

   select a study population
   assign study volunteers to study groups
   achieve blinding

3. Understand the rationale behind defining comparison groups.
4. Recognize the importance of sample size in a randomized clinical trial.
5. Recognize the impact on study validity of exclusions before randomization.
6. Recognize the impact on study validity of exclusions after randomization.
7. Distinguish baseline data analysis from follow-up data analysis.
8. Identify methods used in analyzing data for clinical trials with a short follow-up.
9. Identify methods used in analyzing data for clinical trials with a long follow-up.
10. Recognize advantages and disadvantages of clinical trials.
11. Design a hypothetical randomized clinical trial in outline form.

8 Randomized Clinical Trials

Introduction

Clinical trial defined

The clinical trial is a controlled experiment that is used to assess the safety and efficacy of treatments for human diseases and health problems. The clinical trial is essential to the process of developing and accepting new treatments. When several approved treatments are available for the same condition, the clinical trial is used to determine if one treatment, usually a new treatment, is superior to the standard treatment(s). Two stages are involved in the development of new treatments.

Preclinical studies

Stage 1. The first stage of development involves laboratory experiments. These studies, also called preclinical studies, are conducted in vitro and with animals. Preclinical studies provide information on pharmacology and toxicology in preparation for studies planned in humans.

Stage 2. The second stage involves studies that use human participants. This stage is usually categorized into four phases:

- Phase I studies comprise the initial evaluation among human participants (20–100 subjects). The primary objective is to assess the safety and tolerance of the treatment.

- Phase II studies (100–200 subjects) evaluate the potential effectiveness of the treatment. Frequently, several competing new treatments are screened to select the one with the most potential. The optimal method of administering the treatment is determined in Phase II.

- Phase III studies evaluate the new treatment in a large number of subjects (500–1500) to establish the treatment's effectiveness and to gather additional safety information. The majority of Phase III studies are comparative clinical

Treatment groups

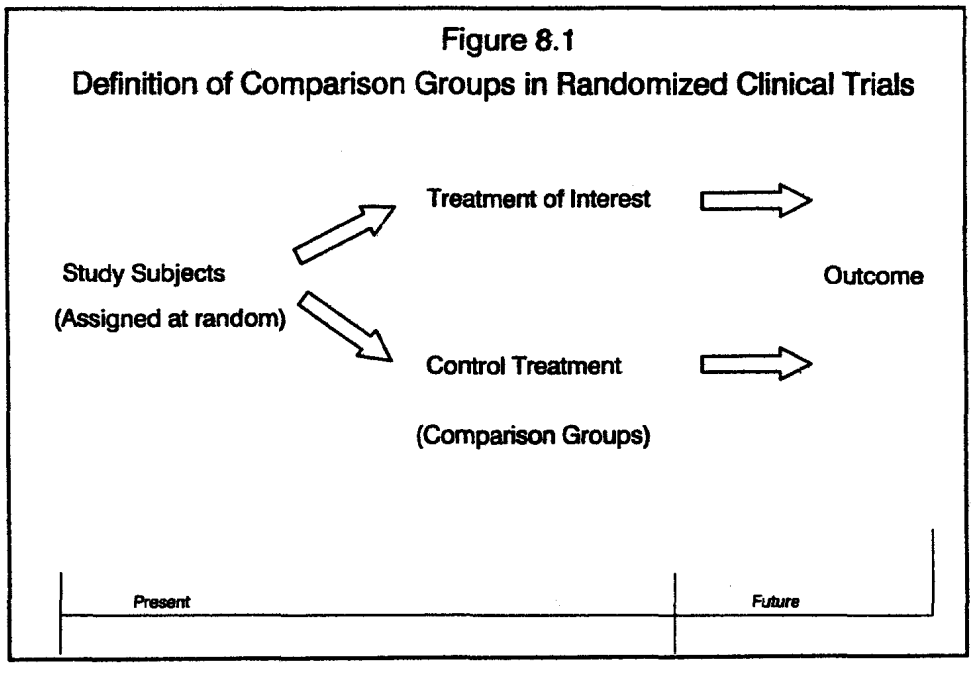
trials in which the comparison groups are the new treatment group (the treatment of interest) and the control treatment group. The research team assigns each study subject to one of these treatment groups. Ideally, treatment is assigned at random, and both the research team and the study subjects are unaware of, or blinded to, the exact treatment assignment.

- Phase IV studies, or postmarketing studies, investigate the long-term effects of the treatment; these studies are conducted after the treatment is approved for general use.

Randomized clinical trial

If the investigator assigns treatments at random, the study is known as a randomized clinical trial (RCT). From a scientific perspective, the RCT, with adequate sample size and blinding, is the preferred study design. Randomization should eliminate bias and is the preferred method of assigning individuals to treatment and control groups. The treatment groups are followed for the same time period to observe the occurrence of a particular event or outcome (Figure 8.1). Research hypotheses should be stated in terms of the anticipated outcome in the new treatment group. For analysis, the research team compares the rate of the outcome in the new treatment group and the rate of the outcome in the control treatment group.

Figure 8.1
Definition of Comparison Groups in Randomized Clinical Trials



In Example 8.2, investigators designed a randomized clinical trial to test whether multivitamins taken daily with oral contraceptives (OCs) reduced the side effects associated with the pill.

Example 8.2

Multivitamins With Oral Contraceptives

Women were assigned at random to the treatment group that took Norinyl with a daily multivitamin supplement or to the control group that took Norinyl with a placebo. Neither the investigator nor the study subjects knew which treatment was being taken. The study subjects were followed at 3-month intervals for one year. During the 3-, 6-, 9-, and 12-month follow-up visits, the investigator collected information about the frequency and severity of certain side effects.

(Basnayake et al., 1983)

Often, several treatments are available for the same condition. To prevent pregnancy, for example, several contraceptive methods are available (e.g., different doses of pills, different types of intrauterine devices (IUDs), barriers, sterilization). Decision makers in government health programs and donor agencies must sometimes choose among treatments; in general, they want to provide treatments that are safe, effective, and also relatively inexpensive. Phase III clinical trials provide decision makers with scientific evidence about the relative effectiveness and safety of competing treatments.

Design and Data Collection Methods

Randomized Clinical Trial Designs

Clinical trial designs have two basic forms: the parallel design and the successive treatment design. Again, adequate sample size, random assignment to treatment groups, and blinding are preferable design features.

Parallel designs. The structure of the parallel design is illustrated in Figure 8.1. In this design, study subjects remain with the treatment, to which they are randomly assigned, as long as they continue in the trial. Both groups are followed forward in time to observe the outcome of interest. The basic comparison groups are the patients receiving the new treatment and the patients receiving the control treatment. In Example 8.3, we present details of a parallel randomized clinical trial that compares expulsion rates for two different IUDs.

Example 8.3

Expulsion Rates of TCu2OO* and Progestasert Intrauterine Devices

Expulsion rates following postpartum insertion of the TCu2OO and the Progestasert intrauterine devices (IUDs) were compared in an RCT among 400 women.

Background: The Progestasert, a T-shaped IUD releasing crystalline progesterone, was designed to reduce bleeding and pain, improve effectiveness, and decrease expulsion. Expulsion rates are higher for all devices inserted immediately after delivery, but insertions performed at this time are easier and motivation may be high.

Research Hypothesis: Among postpartum women, expulsion rates for the Progestasert IUD are less than expulsion rates for the TCu2OO IUD.

Study Design: Parallel randomized clinical trial.

Eligibility Criteria: Postpartum women who delivered in Santiago, Chile, during November 1978 through February 1980.

Treated: Two experimental groups: 100 women receiving a Progestasert that was inserted by hand, and 100 women receiving a Progestasert that was inserted with an inserter. Insertions were performed within 10 minutes of the delivery of the placenta.

Not Treated: Two control groups: 100 women receiving a TCu200 that was inserted by hand, and 100 women receiving a TCu200 that was inserted with an inserter. Insertions were performed within 10 minutes of the delivery of the placenta.

*Throughout this manual, the use of trade names is for identification and does not imply endorsement by the U.S. Department of Health and Human Services, Family Health International, and the World Health Organization.

Example 8.3 (continued)

Outcome: Expulsion of IUD.

Follow-up: Study subjects were followed for up to 12 months to observe bleeding, pain, expulsion, and other relevant symptoms.

Data Collection Methods: Sociodemographic, medical, and follow-up data were recorded on standard forms.

Randomization Scheme: IUD and insertion techniques were randomly assigned, but the methods of assignment were not reported.

Blinding: The investigators did not provide information about blinding.

Data Analysis: Age, parity, education, and sociodemographic characteristics were compared to determine similarity of groups. At the end of 12 months, the investigators compared expulsion rates for the four groups.

(Lavin et al., 1983)

Successive treatment designs. In this type of design, each study subject is randomly assigned to a group that follows a predefined sequence of treatments where each person receives more than one treatment. The most common form of the design, the two-period successive treatment design, features one treatment that is followed by a second treatment. Frequently, a treatment-free period is introduced between treatments to allow any carryover effects from the first treatment to dissipate. The length of follow-up for each treatment should be the same so that events have the same time to manifest for either treatment. The successive treatment design has two variations: the replacement treatment design (Figure 8.4) and the crossover design (Figure 8.5).

Replacement treatment design

The replacement treatment design is used to collect data about the effect of changing from treatment A to one of a pair of alternative treatments—for example, treatment B or treatment C. In the replacement treatment design, the study subjects are equally divided into two groups. Both groups receive treatment A for the first period. After completing the first period, the study subjects in one group are switched or *changed over* to treatment B and those in the other group are switched to treatment C. Both groups are

Figure 8.4
Replacement Treatment Design

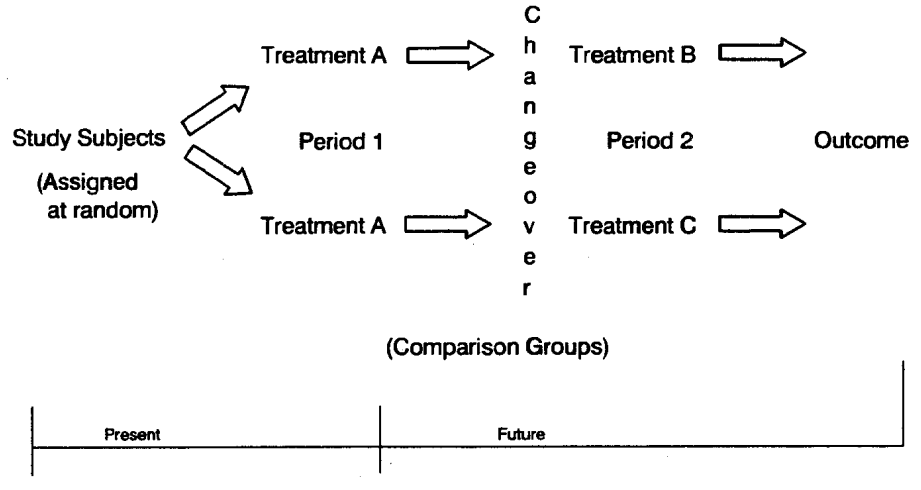
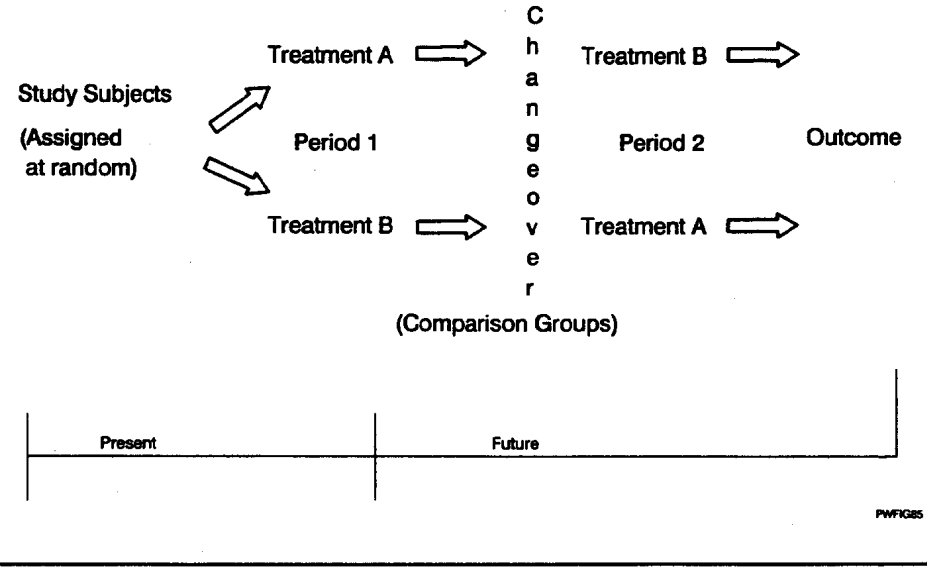


Figure 8.5
Crossover Design



followed forward in time to observe the outcome of interest. The outcomes among patients treated with A and B are compared with the outcomes among patients treated with A and C. Example 8.6 describes a replacement treatment clinical trial conducted in Indonesia.

Example 8.6

Switch from Standard Oral Contraceptives to Low-dose Oral Contraceptives in Indonesia

In many clinics, women who are taking a standard dose oral contraceptive (OC) regimen are later switched to a low-dose regimen, either to reduce side effects or because of changes in the availability of pills. A replacement treatment clinical trial in Indonesia was designed to assess changes in short-term side effects, including breakthrough bleeding, which accompanied the change from a standard dose to one of two low-dose combination pills.

(Badan Kerja Sama Penelitian Fertilitas et al., 1986)

Crossover design

In the crossover design, Group 1 receives treatment A in the first period and treatment B in the second period. Group 2 receives the treatments in the reverse order of Group 1 (Example 8.7).

Participants in clinical trials vary considerably in their initial state of health and in their reaction to treatment. One advantage of the crossover design is that it adjusts for person-to-person variation by having each individual serve as his or her own control. In effect, each study subject is used to gather information on two treatments. Because of this attribute, the crossover design frequently requires fewer study subjects than parallel designs.

Successive treatment designs are not appropriate or even possible in many situations because study subjects receive more than one treatment over both periods of the study. Some outcomes may preclude receiving more than one treatment. Surgical procedures are usually not suitable for study by a successive treatment design. Any clinical trial that has death as a primary outcome (e.g., cancer trials) cannot use the successive treatment design. Similarly, for contraceptive efficacy trials in

Example 8.7

Oral Contraceptives and Sickle Cell Disease

A common belief among physicians is that women with sickle cell disease should not use oral contraceptives (OCs), but there is no evidence to support this concern. On the contrary, progesterone (similar to one of the steroids contained in oral contraceptives) has been shown to have a beneficial effect on sickle cell disease by inhibiting the sickling of red blood cells, thereby reducing the frequency of painful crises and the number of irreversible sickled cells and significantly increasing total hemoglobin and red blood cell mass, count, and survival.

An 18-month study was undertaken to evaluate the effects of OCs on the progression of sickle cell disease. Using a crossover design, the investigators compared OCs and a placebo. The effects of OCs were assessed by monitoring blood parameters that would indicate sickling and red blood cell turnover and by recording the frequency of painful crises, infections associated with sickle cell disease, and other physical complaints.

Thirty women were recruited who had sickle cell disease, who attended sickle cell clinics, and who either had been sterilized, were using IUDs, or were using barrier methods of contraception. The medications—oral contraceptive or placebo—were prescribed by random assignment for each woman during the first 24 weeks of the trial. The women took the assigned medication for 24 weeks, discontinued the medication for 12 weeks, and then received the alternate medication for 24 weeks.

(Bonhomme, 1984)

which pregnancy is the primary outcome, the successive treatment design is not appropriate.

Clinical trials that are likely to have a substantial percentage of subjects who drop out are not suited for the successive treatment design. Participants who drop out during the first period are not available for evaluation in the second period. In general, as the treatment period lengthens, the probability that study participants will drop out increases. To minimize dropout, investigators should design successive treatment trials as short-term studies with period lengths of three months or less.

Study Population

The study population, a subset of the general population, consists of the individuals defined according to unambiguous study criteria. The group of actual study subjects in the trial may be a sample or a subset of the study population. For example, the general population in City A includes 3,000,000 women. The study population in City A is defined as 500,000 healthy women aged 18 to 44 years who reside in City A during a specified time period, and the RCT includes a sample of 500 women from the study population.

Investigators need to specify which individuals were studied and how they were selected. The scientific community needs to know the characteristics of the individuals who responded or failed to respond to the treatment, regardless of the success of the treatment. This information is essential for assessing whether the study results can be generalized to the study and general populations.

Inclusion and exclusion criteria

Eligibility Criteria. These criteria are the characteristics of the study population used to determine which individuals are eligible for inclusion in the clinical trial. Precise eligibility criteria and the reasons for their specification should be detailed in the planning stages of the trial. The impact that these criteria will have on the study design, on the recruitment of study subjects, and on the ability to generalize the findings from the trial must be considered. Eligibility criteria are most often described in two lists: inclusion criteria and exclusion criteria. In general, the number of exclusions should be minimized. In Examples 8.8 and 8.9, detailed inclusion and exclusion criteria for RCTs involving IUDs are provided.

Selection of Study Subjects

Defining the comparison groups

Defining the comparison groups is a critical step in designing the RCT. Treatment definition should focus on the most important question that the investigators think they can answer. The control treatment may be no treatment (a placebo) or the current standard treatment. Some situations offer no suitable

Example 8.8

Pelvic Inflammatory Disease and Intrauterine Devices With and Without Strings

Background: An RCT was designed to evaluate the rate of pelvic inflammatory disease (PID) over a 12-month period among women who received intrauterine devices (IUDs) with strings (control group) and among women who received IUDs without strings (treatment group).

Eligibility criteria

Inclusion criteria: Women in good physical health who want IUDs and can be followed for at least 12 months. Each woman must:

- Be a first-time IUD user.
- Be sexually active.
- Have terminated last pregnancy at least 42 days before IUD insertion.
- Have had at least one normal menstrual period since termination of last pregnancy.

Exclusion criteria: Before the IUDs were inserted, women provided a medical history and submitted to a physical examination that included a detailed pelvic examination, cultures for gonococcus from the cervix and anal crypts, and cultures for chlamydia from the cervix. Women with any of the following conditions were excluded from the study:

- Uterine abnormalities, abnormalities that distort the uterine cavity or cervical canal, evidence of adnexal mass or pelvic adhesive disease, suspicious pelvic mass.
- Abnormal uterine bleeding, menstrual irregularities, history of hyperplastic endometrial disease, polyps, or carcinoma.
- Postpartum endometritis or infected abortion, septic course after incomplete abortion or delivery.
- Using antibiotics or having used antibiotics within three months.
- Pelvic inflammatory disease:
  - fever at time of IUD insertion ($\geq 38^\circ$ C)
  - pelvic pain: excessive tenderness on pelvic palpation.
  - purulent vaginal, urethral, or Skene's glands discharge, detected either visually or by urethral massage; enlarged Bartholin's glands (PID did not include mild vaginitis or cervicitis).

(Wheeler et al., 1985)

Example 8.9

Comparison of Two Copper Intrauterine Devices

Background: Long-term intrauterine device (IUD) use has the advantages of convenience and low cost. One of the newer IUDs, the TCu380Ag, may be as effective or more effective than other popular copper IUDs, such as the Multiload Cu375.

Research Hypothesis: The TCu380Ag performs as well or better than the Multiload Cu375.

Study Design: Parallel randomized clinical trial.

Eligibility Criteria: Healthy, sexually active women who gave informed consent were eligible to participate in the trial. Their last pregnancy terminated at least 40 days before the IUD insertion.

(Cole et al., 1984)

comparison group. For example, in contraceptive trials, the use of placebo controls is not ethical (Example 8.10).

Comparison of only two groups in any trial is the most efficient design and gives the maximum chance for being able to make some definitive conclusions for a limited number of study subjects (Peto et al., 1976). Pressures will always be exerted from various sources to study more than two groups. Efforts to include more than two groups should be resisted. Sometimes, investigators desire to evaluate the equality of more than one new treatment, perhaps because of lower cost, lower toxicity, fewer side effects, and other factors. Even in these instances, studying more than two groups tends to dilute study power and produce a less informative result. With two groups, a positive result is more likely, and a null result is more informative (Example 8.11). Treatments must be sufficiently different from each other so that it is biologically plausible that the outcome related to one treatment could be substantially different from the outcome related to others. Ideally, only treatments likely to differ substantially should be compared in an RCT (Peto et al., 1976).

Exclusions before randomization. Individuals who are excluded from the study before randomization do not affect the treatment

Example 8.10

Early Oral Contraceptive Research
No Suitable Comparison Group

On May 11, 1960, the U.S. Food and Drug Administration (FDA) approved the marketing of oral contraceptives (OCs). Preclinical studies had begun nine years earlier. Between 1954 and 1956, clinicians assessed the pill's safety and the dosage necessary to prevent pregnancy. In 1956, approximately 1,000 women were enrolled to use the pill for at least one year in a large efficacy study. Although the trial had several objectives, the most important objective was to demonstrate that the pill would be very effective in preventing pregnancy. A convincing demonstration would help gain approval for marketing from the FDA. The demonstration was successful and approval was granted even though the study of OCs did not include a control group. There was no standard hormonal contraceptive against which to compare the effectiveness of the pill. Once the original pill was marketed, however, a standard existed to serve as a control for future experimental studies of OCs.

(McLaughlin, 1982)

Example 8.11

Randomized Clinical Trial Comparing Penicillin G and Kanamycin in the Treatment of Endometritis

A variety of antibiotics can be used to treat endometritis (infection of the lining of the uterus) after childbirth. Consider a hypothetical study of the comparative effectiveness of several drugs that are used alone to treat endometritis: penicillin G, ampicillin, tetracycline, and kanamycin. Although it is tempting to study all four drugs in a single RCT, a definitive conclusion is more likely when comparing only two drugs at a time. Kanamycin is the standard treatment, but judging from antimicrobial coverage and cost factors, the researchers chose penicillin G as the most promising alternative. Hence, the preferred RCT design would compare penicillin G with kanamycin.

comparison and do not introduce bias. Subjects might have to be excluded because they have a condition for which the treatment assigned in the trial may be contraindicated, because they are not likely to remain *observable* or because they are taking other medication.

However, exclusions made before randomization may affect the ability to generalize the results. As the number of excluded study subjects increases, the less generalizable the results become. Although the trial may be perfectly executed, it may have no relevance to the persons to whom the researchers wish to relate the findings.

The investigators must make sure that the inclusion and exclusion criteria are clearly specified and that they are applied before randomization. If they are applied before randomization, bias is not introduced. Criteria must be evaluated for sources of selection that would make the study sample atypical or nonrepresentative. Selection criteria that are too restrictive result in the selection of study subjects who are such a restricted subset of the study population that the results will have little meaning. Example 8.12 illustrates problems with restrictive eligibility criteria.

Example 8.12

Intrauterine Devices and Menstrual Blood Loss
Restrictive Eligibility Criteria

An RCT was designed to evaluate the menstrual blood loss associated with two different intrauterine devices (IUDs). Since the investigator knew that menstrual blood loss increased substantially in users of most inert, nonmedicated IUDs, he established rigorous exclusion criteria based on hemoglobin levels. Women with even mild anemia were excluded from participation. Consequently, most women in his region were ineligible for enrollment in the trial. These strict criteria meant that it took a long time to recruit sufficient numbers of women for study and that only affluent, well-nourished women were enrolled. Thus, the generalizability of the findings of the study to the general population of women in the region was limited.

Randomization. Study subjects should be assigned to treatment groups at random. Because randomization should reduce selection

and confounding biases, it is the preferred method of assigning individuals to treatment and control groups. Randomization ensures that study subjects have an equal chance of being assigned to either group and enhances the likelihood of comparability between the groups on all factors, known and unknown, measured and unmeasured. Allocating patients to treatment groups by random chance rather than by the clinician's choice is essential to avoid selection bias. Without effective randomization, clinicians with a preference for or against a particular therapy may subconsciously or consciously exclude a subject or alter a study subject's treatment assignment. Failure to randomize will likely introduce differences between the treatment groups and invalidate the results.

Randomization ensures that subjects are equally likely to be assigned to the treatment or control group. However, not all randomization schemes are equivalent (Table 8.13). Certain schemes that investigators label *random* are not really random and are therefore ineffective for controlling bias. In particular, some commonly reported methods are based on alternate assignment of patients, on chart number (odd or even), or on date of birth. These methods may influence the assignment decision, since personnel who assign treatment or refer patients may know in advance which treatment is next and may therefore decide a patient's eligibility or schedule the patient's referral according to treatment preferences. These are not truly random allocation schemes. Schemes such as shuffling a deck of cards or flipping a coin are random, but these methods can tempt investigators to intervene. For example, they may think, "We have had too long a series of heads," and decide to change the next assignment. Also, these schemes cannot be checked or reproduced. Using a table of random numbers is a preferred method because the table is random and reproducible. Even if the method of selecting study subjects is random, the investigators and the study subjects must not be able to suspect which treatment is next. That is, the randomization must be blinded. See the section on blinding in this chapter for more detail.

Balanced randomization

Balanced randomization, sometimes called *pseudorandomization* or *restricted randomization*, can be used to ensure that the number of study subjects allocated to each treatment group is approximately equal during the entire trial. Not only is this statistically efficient, but it ensures that no time trend in the allocation rates will exist and that bias is minimized even if a time trend in the prognosis of the patients

Table 8.13

An Evaluation of Randomization Schemes

Scheme	Evaluation
None reported by authors	Poor
Alternate assignment (chart number, date-of-birth, etc.)	Poor
Coin flipping or card shuffling, without blinding of the randomization	Fair
Coin flipping or card shuffling, with blinding of the randomization (e.g., sealed opaque envelopes)	Good
Random number table, with blinding of the randomization	Excellent
Random number table, balanced randomization with blinding and disguised block length	Excellent

exists for any reason. Balanced randomization is usually achieved by first determining the number of study subjects per *block*; then individuals are randomly assigned within each block so that half are in the treatment group and half are in the control group.

Balanced randomization is the preferred method of randomization as long as the investigators do not allow the block lengths to become known. Predicting the next treatment is more difficult with long block lengths (block lengths of ten are better than four) and with randomly varying the block lengths.

By design, randomization improves the balance of potentially confounding variables. The equal distribution of the measured and unmeasured variables in the treatment groups minimizes confounding bias. If randomization does not produce balanced groups, statistical techniques exist that allow the investigator to adjust for factors that are measured. However, we cannot adjust for confounding by factors that have not been measured.

Finally, investigators need to completely describe in their research reports the randomization scheme used in their RCT. Failure to describe the randomization scheme should cause readers to question the methods and results.

Steps for determining a balanced randomization include using blocks, using a random number table to select blocks, and concealing the treatment assignment. The following eight steps describe a method that randomizes and allocates approximately the same number of patients to each treatment group during the trial (Adapted from Peto et al., 1976).

Step 1. List all possible randomization blocks for a specified block size and the specified number of treatments (Example 8.14 and Table 8.15).

Example 8.14

Randomization Blocks

For the present example, *A* represents treatment assignment A, and *B* represents treatment assignment B. Blocks are set up to make sure that the randomization balances with every sixth patient; the block size equals 6. All combinations of 3 As and 3 Bs are listed. After every 6th patient, the number of patients assigned to treatment A will approximately equal the number of patients assigned to treatment B.

All possible permutations or treatment sequences have to be listed. To determine how many different sequences are possible, use the following formula:

(8.14.1)

$$\text{Number of possible sequences} = \frac{(\text{Block Size})!}{(\text{Block Size}/2)!*(\text{Block Size}/2)!}$$

For the present example, a block size of six generates 20 sequences, which are listed in Table 8.15.

$$\text{Number of possible sequences} = \frac{6!}{3!*3!} = \frac{6*5*4*3*2*1}{(3*2*1)*(3*2*1)} = 20$$

Table 8.15

Treatment Sequences

AAABBB	ABABAB	BAAABB	BABBAA
AABABB	ABABBA	BAABAB	BBAAAB
AABBAB	ABBAAB	BAABBA	BBAABA
AABBBA	ABBABA	BABAAB	BBABAA
ABAABB	ABBBAA	BABABA	BBBAAA

(Adapted from Peto et al., 1976)

Step 2. Assign sampling numbers to each block of treatment sequences. Every block must have an equal number of sampling numbers so that each block has an equal probability of selection. For this example, five sampling numbers are assigned to each block (Table 8.16).

Table 8.16

Sampling Number Assignment

Sampling Number	Treatment Sequence	Sampling Number	Treatment Sequence	Sampling Number	Treatment Sequence	Sampling Number	Treatment Sequence
00-04	AAABBB	25-29	ABABAB	50-54	BAAABB	75-79	BABBAA
05-09	AABABB	30-34	ABABBA	55-59	BAABAB	80-84	BBAAAB
10-14	AABBAB	35-39	ABBAAB	60-64	BAABBA	85-89	BBAABA
15-19	AABBBA	40-44	ABBABA	65-69	BABAAB	90-94	BBABAA
20-24	ABAABB	45-49	ABBBAA	70-74	BABABA	95-99	BBBAAA

(Adapted from Peto et al., 1976)

Step 3. Before you turn to a random number table, you must decide, once you have obtained a starting number, whether you will proceed up, down, left, or right in the chosen row or column in the random number table. For the present example, we decided to proceed down the column.

Step 4. Shut your eyes and, at random, make a mark on the random number table with a pencil dot. Select the two-digit number nearest to the mark. For the present example, the number 33 was selected (Table 8.17).

Table 8.17

Random Number Table

41	82	79	37	00	45	98	54	52	89	26	34
66	18	76	82	11	18	61	90	90	63	78	57
42	34	00	49	97	53	33*	16	26	91	37	58
90	84	22	16	26	96	54	11	01	96	58	91
33	43	01	32	58	39	19	54	56	37	23	38
17	89	37	04	18	32	13	45	59	03	91	08
36	87	98	73	77	64	75	19	05	61	11	64
15	58	19	68	95	47	25	69	11	90	26	19
45	52	27	35	86	81	16	20	37	60	39	35
72	72	81	84	36	58	05	10	70	50	31	04
06	68	52	50	39	35	92	28	18	89	64	37
95	73	80	92	26	49	54	30	41	21	06	62
96	23	16	46	15	51	60	31	55	27	84	14
34	96	32	68	48	22	40	17	43	25	33	31
07	19	94	46	17	51	03	73	99	89	28	44
37	08	08	46	56	76	29	48	33	87	70	79
18	01	67	19	29	49	58	67	68	56	27	24
08	79	18	78	00	32	86	74	78	55	55	72

(Adapted from Peto et al., 1976)

Step 5. From the two-digit starting point, list the series of two-digit random numbers from that point down (up, right, left) the column (row). The number of random numbers to list is equal to the total sample size desired divided by the block size.

For the present example, the required sample size is 60 participants, or 30 in each treatment group. Therefore, the required number of two-digit random numbers needed to establish a treatment sequence for 60 participants is the number of participants (60) divided by block size (6). Thus, starting with the number 33 (Step 4) list the ten random numbers down the column in Table 8.17: 33 54 19 13 75 25 16 05 92 54.

Step 6. For each random number (Step 5), find the corresponding block or treatment sequence (Table 8.16). For our example, the treatment sequences are:

(33)	(54)	(19)	(13)	(75)
ABABBA	BAAABB	AABBBA	AABBAB	BABBAA
(25)	(16)	(05)	(92)	(54)
ABABAB	AABBBA	AABABB	BBABAA	BAAABB

Step 7. Consecutively number opaque envelopes for the number of participants in the trial. Then, in sequence, insert the letter (A or B: the treatment group assignment corresponding to the letter) into the envelopes.

For the present example, 60 envelopes were numbered from 1 to 60. Letters corresponding to the indicated treatments were inserted into the numbered envelopes to match the treatment sequence (Table 8.18).

Table 8.18

Treatment Assignment

Envelope Number	Treatment Assigned	Envelope Number	Treatment Assigned	Envelope Number	Treatment Assigned
1	A	9	A	17	B
2	B	10	A	18	A
3	A	11	B	19	A
4	B	12	B	...	
5	B	13	A	...	
6	A	14	A	...	
7	B	15	B	59	B
8	A	16	B	60	B

Step 8. As each participant is entered into the trial, determine treatment assignment by opening the next envelope in sequence.

Blinding. In any clinical trial, bias is a major concern and blinding is one method for decreasing bias. Bias can occur at

Blinding decreases bias

several points in time during a clinical trial, from the design stage through data analysis and interpretation. The general solution for managing bias in an RCT is to keep study subjects and the investigator blinded to the exact treatment assignment. Blinding can be applied to the randomization process, the administration of the treatment, and to the assessment, classification and evaluation of the outcome.

The randomization process must be blinded

The randomization process must be blinded. The investigator, the evaluator, and study subjects must not be able to suspect which treatment is next. Several processes may be used to blind the investigator, clinicians, and study subjects about the randomization of treatment assignments:

- Inserting assignments in consecutively numbered, opaque envelopes.
- Making assignments by telephone from a central office.
- Having drugs prepackaged and numbered for consecutive patients according to the randomization scheme.

The use of consecutively numbered, opaque envelopes is one of the most practical methods. Persons who implement the randomization process must not be involved in the enrollment of study subjects, treatment, or evaluation.

Types of blinding

Single blinding occurs when the treatments have been concealed from trial participants. Double blinding occurs when the treatments have been concealed from both the investigator and the study subjects. Triple blinding occurs when the clinician who is evaluating outcome (the evaluator) is not the investigator and when the study subjects, the investigator, and the evaluator are unaware of the treatment assignments. Because both investigators and study subjects are likely to have strong hopes and prejudices about a trial, biases are likely to occur unless precautions are taken.

Subjects and investigators should be blinded to treatment assignment

Blinding study subjects about the treatment assignment is desirable. Blinding study subjects is necessary if dropout or noncompliance is likely to result from patients' knowing their therapy and becoming discouraged.

Blinding investigators about the treatment assignment is also important. Knowing that a patient is receiving what is *suspected* to be the less effective treatment may lead to compensatory efforts that would be detrimental to the study. For example, the investigator may

institute some additional adjunctive therapy not specified in the protocol, or an assistant may decide or be given orders to monitor study subjects receiving certain treatments more closely than required in the protocol. The most serious problem involves the investigator who is inclined to make sure that the study subject who is more severely ill receives the treatment *suspected* to be the most effective.

Subject and evaluator should be blinded to the outcome

Blinding study subjects and the evaluator about the outcome is also critical, especially if the outcome is subjective. In some situations, blinding about the outcome is not possible (e.g., if the outcome is death). A study subject who has knowledge of the expected outcomes may be more likely than an uninformed subject to report data that support the desired outcome. Evaluators who have prejudices about the conclusions of the trial may be inclined to diagnose borderline symptoms and situations as the outcome that supports their expectations about the proposed new treatment.

Of course, in some trials not all types of blinding are possible. For example, if an operative treatment is being compared to a drug treatment, it is impossible to blind the surgeon and patient about the treatment. In this case, using a blinded evaluator may be the most appropriate approach. Examples 8.19 and 8.20 illustrate blinding the study subject and blinding the clinician who evaluates the outcome.

Example 8.19

Estrogen Injection and Postpartum Breast Engorgement

A hypothetical study is designed to compare the efficacy of an injection of estrogen immediately after delivery with no therapy in preventing the pain of postpartum breast engorgement among women who choose not to nurse. The outcome measure is the amount of pain reported by the women each day after delivery. For a given level of discomfort, women who receive the injection may report less pain because they know they have been given a treatment they believe to be effective. This bias (called ascertainment or information bias) may be avoided by using a placebo injection instead of no therapy for the control group.

Ascertainment or information bias

Example 8.20

Daikon Shield Versus Lippes Loop and Pelvic Inflammatory Disease

A hypothetical study compares rates of pelvic inflammatory disease (PID) among women who use the Dalkon Shield and women who use the Lippes Loop. The outcome measure is the physician's diagnosis of PID. Since the type of intrauterine device (IUD) can be determined by viewing the tail protruding from the cervix, the physician evaluating a patient with symptoms would know the patient's treatment group. If the physician had read or heard that the Dalkon Shield was associated with more infections than other devices, this information (or clinical hunch) might influence the diagnosis. Given the same symptoms, a patient known to have a Dalkon Shield in place might be more likely to be diagnosed with PID than a patient with a Lippes Loop. The physician could be blinded by using the same type of tail for both IUDs and by having different persons do the insertions and the follow-up examinations.

Sample Size

The need to determine the requisite sample size is critical in the design of an RCT. (Sample size and power estimation are discussed in Chapter 5.) This is especially true because RCTs are held in such high esteem. Yet many researchers do not design sufficient statistical power into their studies (i.e., their sample sizes are too small to answer, with assurance, the questions posed). A review of 71 *negative* randomized clinical trials revealed that 67 had a greater than 10% chance of missing a true 25% improvement in therapy. Many treatments were labeled as no different from the standard or control treatments in trials that had inadequate samples (Frieman et al., 1978).

Ethical Concerns

Unlike other studies, RCTs allow the investigator to assign study subjects to the comparison and treatment groups. This scientific *meddling* in the course of individual lives forces the investigator to confront important issues concerning the ethics of *human experimentation.*

The term human experimentation recalls for many persons sad chapters in human history when experiments were carried out on unwilling human subjects (e.g., prisoners of war or men and women institutionalized for diagnoses of mental illness). On the other hand, the development of modern medicine has been due, in large part, to human experimentation. Since the first treatment was given to the first patient, practitioners of medicine have been experimenting empirically on patients in uncontrolled settings and have been changing treatments as dictated by the response of the patient. The development of RCTs has been a major advancement for medicine because when properly conducted, RCTs provide valid, useful, and reproducible results. Where disagreement exists concerning choice of treatments, a well-conducted and well-reported RCT offers hope of determining the preferred treatment. In this sense, it may be unethical not to perform an RCT to answer an important medical question (Schafer, 1982).

Conflict of Obligations

RCTs may present a conflict of obligations to the clinician or investigator. Most western physicians are bound by an oath (often a modern version of the Hippocratic Oath) to make the welfare of the individual patient of paramount importance. An RCT may appear to be in direct conflict with this obligation. Although the patient's welfare is the principal responsibility of the clinician, the clinician-investigator has the long-term perspective—the pursuit of knowledge may help other patients in the future.

Conflict exists between the obligation to the individual and the obligation to the groups. In an RCT, individual treatment, considered the patient's right in western medicine, may have to be compromised to adhere to random allocation and other RCT rules. This has been interpreted by some to be a violation of the clinician's obligation to the individual patient. The question can be stated as "When, if ever, is it morally justifiable to sacrifice the patient's right to completely individualized treatment for the benefit of scientific progress?" (Schafer, 1982). When contemplating this question, one must consider two types of RCTs:

- One of the treatments is the accepted *best therapy available today*, and the other is a new therapy that shows promise to be even more effective.
- One treatment is a placebo and the other is a new therapy that shows promise for treating a disease for which no effective therapy existed in the past.

There are consequences to not performing RCTs. Studies performed without adequate controls may seriously mislead clinicians regarding choices of therapy. Clinicians may feel obligated to provide treatments that are, in fact, inferior. Thus, the investigator should carefully consider the potential risks and benefits of a given RCT.

Informed Consent

Informed consent is required for prospective study subjects in an RCT. Study subjects must be informed about all aspects of the trial that might influence their participation. They need to be aware of the known risks and benefits of the proposed treatments. They have a right to choose not to participate, or if enrolled, they have a right to withdraw at any time without penalty. Study subjects being recruited into an RCT clearly have a right to know that they are taking part in an experiment and that their treatment is decided by chance. It is unethical not to advise prospective study subjects of the method by which their treatment will be chosen. Some investigators fear that this disclosure will discourage patients from cooperating. Patients seek out clinicians who will base their therapy on training and clinical experience, not on the *luck of the draw*. Some investigators feel that adequate numbers of patients cannot be recruited if this disclosure is made. However, since the outcome from the proposed therapy is admittedly unknown (excluding placebo studies), no material information is withheld.

When investigators plan an RCT, they likely have doubts about the effectiveness of a treatment but may have a treatment preference (a hunch about which treatment is better, based on clinical experience). When investigators have a treatment preference, it is difficult for them to be truly indifferent to the alternatives being tested. Can they ethically withhold their opinions from patients considering entering the study? On the other hand, clinical impressions in the absence of scientific controls are highly questionable. What are the ethical

consequences of providing well-intended, but possibly incorrect, advice to a patient considering taking part in an RCT?

In summary, the ethical issues need careful consideration before implementing an RCT. Many needed RCTs have not been done because of ethical considerations based on *suggestive evidence.* This suggestive evidence frequently consists of clinical impressions, small inconclusive studies, or merely bias. In other words, ethical considerations can provide an easy justification for not doing an RCT—an action that can in itself be unethical.

Data Analysis Methods

The RCT study design determines the types of analyses that may be performed. In clinical trials, treatment data are collected for groups of study subjects who are comparable except for the treatment assigned. The eligibility criteria are used to select study subjects based on a specific select set of characteristics. For other characteristics, investigators rely on randomization to help ensure the comparability of study subjects. However, even when eligibility criteria are used and randomization is properly implemented, there is no guarantee that comparison groups will be homogeneous in characteristics other than the treatments.

Baseline and Follow-up Data

Data from clinical trials can be divided into two types: information collected from study subjects when they are admitted to the study (baseline data), and information collected from study subjects during follow-up (follow-up data). The baseline data provide researchers with health or demographic profiles of study subjects and with a description of the study group(s). With follow-up data, researchers are able to evaluate each participant's response to his or her assigned treatment and to determine the overall response of each treatment group. Together, baseline and follow-up data provide information on changes in a study subject's health that could be attributed, for example, to a treatment or, alternatively, to a consequence of a baseline characteristic.

Baseline Data Analysis

Baseline data are collected and analyzed to assess comparability between the treatment groups. The distributions of the baseline characteristics in the study groups should be compared. When a difference in outcomes between study groups is observed, this difference may have one of two causes: one treatment may be superior to the alternative treatment, or the group that exhibited the superior performance may be composed of subjects who would have done better regardless of the treatment they had been assigned.

Analysis table of characteristics of study subjects. Comparison of baseline characteristics between treated and untreated study subjects may be visual. Statistical testing to determine if the distributions are different is not essential. Even the successful use of randomization is compatible with some differences in variable distributions; and if distributions of important (prognostic) baseline variables differ, regardless of whether or not the differences are statistically significant, the differences may be related to the outcome.

Table 8.21 presents arithmetic means for patient characteristics according to treatment groups (doxycycline and placebo) and women lost to follow-up in an RCT designed to evaluate the effectiveness of 200 mg of doxycycline given orally at the time of IUD insertion in reducing the incidence of PID. Treatment groups and women lost to follow-up had similar ages, education, number of live births and spontaneous abortions, and coital frequency per week (Sinei et al., 1990). Statistically significant but practically unimportant differences existed for education and live births.

Follow-up Data Analysis

In most Phase III clinical trials of contraceptives, the primary outcome variable provides incidence data: the number of pregnancies, the number of discontinuations of specific contraceptives, the number of IUD expulsions, etc. Outcome frequencies, the number of individuals receiving the study treatments, and how long the individuals participated in the study are used to evaluate the treatments. Some contraceptive clinical trials have a relatively short period of follow-up—days or a few weeks.

Table 8.21

Means of Patient Characteristics by Treatment Group

Characteristic	Prophylactic doxycycline (n=827)	Placebo (n=828)	Lost to follow-up (n=158)	p*
Age	25.5	25.3	25.6	0.51
Education	9.5	9.7	9.2	0.02
Live births	2.5	2.3	2.4	0.03
Spontaneous abortions	0.2	0.2	0.2	0.78
Coital frequency/week	2.4	2.4	2.5	0.48

*One-way analysis of variance on the three groups.

(Sinei et al., 1990)

Short period of follow-up. Contraceptive clinical trials, which are generally designed to evaluate long-term contraceptive use, may produce results within a few days or weeks after study subjects have received their treatment. Clinical trials of sterilization techniques, for example, represent an important class of contraceptive studies, in which evaluations after a short period of follow-up are important. These studies typically focus on complications associated with various sterilization techniques or their use by different types of family planning professionals. In other clinical trials, secondary outcome variables may produce results within a few days or immediately. For instance, in an IUD study focusing on expulsion as the primary variable, investigators may be interested secondarily in the percentage of women reporting pain at IUD insertion.

Lengthy period of follow-up. Most contraceptive clinical trials have an extended period of follow-up (months or even years). In

these studies, how the events were distributed throughout the follow-up period is important. For clinical trials with a lengthy follow-up (three months or more), the investigators are interested not only in whether the event of interest occurs but also in how long it takes for the event to occur. The analysis of the time until the occurrence of the outcome of interest is called life-table analysis. The methods for conducting a life-table analysis are described in detail in Appendix 3.

Life-table analysis

Analysis table for proportions and relative risk. The symbol p represents a proportion calculated from the study data and the symbol n represents the sample size used in the calculation of the proportion. p is number of study subjects who exhibited the primary outcome event divided by the total number of study subjects. The proportions (or percentages) of the primary outcome event for the different treatment groups and the relative risks are calculated as shown in Table 8.22.

In Example 8.23, the proportion of women with postoperative morbidity who were sterilized by a physician is 9/149 = 0.06. For women sterilized by a nurse-midwife, the proportion with post-operative morbidity is 10/143 = 0.07. Thus, 7% of the women who had had sterilization surgery by a nurse-midwife had a postoperative complication, whereas 6% of the women who had had sterilization surgery by a physician had a complication.

Relative risk

In RCTs, the relative risk may be used to measure the magnitude of the association between the treatment and outcome under study. Relative risk is the incidence of the outcome among the study subjects who received the specified treatment (8.22.1) divided by the incidence of the outcome among the study subjects who received no treatment (placebo) or who received the standard treatment (8.22.2). This relative risk, also known as cumulative incidence relative risk (CIR), measures the risk of the study subjects developing the outcome during the entire study period (8.22.3). In Example 8.23, the CIR = 0.07/0.06 = 1.2. Because the confidence interval includes 1.0, we would conclude that women who had a tubal sterilization performed by a nurse-midwife had the same risk for postoperative morbidity as women who had a tubal sterilization performed by a physician.

Cumulative incidence relative risk (CIR)

Person-time

Incidence density relative risk (IDR)

Sometimes, person-time of observation is used as the denominator in the relative risk calculation instead of the number of people enrolled in the study. This type of relative risk is known as incidence density relative risk (IDR). The person-time denominator

Table 8.22

Analysis Table for Cumulative Incidence Relative Risk

Outcome	Treatment Yes	Treatment No
Present	a	b
Absent	c	d
Total	n_1	n_0

(8.22.1)

Proportion with outcome in treatment group = Number of study subjects with outcome in treatment group / Total study subjects in treatment group

$$p_1 = \frac{a}{n_1}$$

(8.22.2)

Proportion with outcome not treated = Number of study subjects with outcome in no treatment group / Total study subjects not treated

$$p_0 = \frac{b}{n_0}$$

(8.22.3)

Cumulative Incidence Relative Risk (CIR) = Proportion of outcome in treatment group / Proportion of outcome in no treatment group

$$= \frac{a/n_1}{b/n_0}$$

(Rothman and Boice, 1979)

Example 8.23

Tubal Sterilization by Nurse-Midwives and by Physicians

In a study of complications occurring with postpartum minilaparotomous tubal sterilization procedures, those performed by nurse-midwives were compared with procedures performed by physicians. The event of interest was a postoperative problem including mild pyrexia, respiratory infection, cystitis, or wound breakdown. Volunteers who had requested sterilization before delivery were randomly allocated to a nurse-midwife or a physician for surgery. The evaluation of the outcome was done five days postoperatively.

	Exposure				
Outcome	Sterilization procedure by nurse-midwife		Sterilization procedure by physician		
Women with postoperative morbidity	10	(a)	9	(b)	
Women without postoperative morbidity	133	(c)	140	(d)	
Total	143	(n_1)	149	(n_0)	192(t)
Proportion with postoperative morbidity	10/143 = 0.07		9/149 = 0.06		

CIR = 0.07/0.06 = 1.2 (95% CI: 0.5 - 2.8)

(Dusitin et al., 1980)

simultaneously considers the number of persons under observation and the duration of observation for each person; that is, the actual time of observation must be computed for each individual in the group (Table 8.24). For example, if 10 persons participate in a study for 10 years, they are said to have contributed 100 (10 persons * 10 years) person-years of observation. The same figure may be derived if 100 persons were under observation for one year or 200 persons for

Table 8.24

Analysis Table for Incidence Density Relative Risk

	Treatment	No Treatment	Total
Cases	a	b	m_1
Person-time	n_1	n_0	t

(8.24.1)

$$\text{Incidence Density Relative Risk (IDR)} = \frac{a/n_1}{b/n_0}$$

where
a is the number of study subjects with the outcome in the treatment group, *b* is the number of study subjects with the outcome in the group that received no treatment, n_1 is the person-time in the treatment group, and n_0 is the person-time in the group that received no treatment.

(Rothman and Boice, 1979)

six months. This method allows the investigator to more satisfactorily manage the situations in which the dates when the study subjects enter the trial vary, or in which some study subjects are no longer under observation during the course of the study because of death, loss of contact, or other reasons.

Exclusions after randomization. As a general rule, all patients who are randomized should be analyzed. Furthermore, the patients should be analyzed as part of the treatment group to which they were initially assigned. Every effort must be made to determine the outcome for all study subjects randomized. Exclusions after randomization (e.g., withdrawals, losses, and deviations) can bias the randomized treatment comparison and thus should be carefully scrutinized. Inappropriate approaches to the handling and analysis of these patients can lead to subtle, well disguised, yet serious errors in research design and results.

Ineligibility

Investigators have given ineligibility as a stated reason for exclusion after randomization when they have eventually discovered that the study subject did not meet the inclusion criteria. Decisions to withdraw study subjects under these circumstances, however, will likely bias the results. For example,

clinicians who have a treatment preference for a specific patient may withdraw the patient if randomized to the *wrong* group. As another example, the patient whose health continues to decline on the assigned treatment would likely attract more attention and therefore be more likely to be categorized as ineligible. The valid approach is not to allow any exclusions after randomization for the eventual discovery of patient ineligibility.

One exception is when differential diagnosis of eligibility criteria is difficult. The investigators might develop a process so that 1) the same information is collected from each patient at the time of randomization, 2) the information is reviewed at a central location, and 3) the clinician is blinded to the treatment allocated to the patient. Patients not satisfying the eligibility criteria could then be withdrawn.

Postrandom -ization pretreatment outcome

Postrandomization pretreatment outcome pertains to any outcome (such as death) that occurs after randomization but before the treatment begins, before the treatment regimen is completed, or before the treatment theoretically could produce an effect. Decisions to withdraw study subjects who were diagnosed with the outcome after randomization but before treatment also introduce bias. For example, in an RCT of the effect of a drug on death, the investigator decided to withdraw as *not analyzable* all patients who died after randomization but before treatment began and also all patients who died without receiving at least seven days of the treatment. This rule is called the *7-day rule* since the drug would not have any effect, theoretically, for at least seven days. This seems intuitively attractive, since none of the deaths could be attributable to the treatment. In essence, however, the same argument could be used to exclude all the deaths during the entire study interval in the placebo group, since none of them, theoretically, would be related to treatment.

When such rules are instituted as part of the protocol before the trial has begun, only the impact of the randomization may be reduced. If rules are instituted after the trial has begun, the actions may lead to biased and invalid results. Therefore, all randomized patients should be analyzed. This is especially true since it is often difficult to determine whether the rules were instituted before or after the start of the trial. Planned or unplanned, the exclusion of *not analyzable* outcomes is not acceptable in the analysis of randomized clinical trials (Meier, 1981).

Analyze all randomized patients

When study subjects are followed up, sometimes they are lost to

Conduct thorough followup

further involvement in the trial for various reasons, such as moving away or disinterest. The study must retain as many study subjects as possible. Bias is likely to occur if the two treatments differ in unpleasantness, toxicity, efficacy, or any way that affects loss differently. There are no acceptable reasons for losing study subjects.

Although losses may not be completely eliminated, methods to minimize them must be developed. For example, study subjects who are likely to be lost should be excluded before randomization. When study subjects do not return, extensive procedures can be implemented to locate them by telephone inquiries, letters, or special visits by research assistants. Analysis should consider differential rates of loss between the treatment groups. If losses do occur, analyses should include their outcomes up to the time of loss.

Do not omit subjects who deviate from the protocol

Some investigators suggest that if a study subject deviates from the protocol or the assigned treatment, that person should not be included in that treatment group in the final analysis (or should be included only up to the point of deviation). Again, this approach is intuitively attractive. However, omitting study subjects who deviate from the protocol is a serious error, since the group that deviates from one protocol and the group that deviates from the other protocol may be so different that the treatment comparison in the remaining study subjects may be seriously biased (Peto et al., 1976).

All the study subjects who deviate from the protocol should be followed and analyzed in the group to which they were originally assigned. In Example 8.25, the study subjects who deviated from the antibiotic group, therefore, must be categorized in the antibiotic group for the analysis. The correct procedure is to compare the group that was randomly assigned to antibiotics with the group that was randomly assigned to placebo, no matter what happened to them after assignment. Then the results address the issue of whether a policy of antibiotics for PID is superior to a policy of no antibiotics (Figure 8.26).

It might seem difficult to deal with exclusions, withdrawals, losses, and protocol deviants because of the many options that seem logical, ethical, and medically sound. However, with rare exceptions, all subjects randomized should be analyzed, and

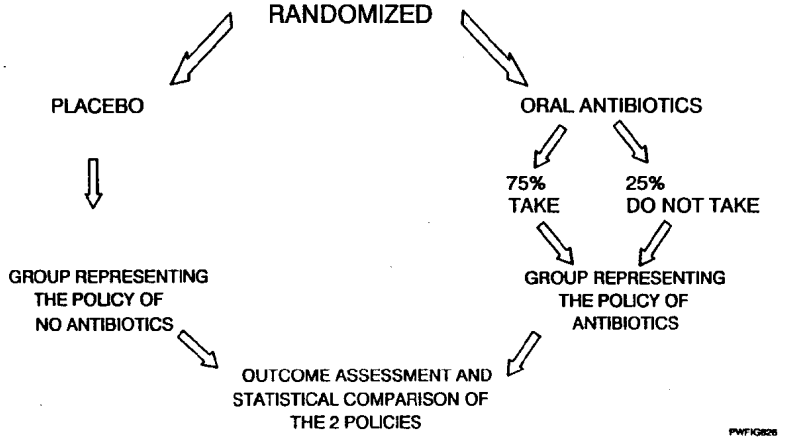
Example 8.25

Oral Antibiotics and Pelvic Inflammatory Disease Febrile Morbidity

To determine whether the incidence of febrile morbidity from pelvic inflammatory disease (PID) is decreased by oral antibiotic treatment, patients with PID are randomly assigned to either the antibiotic group or the placebo group. However, 25% of the patients in the antibiotic group deviated from the protocol and did not take their medication. In effect, the patients in the antibiotic control group who deviated from the protocol received the same treatment (i.e., nothing as the placebo group). Should the patients who deviated from the protocol then be excluded totally or should they be merged with the placebo group and compared only to the antibiotic patients who adhered to the protocol? The response for both options is emphatically negative because the members of the two groups would no longer be comparable. The patients who did not take the antibiotics were likely to be in better health generally or were likely to have felt better because of less serious PID. Exclusion of the patients who deviated from the protocol potentially would leave only the more serious cases in the antibiotic group and produce a bias in the treatment comparison. If these patients were included in the placebo group, then not only do the more serious cases remain in the antibiotic group, but the placebo group has been infiltrated with potentially less serious cases. Again, the treatment comparison is biased. The appropriate paths for treatment assignment and analysis for patients who have been randomized is illustrated in Figure 8.26.

Figure 8.26

Flow Diagram of Patients Randomized in a Randomized Clinical Trial of Pelvic Inflammatory Disease and Oral Antibiotics



analyzed in the group to which they were randomized. Failure to follow this rule may introduce bias (Examples 8.27 and 8.28).

The exclusions described in Example 8.27 biased the results. The study reported a 32% reduction in death from cardiac causes (Anturane Reinfarction Trial Group, 1978). When *ineligible* and *unanalyzable* subjects were included, the effect was adjusted to a 21% reduction (Temple and Pledger, 1980). Subsequently, the Food and Drug Administration (FDA) advisory committee announced that sulfinpyrazone could not be labeled and advertised for the prevention of death in the critical months after a heart attack. The detailed review of the study indicated that the original exclusions were inappropriate (Temple and Pledger, 1980).

In Example 8.28, the authors concluded that *one can justify almost any conclusion, depending on the analysis chosen* (Coronary

Example 8.27

Anturane and Myocardial Infarction

Anturane (sulfinpyrazone) is a drug used to treat gouty arthritis; it also appears to have antithrombotic (anticlotting) effects. To test the ability of this drug to prevent repeat myocardial infarction (heart attack) in persons having had a recent myocardial infarction, a randomized clinical trial was conducted to compare Anturane and a placebo. Subjects were followed for a mean of eight months, and rates of death from cardiac causes were compared for the two groups.

This apparently well-conceived and well-executed RCT contains good examples of inappropriate exclusions. For example, seven patients who had received treatment until or almost until the day of death were withdrawn as ineligible. Six were in the sulfinpyrazone group and one in the placebo group. Moreover, in a detailed Food and Drug Administration (FDA) staff review, auditors found that patients from the placebo group who could have been declared ineligible, based on similar criteria, had not been declared ineligible. The study plan never mentioned an intent to exclude ineligible patients after entry, particularly patients who had died. Also, the Anturane Trial used the *7-day rule*. Any death to a patient who had not received treatment for at least seven days or who died more than seven days after termination of treatment was declared as *not analyzable*.

(Anturane Reinfarction Trial Research Group, 1978)

Drug Project Research Group, 1980). Clearly, when analyses are conducted on subgroups of randomized patients, the many advantages provided by randomization are lost. The only valid comparison includes all randomized patients.

Example 8.28

Clofibrate and Cardiac Deaths

An RCT compared the effectiveness of various cholesterol-lowering drugs for preventing cardiac deaths in men who had survived a myocardial infarction. Clofibrate was compared with a placebo. The five-year total mortality rate among clofibrate patients was 20.2%; this rate was not significantly different ($p = .55$) from the placebo patients (20.9%). Had patients who deviated from the protocol in the clofibrate group been eliminated (80% adherence), the resulting rate, 15.0%, would have been significantly different from the placebo group. However, these exclusions were demonstrated to be unacceptable when elimination of the patients who deviated from the placebo group resulted in a mortality rate of 15.1%—approximately equal to the 15.0% for the clofibrate group.

(Coronary Drug Project Research Group, 1980)

Controlling for Confounding in the Analysis

By design, confounding will not likely occur in an RCT, especially one that has been well conceived and carefully conducted. In general, the random assignment of study subjects to treatment groups effectively balances potential confounding variables by equally distributing the measured and unmeasured variables between the two groups. Although confounding could result from disparate distribution that occurred by chance, the possibility that confounding could occur is small. If the investigator suspects that the association between the treatment and the outcome is confounded by a third variable, standard statistical methodologies may be used to check and adjust for the confounding.

Effect Modification

Each subject who enters a clinical trial brings a set of personal characteristics (age, parity, etc.) that may be related to the study

subject's response to the treatment. We have already discussed these concepts in the section on baseline data. It is often informative to compare the effectiveness of treatments in subgroups of study subjects whenever we have reason to suspect that the reaction to the treatments may be different among these subgroups. Table 8.29 presents the relative risks of unscheduled IUD-related visits according to sexually transmitted disease (STD) culture status at time of IUD insertion: no gonorrhea and no chlamydia, gonorrhea positive only, and chlamydia positive only (Sinei et al., 1990). For this RCT of doxycycline prophylaxis at time of insertion and the risk of IUD-related pelvic infection, the risk of an unscheduled IUD-related visit differed according to STD culture status at the time the IUD was inserted. Among the women whose cultures were negative for both gonorrhea and chlamydia, those who were treated with doxycycline had a statistically significant lower risk of an unscheduled IUD-related visit [CIR = 0.62 (95% Confidence Interval: 0.45 - 0.86)] than women who received the placebo. Among the women whose cultures were positive only for gonorrhea, the treated women also had a reduced risk of an unscheduled visit, although the estimate was not statistically significant. Finally, among the women whose

Table 8.29

Rates and Relative Risks of Unscheduled Intrauterine Device—Related Visits During the First Month by Culture Status at the Time of Intrauterine Device Insertion and by Treatment Group

	Unscheduled Visit Rate			
Culture Status	Doxycycline	Placebo	Relative Risk	95% Confidence Interval
No gonorrhea or chlamydia	8.1%	13.0%	0.62	(0.45 - 0.86)
Gonorrhea positive only	10.5%	25.9%	0.41	(0.10 - 1.7)
Chlamydia positive only	10.4%	11.1%	0.94	(0.41 - 2.1)

(Sinei et al., 1990)

cultures were positive only for chlamydia, the treated women had approximately the same risk of an unscheduled IUD-related visit as women who received the placebo. Each STD culture status group was analyzed separately. Because the risk of an unscheduled IUD-related visit differed according to culture status at the time of IUD insertion, STD culture status is considered an effect modifier.

Advantages and Disadvantages of Randomized Clinical Trials

Clinical trials have more advantages than disadvantages:

- First and foremost, randomization is the only effective method known to control selection bias.
- Randomization will also likely balance the potential confounding variables.
- An RCT allows standardization of eligibility criteria, the exposures, and outcome assessments.
- An RCT is statistically efficient since equal numbers of exposed and unexposed are studied.
- An RCT is statistically efficient since not much statistical power is lost if and when confounding is controlled for in the analysis.
- An RCT is theoretically attractive since many statistical methods are based on the assumption that subjects are randomly assigned or selected.
- An RCT has concurrent comparison groups; any outside intervention is less likely to influence results since it should affect both groups equally.

Disadvantages include the following:

- Design and implementation of RCTs may be complex and expensive.

- RCTs can be subject to a lack of representativeness: patients who volunteer may differ substantially from the general population and the study population.

- An RCT may be open to ethical challenges: is it ethical to withhold a treatment from one group?

- Sometimes RCTs are impractical.

Scientifically, RCTs with blinding and adequate sample size and power are the ideal study design. Although an RCT is conceptually a more difficult design than a cohort study, it is probably easier to conduct correctly. In other epidemiologic study designs, selection bias and confounding may be difficult to manage. In general, a simple analysis suffices in the RCT, and selection and confounding bias are addressed adequately. Hence, the increased complexity in designing and implementing an RCT (including the randomization process, blinding, and the need for an adequate sample size) is more than compensated for by the ease of analysis and by the potential for increased validity of the results.

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F Study subjects are allowed to select their treatments.

   (b) T/F In the crossover design, each subject receives both treatments.

   (c) T/F A potential problem with the parallel design is the carryover effect.

   (d) T/F A major appeal of the crossover design is its efficiency.

   (e) T/F In a randomized clinical trial, the study population is composed of study subjects randomly assigned as treatment or control.

   (f) T/F In a randomized clinical trial, inclusion and exclusion criteria are implemented after randomization but before treatment.

   (g) T/F Random assignment to a treatment or a control group takes place after screening for eligibility and consent to participate.

   (h) T/F With a limited number of patients, a definite conclusion is more likely if the main comparisons are between two treatment groups.

   (i) T/F When attempting to determine if a new treatment is better than the standard treatment, it must be biologically plausible that the outcomes are substantially different from each other.

   (j) T/F Sample size issues minimally affect conclusions in a randomized clinical trial.

   (k) T/F Exclusions before randomization do not bias randomized treatment comparison.

   (l) T/F Exclusions before randomization do not affect the extrapolation of the results.

   (m) T/F Exclusions after randomization do not affect the extrapolation of the results.

(n) T/F Bias can result if study subjects in a randomized clinical trial are lost to follow-up.

(o) T/F Study subjects who deviate from protocol in a randomized clinical trial need not be followed.

(p) T/F Study subjects who do not adhere to the prescribed treatment in a randomized clinical trial should be analyzed as a separate group.

2. Multiple choice. Select one response.

2.1 Which two of the following approaches to randomization are recommended?

(a) Alternate assignment of patients by chart number
(b) Alternate assignment of patients by date of birth
(c) Coin flipping
(d) Card shuffling
(e) Use of random number table
(f) Balanced randomization

2.2 Which type of blinding is recommended and feasible when a surgical treatment is being compared to a drug treatment? Choose one.

(a) Single blinding: surgeon
(b) Single blinding: clinician
(c) Double blinding: patient and clinician
(d) Double blinding: patient and surgeon
(e) Triple blinding: surgeon, patient, and clinician

3. Arrange these steps of the recommended randomization method in the correct order by numbering them 1 through 4.

____ Use a random number table to select blocks from the master list of randomization blocks.

____ Based on the randomly selected blocks, put individual treatment allocations in opaque envelopes.

____ As study subjects enter trial, open envelope.

____ Set up a master list of randomization blocks.

4. For an RCT that compares a new treatment A with a standard treatment B, answer the following questions:

   (a) Calculate the number of possible sequences of block size four.

   (b) List all possible sequences for treatments A and B with the block size of four and number the sequences consecutively, using integers beginning with the number 1.

   (c) Your study requires 20 subjects in each of two treatment groups. How many blocks will you need to select from the master list?

(d) The following is a list of one-digit random numbers. Beginning with the first number, underline the numbers you can use to select blocks from the master list.

8 7 4 1 8 2 7 9 3 7 0 0 4 5 9 8 5 4 5 2 7 9 0

(e) Using the random numbers in (d), list the blocks from (c) in the order you selected them.

5. Circle true (T) or false (F).

(a) T/F Follow-up data describe the characteristics of the study subjects.

(b) T/F Baseline data analysis evaluates whether study groups are comparable.

(c) T/F Follow-up data analyses allow researchers to evaluate the response to treatment.

(d) T/F Statistical tests are important for comparing the baseline characteristics in the treatment and the control groups.

(e) T/F When the 95% confidence interval for the difference in proportions includes zero, the results are not statistically significant.

(f) T/F The RCT is an inexpensive type of research design.

(g) T/F One disadvantage of RCTs is that they do not ensure that each eligible subject has an equal chance of being in the treatment or the control group.

(h) T/F Outside intervention is unlikely to affect results, since it would affect treatment and control groups equally, especially if the investigator uses balanced randomization.

6. Consider the following randomized clinical trial of oral contraceptives and possible side effects and continuation of use.

Example 8.30

Norinyl Versus Brevicon in Sri Lanka

Background: It has been hypothesized that low-dose oral contraceptives (OCs) may be more suitable for Asian women who, in general, have a smaller body mass index than women who live in developed countries.

Research Hypothesis: Low-dose OC users are more likely to continue using OCs than standard-dose OC users.

Study Design: Randomized parallel clinical trial that is to be conducted in Colombo and two rural sites in Sri Lanka.

Eligibility criteria: Physically healthy women 18 to 40 years old were eligible for inclusion. Women were excluded if they were currently using OCs for therapeutic reasons, had been using OCs six months before the study, or had any known or suspected contraindications to OC use. Women were also excluded if they were currently breast-feeding or if their last pregnancy ended within three months before the study.

Treated: Brevicon (low-dose OC).

Not treated: Norinyl (standard-dose OC).

Outcome: Termination of OC use.

Follow-up: Follow-up data were collected at 1-, 3-, 6-, 9-, and 12-month visits.

Data Collection Method: Study subjects were visited at home to discuss possible side effects. All data were recorded by the study subject on a symptom calendar.

Randomization Scheme: Study subjects were assigned at random.

Blinding: OCs were packaged similarly. Study subjects were not aware of which OC they received.

(Basnayake et al., 1983)

(a) In Colombo, Sri Lanka, women who used Norinyl were significantly more likely to continue use than women who used Brevicon. However, in Colombo the follow-up staff were aware of treatment assignment and exclusively followed study subjects treated with one OC. Fewer Norinyl users were lost to follow-up; Norinyl users completed the symptoms calendar more frequently; and more home visits were made to homes of Norinyl users. In retrospect, how could procedures be modified to eliminate the problems that occurred here?

(b) In this study, Norinyl and Brevicon were compared at three locations in Sri Lanka. At the two rural locations, continuation rates for Norinyl and Brevicon were similar. In Colombo, Norinyl users were more likely to continue use than Brevicon users. Considering these results and the methodologic difficulties at the Colombo site, what would you conclude about continuation?

7. Consider the following randomized clinical trial of the effectiveness of perioperative Cefazolin in preventing infection after hysterectomy.

Example 8.31

Perioperative Cefazolin and Preventing Infection After Hysterectomy

Background: Approximately 700,000 elective hysterectomies are performed in the United States each year. The morbidity rate for these procedures ranges from 20% to 60% and is primarily related to postoperative infections. Several clinical trials have shown prophylactic antibiotics to be effective at reducing the number of infections after vaginal hysterectomy. Minimal information exists on the efficacy of antibiotic prophylaxis for elective abdominal hysterectomy.

Research Hypothesis: Perioperative Cefazolin by injection reduces pelvic and wound infections after elective abdominal hysterectomy.

Study Design: Randomized clinical trial.

Eligibility Criteria: All patients booked for elective nonradical abdominal hysterectomy were eligible. Patient refusal, physician refusal, evidence of infection, use of antibiotic within two weeks of hysterectomy, and a history of hypersensitivity to penicillin or any cephalosporin were reasons for exclusion.

Treated: Patients received Cefazolin 1.0 g intramuscularly one to two hours before surgery and again at 6 and 12 hours after surgery.

Not treated: Patients received equal volume of placebo by same route on same schedule.

Outcome: Infection.

Follow-up: Hospital records were reviewed daily until discharge. Letters were sent to physicians 6 weeks after discharge to inquire about subsequent infection.

Data Collection Methods: Patients and records were reviewed daily for data relevant to infection. When catheter was removed, urine specimen was obtained from catheter. A urine culture was performed on day of discharge.

Randomization Scheme: Random allocation (type not given).

Example 8.31 (continued)

Blinding: Physicians and patients were unaware of treatment assignment.

Data Analysis: The data were treated as a clinical trial with a short follow-up. Percent differences among women with infection were tested for statistical significance.

	Regimen	
Outcome	Received Cefazolin	Received placebo
Number infected	29	47
Number not infected	177	176
Totals	206	223
Percent infected	14%	21%

No weaknesses were described. This study was based on several earlier studies and could thus avoid their difficulties. However, some of the points raised below may indicate potential problems with the study and perhaps should have been addressed.

This study included vaginal and abdominal hysterectomies (elective, nonradical). The majority of the patients received abdominal hysterectomies (85%) and their results were the focus of our outline of the study.

During the study period, 1,511 women were eligible for the study, but only 567 (38%) were enrolled. The primary reasons for nonenrollment were patient refusal (32%), physician refusal (30%), recent antibiotic therapy (14%), drug allergy (8%), and other (16%). Of the 567 who were enrolled, 52 (9%) were withdrawn (26 treatment, 26 control) and given other treatments, usually because of prolonged or complicated surgery.

(Polk et al., 1980)

(a) Do you think the women enrolled in the trial were a random sample of those eligible for the study? Why or why not?

(b) Were they a sample that was representative of those eligible? Why or why not?

(c) What effects do you think the withdrawal of 52 patients had on the study results?

8. Design an RCT to test whether the Today sponge is as safe a method of contraception as other vaginal methods. State the problem and the hypothesis. Describe the eligibility criteria. Define the treatments and the outcome in measurable terms. Describe randomization, blinding, and follow-up procedures. Outline analysis tables.

Problem:

Research Hypothesis:

Study Design:

Study population:

Eligibility Criteria:

Treated:

Not treated:

Outcome:

Follow-up:

Data collection methods:

Randomization Scheme:

Blinding:

Suggested Answers to Practice Exercises

1. True or false.

   (a) F Introduction. Investigators assign treatments to study subjects.

   (b) T Successive Treatment Designs

   (c) F Successive Treatment Designs. Carryover effect is a problem of crossover design.

   (d) T Successive Treatment Designs. Since study subjects receive more than one treatment, a crossover design requires fewer subjects than a parallel design.

   (e) F Study Population. No, the study population are the individuals from the general population defined according to the study criteria, and the study subjects in the trial are usually a sample of the study population.

   (f) F Eligibility Criteria. No, inclusion and exclusion criteria must be defined and implemented before randomization.

   (g) T Selecting the Study Subjects

   (h) T Defining Comparison Groups. Comparison of only two groups gives the maximum chance of being able to make some definitive conclusions for a limited number of patients.

   (i) T Defining Comparison Groups

   (j) F Sample Size. No, many researchers design studies that are too small to answer the questions posed.

   (k) T Exclusions Before Randomization

(l) F Exclusions Before Randomization. No, exclusions before randomization do affect the extrapolation of results. As the number of excluded study subjects increases, the results become less relevant to the general population.

(m) F Exclusions After Randomization. No, exclusions after randomization can bias the randomized treatment comparison and can affect the extrapolation of results.

(n) T Exclusions After Randomization

(o) F Exclusions After Randomization. No, study subjects who deviate from protocol should be followed and analyzed in the groups to which they were originally assigned.

(p) F Exclusions After Randomization. See 1 (o).

2. Multiple choice.

2.1 e Randomization
f Randomization

2.2 b Blinding

3. Steps of recommended randomization method.

2, 3, 4, 1

4. Comparing treatments A and B.

(a) $n = \frac{4!}{2!2!} = \frac{24}{4} =$ 6 different sequences

(b) Number sequence. (Note: You may have listed your sequences in a different order.)

1 A A B B
2 A B A B
3 B A A B
4 B A B A
5 B B A A
6 A B B A

(c) Ten blocks required.

(d) 8 7 4 1 8 2 7 9 3 7 0 0 4 5 9 8 5 4 5 2 7 9 0

(e) (4) B A B A
(1) A A B B
(2) A B A B
(3) B A A B
(4) B A B A
(5) B B A A
(5) B B A A
(4) B A B A
(5) B B A A
(2) A B A B

5. True or false.

(a) F Follow-up Data Analysis. No, follow-up data provide the incidence data—the number of study subjects treated and the number of study subjects not-treated who develop the outcome or health problem of interest.

(b) T Baseline Data Analysis

(c) T Follow-up Data Analysis

(d) F Baseline Data Analysis. No, differences in distributions of prognostic variables, regardless of statistical significance, may be related to the outcome.

(e) T Follow-up Data Analysis

(f) F Advantages and Disadvantages

(g) F Advantages and Disadvantages

(h) F Advantages and Disadvantages

6. Randomized Clinical Trial—Norinyl and Brevicon

(a) Apparently the follow-up staff for Norinyl users was more effective in locating study subjects than the follow-up staff for Brevicon users. This difference likely affected study results. To avoid this problem, study subjects followed by clinic staff should be evenly divided between Norinyl and Brevicon users.

(b) The investigators were not able to demonstrate a convincing difference in continuation between Norinyl and Brevicon users.

7. Randomized Clinical Trial—Cefazolin

(a) No. Women volunteered for the study and as volunteers were not selected at random.

(b) They are not likely to be representative of those eligible because they are volunteers.

(c) Very little. The withdrawal was balanced (same number in each group) and the reasons for withdrawal appear to be the same for both treatment and control groups. If randomization had been delayed until the surgery was completed, prolonged or complicated surgery could have been used as an exclusion criterion.

8. Randomized clinical trial—Today Sponge. The answer given here is a suggested answer and is based on the report of the actual study.

Example 8.32

A Trial of the Contraceptive Sponge and Diaphragm

Background: The Today sponge was designed to be easy to use, provide 24-hour protection against pregnancy, and be available without prescription.

Research Hypothesis: Sponge users are not as likely as diaphragm users to be compliant users.

Study Design: Randomized multiclinic trial.

Study Population: At least 50 volunteers attending each of 13 clinics.

Eligibility criteria: Women aged 18 to 40 years who were in good health, had no anatomic abnormalities, were sexually active, had had at least one menstrual period since the termination of last pregnancy, and agreed to return for regular follow-up visits were eligible for participation in the trial. Women were excluded after randomization if they were not sexually active, had had no menses since last pregnancy, or were older than 40 years.

Treated: Sponge.

Not Treated: Diaphragm.

Outcome: Product-related problems, discontinuation, pregnancy.

Follow-up: Every three months for one year.

Data Collection Methods: Common protocol. Data were recorded on standard forms using a common protocol. Lab tests on admission to study and at the final 12-month clinic visit provided hematocrit, complete blood count, urinalysis, Veneral Disease Research Laboratories (VDRL), gonorrhea culture, Papanicolaou smear, and pregnancy test.

Randomization Scheme: Study subjects were volunteers who were randomized to use the sponge or diaphragm.

Blinding: Other details about the randomization were not provided.

(Edelman et al., 1984)

References

Anturane Reinfarction Trial Research Group. Sulfinpyrazone in the prevention of cardiac death after myocardial infarction. N Engl J Med 1978;298:289-295.

Basnayake S, de Silva SV, Miller P, Rogers S. A comparison of Norinyl and Brevicon in three sites in Sri Lanka. Contraception 1983;27:453-464.

Badan Kerja Sama Penelitian Fertilitas, Badan Koordinasi Keluarga Berencana Nasional, United States Agency for International Development. Perbandingan Pil Kontrasepsi, Noriday-Kimia Farma-Ovostat 28 DI: Indonesia (Oktober 1982 s/d Juli 1985). Bandung, Indonesia: Badan Kerja Sama Penelitian Fertilitas Indonesia, Februari 1986. Indonesia, March, 1986.

Bonhomme, M. [Unpublished study protocol]. Clinical trial of oral contraceptive use among women with sickle cell disease. [1984]. Located at: Family Health International, Research Triangle Park, North Carolina.

Cole CH, Lacher TG, Bailey JC, Fairclough DL. Vaginal chemoprophylaxis in the reduction of reinfection in women with gonorrhoea. Br J Venereal Disease 1980;56:314-318.

Cole LP, Potts DM, Aranda C, et al. Comparative copper IUD trials. In: Zatuchni GI, Goldsmith A, Sciarra JJ eds. Intrauterine contraception: advances and future prospects. Philadelphia: Harper and Row Publishers, 1984:95-100.

Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 1980;303:1038-1041.

Dusitin N, Varakamin S, Ningsanon P, Chalapati S, Boonsiri N, Gray R. Post-partum tubal ligation by nurse-midwives and doctors in Thailand. Lancet 1980;1:638-639.

Edelman DA, McIntyre SL, Harper J. A comparative trial of the Today contraceptive sponge and diaphragm. Am J Obstet Gynecol 1984;150:869.

Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized control trial. N Engl J Med 1978;299(13):245-249.

Lavin P, Bravo C, Waszak C. Comparison of TCu 200 and Progestasert IUDs. Contracept Deliv Syst 1983;4:143.

McLaughlin L. The pill, John Rock, and the church. Boston: Little, Brown and Company, 1982.

Meier P. Anturane reinfarction trial [Letter]. N Engl J Med 1981;304:730.

Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Br J Cancer 1976;34:585-612.

Polk BF, Tager IB, Shapiro M, Goren-White B, Goldstein P, Schoenbaum SC. Randomized clinical trial of perioperative Cefazolin in preventing infection after hysterectomy. Lancet 1980;1:437-440.

Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Bethesda, Maryland: National Institutes of Health, 1979; DHEW publication no. (NIH)79-1649.

Schafer A. The ethics of the randomized clinical trial. N Engl J Med 1982;307:719-724.

Sinei SKA, Schulz KF, Lamptey PR, et al. Preventing IUCD-related pelvic infection: the efficacy of prophylactic Doxycycline at insertion. Br J Obstet Gynaecol. 1990;97:412-419.

Temple R, Pledger GW. The FDA's critique of the anturane reinfarction trial. N Engl J Med 1980;303:1488-1492.

Wheeler RG, Cole LP, Galvez RS. IUDs with or without strings. In: Zatuchni GI, Goldsmith A, Sciarra JJ eds. Intrauterine contraception: advances and future prospects. Philadelphia: Harper and Row; 1985:420-426.

Chapter 9 - Learning Objectives

After completing this chapter you should be able to:

1. Define the terms:

   prospective cohort study
   retrospective cohort study
   exposed group
   unexposed group
   selection bias
   information bias

2. Describe how study subjects will be identified with respect to exposure status.

3. Define the study outcome in measurable terms.

4. Specify the data collection methods to be used in the study.

5. Develop a data table for analysis of the association between the exposure and outcome.

6. Calculate relative risk and confidence limits for cumulative incidence relative risk and incidence density relative risk.

7. Interpret relative risk.

8. Recognize advantages and disadvantages of cohort studies.

9. Design a hypothetical cohort study in outline form.

9 Cohort Studies

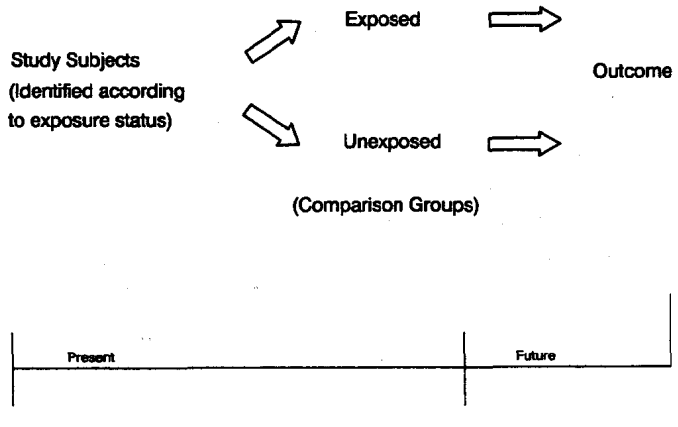
Introduction

Cohort study defined

The cohort study is an analytic epidemiologic research design in which the study population is composed of individuals who are classified as exposed or not exposed to a particular risk factor (comparison groups). These groups are followed forward for a specific period of time to estimate the incidence rates of an outcome or the development of a health problem (Figure 9.1). The exposed group is composed of individuals who have been exposed to a postulated causal or protective factor for a health problem. The unexposed group is composed of individuals who are similar to the exposed group but who are known not to have been exposed to the causal or protective factor. Depending on whether the study is designed to examine a postulated causal or protective factor, the individuals in the exposed group are hypothesized to be at greater or lesser risk of developing the

Figure 9.1

Definition of Comparison Groups in Prospective Cohort Studies



health problem than the individuals in the unexposed group. In designing a cohort study, investigators should clearly state the research hypothesis and specify the expected relationship between the exposure groups and the health problem of interest.

Cohort studies are similar to randomized clinical trials in that they proceed from exposure to outcome. Unlike the randomized clinical trial, however, the investigator observes rather than assigns exposure status. The study groups are identified by exposure status before outcome status. Each person in both the exposed and unexposed groups (study subjects) is followed in an identical manner until one of the following outcomes occurs:

- The study subjects develop the health problem under study.
- The study subjects die.
- The study ends.
- The study subjects are lost to follow-up.

In the randomized clinical trial, the investigator determines the exposure status. In the cohort study, the exposure status is determined by:

- genetics or biology (i.e., male or female, presence or absence of sickle cell anemia)
- a conscious choice made by the subject (smoker or nonsmoker; contraceptive user or nonuser)
- circumstances (living in a poorly served rural area or in a large city).

Prospective cohort study

Historical cohort study

Cohort studies may be either prospective or historical. In a prospective cohort study, exposure status is determined when the study begins, and the exposed and unexposed groups are followed forward in time to observe whether they develop the study outcome. In a historical cohort study, the study subjects have developed the health problem or outcome and have been exposed before the study begins. The exposure and the health problem are determined from existing records. A historical cohort study is essentially the reconstruction of a cohort study that has already taken place. In Example 9.2, we use a hypothetical study of maternal thalassemia and fetal death *in utero* to illustrate a prospective and a historical cohort study design.

Example 9.2

Maternal Thalassemia and Fetal Death in Utero

A hypothetical cohort study is to be conducted to assess the association between maternal thalassemia and fetal death in utero.

Design 1: A Prospective Cohort Study

In a hypothetical prospective cohort study, women with thalassemia who come to a clinic for their first prenatal visit would compose the exposed group. The unexposed group would be women without thalassemia who also come to the clinic for their first prenatal visit. In both groups, each woman would be followed throughout her pregnancy to observe the occurrence of fetal death in utero (outcome).

Design 2: A Historical Cohort Study

The association between maternal thalassemia and fetal death in utero could also be evaluated using a historical cohort study design, if the quality of the records and the follow-up procedures at the study site were good. Using this design, investigators would identify all women with thalassemia who came to the clinic for their first prenatal visit at some specified time in the past (exposed), and all women without thalassemia who had their first prenatal visit during the same time period (unexposed). The medical records of these women would be reviewed to determine whether fetal death in utero (outcome) had occurred. In a historical study, the exposure (thalassemia) would be identified first; then the outcome (fetal death) would be ascertained. Both the exposure and outcome may have occurred months or even years in the past. This design is a cohort study because exposed and unexposed groups are the comparison groups; it is retrospective because both exposure and outcome occurred before the study was initiated.

In cohort studies, investigators can examine more than one health problem that may be the result of the exposure under study. For example, in a study of the effects of oral contraceptives, investigators may be primarily interested in cardiovascular disease, but the study could provide the opportunity to examine a variety of other outcomes hypothesized to be related to oral contraceptive exposure, including deep venous thrombosis, myocardial infarction, and cerebrovascular accidents.

Design and Data Collection Methods

Population Sources and Follow-up

An important step in designing a cohort study is selecting exposed and unexposed populations that can be adequately followed over time. A high rate of follow-up for both exposed and unexposed study subjects is needed if the conclusions of the study are to be considered valid. Both exposed and unexposed groups will need to be followed in an identical manner to determine if they develop the study outcome. For example, if the exposed subjects are followed up with in-person interviews and physical examinations, the unexposed subjects should be followed up by identical data collection methods to minimize bias in the detection of the health problem.

The population used in a given study will depend on the exposure and outcome of interest and on the ease of gathering sufficient information on exposure and outcome in the population. The choice of the study population may be affected by the delay between exposure and outcome (latency). If latency is long, extensive follow-up will be necessary. Potential sources for the study population may include the general population, a sample of the general population, special groups (such as nurses or government employees who can be readily followed over time), attendees to clinics with good record-keeping and follow-up procedures, or occupational groups with different levels of exposure. Populations or samples of populations are most appropriate for a cohort study design when the exposure is relatively common. If the study exposure is uncommon, investigators should select special groups who have experienced a higher level of exposure than the general population. For example, investigators who study a possible association between injectable contraceptives and cancer might use a population of family planning clinic attendees instead of women in the general population, especially if the prevalence of contraceptive use is low in the population.

Latency

Selection of the Exposed Group

The exposed group

The exposed group in a cohort study are the individuals who have the exposure of interest. Before the exposed individuals can be identified, the investigators must develop an unambiguous and objective description of what constitutes exposure. Where applicable,

the definitions should include the minimal acceptable levels of the exposure (e.g., more than ten cigarettes per day) and the minimal duration of exposure (e.g., use of oral contraceptives continuously for at least a year). Other eligibility criteria for entry in the study, such as age, sex, and absence of preexisting medical problems, should also be determined. In particular, the individuals should not have a history of the outcome.

A cohort study may have more than one exposed group; the merits of using more than one exposed group should be considered in the initial study design. For example, in a study of smoking and intrauterine growth retardation (IUGR), the investigators may wish to compare the effects of heavy smoking ($\geq$ 20 cigarettes per day) with the effects of lighter smoking ($<$ 20 cigarettes per day) on IUGR. The study could include two exposure groups: the group that smokes $<$ 20 cigarettes per day, and the group that smokes $\geq$ 20 cigarettes per day. In the analysis, both exposure groups would be compared with the women who do not smoke during pregnancy (unexposed group). However, the number of heavy smokers may be too small to make valid conclusions about the association. In general, if detailed information about the effects of different levels of exposure is needed, the investigators should consider using more than one exposure group. Example 9.3 presents a cohort study of the association between oral contraceptive use and breast cancer among nurses (Romieu et al., 1989). In this study, women who were current users of oral contraceptives (OC) and women who were past users were followed up and compared to women who had never used OCs.

Preferably, only the study subjects who are potentially at risk for the outcome of interest should be enrolled in the study. For example, investigators studying smoking and adverse pregnancy outcomes should exclude women who have had a tubal ligation because these women are not at risk of pregnancy. The decision to include or exclude individuals from the study population will depend on the exposure and outcome of interest and on whether heterogeneity is managed by restricting admission to certain groups or by performing stratified analyses after the data are collected. In general, the more restrictive the admission criteria, the more difficult it is to assemble an appropriate cohort.

Example 9.3
Oral Contraceptives and Breast Cancer

Problem: Are oral contraceptives associated with an increased risk of breast cancer?

Research Hypothesis: Women who use oral contraceptives have a higher risk of breast cancer than women who have never used oral contraceptives.

Study Design: Prospective cohort study of 121,700 female registered nurses 30 to 55 years old who are living in 11 states of the United States.

Exposed: Women who had used oral contraceptives for at least one month in the past and who had never experienced angina pectoris, myocardial infarction, or stroke before the beginning of the study in 1976. Women who had used oral contraceptives were further categorized as women who were currently using oral contraceptives and women who had used oral contraceptives in the past but were no longer currently using them.

Unexposed: Women who had never used oral contraceptives and who had never experienced angina pectoris, myocardial infarction, or stroke before the beginning of the study in 1976.

Outcome: All breast cancer cases reported by the woman (or next of kin for decedents) for whom relevant hospital records confirmed the diagnosis.

Follow-up: The follow-up was conducted every two years using a mailed questionnaire. A repeat mailing was done for those who did not respond. Approximately 5% of the women were lost to follow-up.

Data Collection Methods: Self-administered questionnaires included information on medical conditions and life-style practices. Follow-up data were collected with self-administered questionnaires every two years. Outcomes reported by women were verified by review of medical records; deaths were reported by next of kin or postal authorities (when the questionnaires could not be delivered) or identified from vital records.

Results:

	Oral Contraceptive Use		
	Current	Past	Never
Breast cancer	32	685	1,041
Person-years	22,622	472,828	592,364
Rate/1,000 person-years	1.41	1.45	1.76
Crude IDR	0.8	0.8	
95% CI	(0.6, 1.1)	(0.7, 0.9)	

(Romieu et al., 1989)

Selection of the Unexposed Group

The unexposed group

The unexposed subjects should be similar to the exposed in all ways except they should not have the exposure under study. The unexposed group is used to determine the incidence of the health problem among those without the exposure. Then, the investigators compare the incidence of the outcome among the individuals in the exposed and unexposed groups. Thus, the exposed and unexposed populations should be as similar as possible.

In a cohort study, the unexposed subjects should have the same general criteria for entry into the study as the exposed (e.g., age, sex, absence of preexisting medical problems, no history of the outcome of interest). They should also be at potential risk of developing the outcome under study. Finally, the unexposed study subjects should have the same opportunity as the exposed to be diagnosed with the outcome.

Selection processes

Ideally, the unexposed study subjects should be chosen using the following criteria:

- Take the total population or a sample of the population (e.g., all the attendees at a specific clinic or a random sample of the population in a rural village).
- Determine the exposure status of each individual.
- Classify each individual into the appropriate exposure category.

By using these built-in or internal comparison groups, investigators are afforded considerable advantages that include decreasing the potential for bias attributable to differences in the study populations and to differences in ascertaining the outcome between the exposed and unexposed study subjects. In the study outlined in Example 9.3, participants were asked on the initial questionnaire whether they currently used or whether they had ever used oral contraceptives (Romieu et al., 1989). Based on this information, the investigators divided the cohort into women who had ever (currently or in the past) used oral contraceptives (exposed) and women who had never used oral contraceptives (unexposed).

Consider the following hypothetical cohort study that also uses an internal comparison group to examine the effect of low birthweight on infant survival. The investigator designed a study in which all live-born infants in five villages were weighed at birth by midwives using simple scales and followed for one year. The total population of babies could be divided into an exposed group (birthweight $< 2{,}500$ g) and an unexposed group (birthweight $\geq 2{,}500$ g) and followed at monthly intervals to determine the outcome of survival at the end of the study period (the child's first birthday). Alternatively, all the low-birthweight babies could be used as the exposed group and only a sample of the $\geq 2{,}500$ g babies would be used as the unexposed group, since these larger babies will considerably outnumber the low-birthweight babies.

In another hypothetical cohort study, investigators used an internal comparison group to investigate the effects of anemia on premature births among a population of women who attend a prenatal clinic. The investigators could screen all women for anemia who come to the clinic for the initial prenatal exam in their first trimester during a specified year. Those women with low hemoglobin could comprise the exposed group, and those with normal hemoglobin could comprise the unexposed group. The women in both groups could be followed to determine who delivered premature infants.

When an appropriate unexposed group cannot be selected within the same population, two less-than-ideal methods are sometimes used. The first method compares the exposed group with another group that is similar in composition but does not have the exposure. In Example 9.4, investigators used this method to select a sample of unexposed women.

The second method compares the outcomes among exposed study subjects and a population rate. Using this method, investigators compare the outcomes in a study cohort and the experience of the general population while the cohort is being followed. For example, investigators could compare the mortality rates of women who have used oral contraceptives and died of thromboembolism and the mortality rates of women of comparable age in the general population. This method is only possible when population rates are available (e.g., death rates).

Example 9.4

Selection of an Unexposed Group from a Different Population

Investigators designed a study to examine the effects of malaria on placental weight among women in certain villages; almost all of the women are infected with malaria. To obtain the unexposed group, the investigators select women from villages located in nonmalarious areas. Although the selected villages are demographically similar, and the villagers have access to comparable health services, the investigators' decision to use women from different villages is not ideal. The exposed and unexposed populations may differ in tangible and intangible ways, and they may not be strictly comparable. There may not be other options in this situation, and the investigators may have difficulty determining whether differences in the relative risk for the health problem under study are related to the exposure or to differences in the two populations.

Matching

Individual and frequency matching

Matching refers to procedures for selecting a study group and a comparison group so that they are comparable with respect to extraneous factors (Last, 1988). *Individual matching* involves identifying individuals for the comparison group such that each resembles a certain study subject on the matched variable(s). *Frequency matching* involves matching on selected variables so that the frequency distributions of the matched variables are similar in the study and comparison groups (Last, 1988) (Example 9.5). In frequency matching, the investigator estimates the number of exposed subjects in a given subgroup before the study begins. The estimate is based on a preliminary analysis of the distribution of individuals with the exposure of interest; then the requisite number of unexposed subjects in the subgroup is enrolled in the study (Last, 1988).

In most circumstances, the preferred approach to matching is simply to take a population or population sample, separate the selected individuals into exposed and unexposed groups, and then in the analysis stratify by other variables that are likely to affect the relationship between exposure and outcome. Investigators should be cautious about matching for several reasons:

Example 9.5

Individual and Frequency Matching

A Hypothetical Study

Investigators design a study to examine the effects of anemia on pregnancy outcome among women at a certain prenatal clinic. They decide that parity is likely to affect the relationship between the two variables and that matching may be necessary. Two methods of matching are considered.

Frequency matching. The investigators determined in a preliminary study that 20% of the women with anemia who come to the clinic for their first prenatal visit have had no previous pregnancies. They decide to enroll 100 women with anemia and 100 women without anemia. For the unexposed group, they would identify women without anemia who have had no previous pregnancies as they come into the clinic, until 20 women (with no children) are enrolled.

Individual matching. If this method is used, the investigators could match the first woman diagnosed with anemia who has had no previous pregnancy (exposed) with the next woman seen at the clinic who is not anemic and who has had no previous pregnancy (unexposed). If the next woman to be seen at the clinic has anemia and has had three children, she would be matched with the next woman who has had three children but who is not anemic. This enrollment process would continue until the required number of women is obtained for the study.

Cautions about matching

- Matching on a particular variable prohibits studying its association with the outcome.
- Matching may greatly increase the amount of work required to find appropriate unexposed individuals with the same characteristics as the exposed.
- A matched variable that is in the causal pathway between exposure and disease, or a variable that is related to the outcome but not the exposure may cause problems in interpreting the results of a study.
- Individual matching requires a matched analysis to ensure that the individuals in the matched sets are compared to one another.

Since matched analyses can be difficult, matched designs should be used only when the advantages outweigh the disadvantages.

Measuring Exposure

In a prospective cohort study, investigators measure exposure at the time of enrollment. In some studies, persons may be enrolled at the time they first experience the exposure (e.g., at the time they begin using a certain contraceptive or have a sterilization procedure). More commonly, the exposure has been occurring for some time (e.g., smoking), and both current and historical exposure data are generally collected. Exposure status may be determined from interviews, self-administered questionnaires, or existing records, such as clinic or hospital charts and laboratory or employment records. To adequately describe the relationship between an exposure and an outcome, investigators should include measures related to frequency, duration, dose, and timing (e.g., dates of first and last exposure).

Sources of exposure data

Some exposures are one-time events (e.g., a surgical procedure such as a tubal ligation or vasectomy) or are biologically determined and unchangeable (e.g., hemoglobinopathies or maternal height). Others, however, may change after an individual is enrolled in the study. For example, a woman who is using an intrauterine device (IUD) might decide to change methods or might choose not to use a contraceptive; a person classified as a smoker at the beginning of the study may quit during the study. Alternatively, an initially unexposed person may begin to use an IUD or may begin smoking after being enrolled in the unexposed group. In these instances, maintaining the original designation of exposed or unexposed is a conservative approach that tends to bias the results toward the null effect.

From a conservative perspective, the exposure status of the study subjects should not be changed, for purposes of analysis, after initial classification, but investigators should note changes in exposure status over time. If such information is available, investigators can look at the effect of differing durations or intensities of exposure on the health problem under study. For example, it might be interesting to determine if the development of a myocardial infarction is dependent on the length of time a

woman uses OCs or how long the risk of an infarction persists after a woman stops using them.

Determination of Outcome

When the study begins, the outcome must be defined precisely and as unambiguously and objectively as possible. The definition must be applied uniformly for both the exposed and unexposed. If many medical providers will be diagnosing the health problem or identifying the outcome under study in a number of different study sites, the investigators should consider the skills of the providers and the likelihood that certain diagnostic tests will be performed when they define the outcome. If possible, the processes by which an outcome is defined and diagnosed should be standardized and outlined in a training program. However, such a program must be designed carefully. If the medical providers are made aware of the hypothesis under study, they may be more likely to diagnose the outcome among the individuals in the exposed group than among the individuals in the unexposed group.

In a historical cohort study the medical providers who diagnose the outcome should be unaware of the exposure status of the study subjects since such knowledge may potentially bias the investigators' determination of the outcome. Restricting knowledge of any aspect of the study from anyone involved in the study is referred to as blinding.

Determining outcome may be logistically more complicated in a prospective cohort study. For example, if the disease (e.g., cancer) has a long latency period, the follow-up will have to be conducted over a prolonged period of time. When the study is designed, the investigators must develop methods to ensure the highest possible follow-up rates for both exposed and unexposed study subjects.

Follow-up methods

The specific methods used to determine outcome will depend on the outcome under study. If the outcome involves changes in behavior or minor illnesses that are unlikely to require hospitalization, for example, follow-up would involve recontacting the subjects on a periodic basis. For individuals with more severe outcomes that typically require hospitalization, investigators may periodically review the admissions records at the hospitals where the study subjects are likely to go if they become ill. For outcomes such as death or cancer, the best method of follow-up may be to review death certificates or cancer registries on a regular schedule.

In the design of the study, the investigators must carefully consider the methods that will be used to adequately ascertain the outcome, since the resources required to perform the study and the validity of the study's results are highly dependent on the follow-up methods. If the follow-up is to be conducted through regular reviews of vital records (or cancer registry records), the detailed personal identifiers for each individual should be collected at the beginning of the study and attempts should be made periodically to update this information. If follow-up is conducted through contact with the study subjects or their health care providers, the investigators may want to gather not only personal identifiers on the individual but the names of relatives or friends who are likely to know where to locate the individual in the future. Well-documented specifications on what constitutes adequate follow-up is not available, but if more than 40% of the study subjects are lost to follow-up, results of the study are likely to be seriously questioned. Even a 20% loss to follow-up may introduce doubts about a study's validity. As in historical cohort studies, the outcome should be ascertained without knowledge of the exposure, and identical methods must be used to follow up and ascertain outcomes for individuals in both the exposed and unexposed groups. These measures are necessary to avoid biases that may lead to erroneous conclusions.

Loss to follow-up

Examples 9.6 and 9.7 illustrate follow-up procedures and methods used to ensure the complete and accurate assessment of outcomes.

Example 9.6

Risk Factors for Maternal Mortality in Bangladesh

Problem: Is increased maternal age associated with a higher risk of maternal death in Bangladesh?

Research Hypothesis: In Bangladesh, women who are at least 30 years old at delivery have a higher risk of maternal death than younger women.

Study Design: Prospective cohort study

Exposed: Women in the study population who are at least 30 years old.

Unexposed: Women in the study population who are younger than 30 years old.

Example 9.6 (continued)

Outcome: Death from any cause during pregnancy, at delivery, or in the 42 days after termination of the pregnancy.

Follow-up: Follow-up was conducted through personal interviews every two weeks until 42 days after delivery to determine pregnancy outcomes.

Data Collection Methods: Interviewers identified pregnant women during visits to each of the study villages and tracked their outcomes.

(Alauddin, 1986)

Example 9.7

Instillation Methods for Second Trimester Abortion

Problem: Is urea-prostaglandin instillation a safer method than saline instillation for second trimester abortion.

Research Hypothesis: Women who undergo urea-prostaglandin abortions have less risk of complications than women who undergo saline instillations.

Study Design: Prospective cohort study

Exposed: Women at least 20 years old who had had urea-prostaglandin instillation abortions at one of 13 study institutions during 1975-1978.

Unexposed: Women at least 20 years old who had had saline instillation procedures at the same institutions.

Women in both groups who had concurrent sterilization procedures, ectopic pregnancies, and hydatidiform moles were excluded from the analysis.

Outcome: The study outcomes were defined as the following major complications: fever $\geq$ 38°C for 3 or more days, hemorrhage requiring transfusion, or unintended abdominal surgery.

Follow-up: Complications were recorded on routine follow-up visits 2 to 6 weeks after abortion.

Data Collection Methods: Data were collected through regular reviews of records from participating clinics.

(Binkin et al., 1983)

Validity and Bias

When conducting epidemiologic studies, investigators must be concerned about making false conclusions that are a result of the research methodology used. The purpose of the cohort study design is to allow the investigator to determine what effect the exposure has on the outcome. If the methodology used by the investigator alters the true effect of an exposure, this misrepresentation is called bias. If no bias is present in the study, the measure of effect calculated is said to be a valid estimate of the effect of the exposure on the health problem.

Potential sources of bias

Two important types of bias can occur in cohort studies—selection bias and information bias. Potential sources of bias should be carefully considered in the design of a study. Except for the exposure of interest, the study subjects should have comparable risks of developing the outcome of interest.

Selection bias. This type of bias may be present if the individuals who are enrolled in the study have different characteristics than the population they are supposed to represent. Selection bias occurs if the individuals selected for participation in the study differ somehow from the individuals who were not selected (Examples 9.8 and 9.9).

Example 9.8

Delivery During the Harvest Season and the Risk of Complications
Selection Bias

A Hypothetical Study

A physician observes that women who deliver their infants during the harvest season tend to have more labor and delivery complications than women who deliver their infants at other times of the year. He hypothesizes that the complications may be attributed to high levels of physical activity in the weeks before labor since most women are working in the fields up to 12 hours per day during the harvest season. He designs a hospital-based study to compare women who have their infants during the harvest (the exposed) and women who deliver 3 months after the harvest, when routine daily activities have returned to normal (unexposed); the women are matched on age and parity. He finds that complications are, in fact, greater for women who deliver during the harvest. He also notes that the number of hospital-based

Example 9.8 (continued)

deliveries declined considerably during the harvest months. In a discussion with local midwives and trained birth attendants, he discovers that during the harvest season, women tend to deliver in their villages to minimize the time away from work. Most of the women who do make the long trip to the hospital during this period are those with prolonged or difficult labor.

This study provides an example of selection bias. Although women who deliver in the hospital after the harvest may be fairly representative of the women in the hospital's catchment area, those women who deliver in the hospital during the harvest season are not representative of all women who have their babies during this time period. As a result, the study probably overestimates the risk of complications related to increased physical activity. To conduct this study properly, the physician should conduct a population-based study to examine complication rates among women who deliver at hospitals, at health centers, and at home.

Example 9.9

Intrauterine Devices and Pelvic Inflammatory Disease
Selection Bias

Another example of potential selection bias is a hypothetical cohort study in which an association between intrauterine device (IUD) use and pelvic inflammatory disease (PID) is investigated. If women using oral contraceptives and barrier methods composed the unexposed group, an erroneous conclusion might be made, since both methods of contraception are known to reduce the risk of PID. A better alternative for the unexposed group would be women who are sexually active but not using contraceptives. The study could be matched or stratified by number of sexual partners and any other variables that may be different between IUD users and women who do not use contraceptives, or by variables that may affect the incidence of PID independent of exposure to IUDs.

Information bias

Information bias. If information on exposure or outcome is obtained differently for the exposed and unexposed groups, information bias may result. In prospective cohort studies, information bias related to exposure is minimal because exposure status is determined before the disease occurs. However, in prospective cohort studies, the outcome may be subject to information bias because exposure status is known before the outcome is

determined. In historical cohort studies, both exposure and outcome information are potentially subject to this bias because information on both variables is collected after the outcome has occurred. For example, information bias might occur if ascertainment of outcome among the individuals in the exposed group was based on periodic attempts to recontact the individual and a review of death certificates, while outcome for the unexposed individuals was obtained by only reviewing death certificates. Information bias is best avoided by ensuring that the same attention is given to each study subject in the exposed and the unexposed groups. Preferably, the persons who collect outcome information should be blinded to the exposure status of the study subjects (Example 9.10). Information bias generally is best managed while the study is in the field and not during the analysis phase.

Example 9.10

Oral Contraceptives and Cervical Carcinoma in Situ
Information Bias

In a hypothetical study of oral contraceptive (OC) use and cervical carcinoma *in situ*, women who use OCs compose the exposed group and women who use the barrier method compose the unexposed group. The women are identified through visits to a family planning clinic. Women with cancer *in situ* are identified from hospital laboratory records. A potential concern is that OC users who return more frequently for prescription renewals are more likely to receive a Pap smear and to be diagnosed with cancer *in situ* than women who use barrier methods. To overcome this problem, the study protocol states that women in both the exposed and unexposed groups should receive Pap smears at the same intervals.

Another form of information bias could occur if, during the study, results showing a positive link between OC use and carcinoma *in situ* were published by other investigators. As a result, physicians might be more likely to perform Pap smears more often on women using OCs. Additionally, the pathologists who read the Pap smears may be more likely to read borderline smears in an OC user as positive if they are aware of the woman's contraceptive status. The former problem is best prevented by requiring that the Pap smears be done at comparable intervals for women in the exposed and unexposed study subjects. The problem of exposure status affecting the pathologist's reading is best managed by having the pathologist read the smears without knowledge of the exposure status.

Data Analysis Methods

Ideally, the analytic methods are determined concurrently with the design of the study protocol and the data collection forms to ensure that information is collected for all required variables and that the sample size is adequate. After data collection is complete, the first steps of the analysis are to organize the data into the tables developed when the analysis was planned. Several types of tables are used to organize cohort study data and to make comparisons between exposed and unexposed study subjects. The first tables presented for

Table 9.11

Characteristics of Women Undergoing Abortion by Intra-Amniotic Instillation

Characteristic	Urea-prostaglandin (n = 2,805) %	Saline Solution (n = 4,778) %
Age (year)		
≤ 19	55.3	49.0
20-24	27.3	31.3
≥ 25	17.4	19.7
Race		
White	54.7	59.6
Black and other	45.3	40.4
Marital status		
Currently married	10.6	10.5
Unmarried	89.4	89.5
Prior pregnancies		
None	56.3	53.7
1 or more	43.7	46.3
Prior abortions		
None	82.4	83.4
1 or more	17.6	16.6

(Binkin et al., 1983)

the analysis of cohort studies are usually demographic characteristics and risk factors for the study subjects in the exposed and unexposed groups (Table 9.11). Additional tables present information about the magnitude of the association between the exposure and the health problem under study and information about the risk of the health problem among subgroups of the exposed and unexposed study subjects.

Analysis Table of Characteristics of Exposed and Unexposed Study Subjects

The analysis should begin with a description of the demographic and medical characteristics of the exposed and unexposed study subjects. These data permit a comparison of the individuals in the exposed and unexposed groups. For example, in Table 9.11 women who underwent abortion by intra-amniotic instillation differed slightly by age, race, and type of instillation (Table 9.11). Otherwise, the two groups of women were similar with respect to marital status and pregnancy history.

Cohort Study Analysis Table for Relative Risk and Confidence Interval

Analysis of data from cohort studies involves the comparison of rates of the health problem among the exposed and unexposed study subjects. Relative risk is the most commonly used measure of association between exposure to a particular factor and risk of specified outcome. Relative risk is the incidence of the outcome among the exposed study subjects divided by the incidence of the outcome among the unexposed study subjects. The relative risk shown in Table 9.12 is known as *cumulative incidence* relative risk (CIR); it measures the risk of the study subject developing the health problem during the entire study period.

Cumulative incidence relative risk (CIR)

For example, pregnant women in Bangladesh who were at least 30 years old had almost two times the risk (CIR = 1.8) of maternal death as pregnant women who were younger than 30 years old (Example 9.13). In the study of the relationship between complications after abortion and the type of instillation procedure, the overall risk of serious complications was more than two times (CIR = 2.3) greater for saline abortions than for urea-prosta-

Table 9.12

Cohort Study Analysis Table for Cumulative Incidence Relative Risk

	Exposed	Unexposed	Total Subjects
Subjects developing outcome	a	b	m_1
Subjects not developing outcome	c	d	m_0
Total subjects	n_1	n_0	t

Cumulative Incidence Relative Risk (CIR) = The proportion of the outcome among the exposed study subjects / The proportion of outcome among the unexposed study subjects

$$= \frac{a/n_1}{b/n_0} \qquad (9.12.1)$$

where *a* is the number of persons with the exposure who have the outcome, *b* is the number of persons without the exposure who have the outcome, *c* is the number of persons with the exposure who do not have the outcome, and *d* is the number with neither the exposure nor the outcome. n_1 represents the total number of persons exposed, and n_0 the total number of persons unexposed. m_1 represents the total number of persons who have the outcome, m_0 represents the total number of persons who do not have the outcome, and *t* is the total number of persons under study.

(Rothman and Boice, 1979)

Example 9.13

Risk Factors for Maternal Mortality in Bangladesh

Results:

Outcome	Maternal Age (years) ≥ 30	< 30
Number of pregnant women who died within 42 days of delivery	18	11
Number of pregnant women still living 42 days after delivery	4,318	4,774
Total	4,336	4,785

$$\text{CIR} = \frac{\text{Incidence of maternal deaths in women} \geq 30 \text{ years}}{\text{Incidence of maternal deaths in women} < 30 \text{ years}}$$

$$= \frac{18/4{,}336}{11/4{,}785}$$

= 1.8 (95% CI: 0.9 - 3.8)

(Adapted from Alauddin, 1986)

Table 9.14

Complication Rates and Additional Treatments Associated with Urea-Prostaglandin and Saline Solution Instillations

Complication or Treatment	Rate* Urea-Prostaglandin	Rate* Saline Solution	Adjusted Relative Risk†	95% Confidence Interval
Serious complications:	1.03	2.18	2.3	(1.4 - 3.6)
Hemorrhage necessitating transfusion	0.32	1.72	5.9	(2.7 - 13.0)
Fever ≥ 38° C for ≥ 3 days	0.71	0.50	0.84	(0.43 - 1.6)

Table 9.14 (continued)

Complication or Treatment	Rate* Urea-Prostaglandin	Rate* Saline Solution	Adjusted Relative Risk†	95% Confidence Interval
Other complications:				
Hemorrhage	6.20	2.20	0.23	(0.17 - 0.30)
Fever ≥ 38° C for ≥ 1 day	6.27	4.96	0.77	(0.61 - 0.96)
Endometritis	5.31	3.26	0.53	(0.41 - 0.69)
Cervical injury	1.32	0.42	0.12	(0.06 - 0.25)
Treatments:				
Uterine evacuation	10.34	14.82	1.4	(1.2 - 1.6)

* Per 100 abortions

† Adjusted for follow-up, previous pregnancies, use of laminaria, prophylactic antibiotics, and use of oxytocin at or before instillation

‡ Numbers too small for logistic regression analysis; using an exact procedure, relative risk = 2.3 with a 95% confidence interval of 0.23 to 120

(Binkin et al., 1983)

glandin instillations (Table 9.14). The risk of hemorrhage requiring transfusion was approximately six times greater (CIR = 5.9).

Person-time

Incidence density relative risk (IDR)

Some studies use person-time as the denominator in the relative risk calculation instead of the number of persons enrolled in the study. This type of relative risk is known as incidence density relative risk (IDR). The person-time denominator simultaneously considers the number of persons under observation and the duration of observation for each person. For example, if 10 persons participate in a study for 15 years, they are said to have contributed 150 (10 persons * 15 years) person-years of observation. The same figure may be derived if 150 persons were under observation for one year or 300 persons for six months. This method allows the investigator to more satisfactorily manage situations when the dates the study subjects enter the cohort study vary, or when, during the course of the study, some study subjects are no longer under observation because of death, loss of contact, or other reasons. The basic analytic table for computing the IDR is presented in Table 9.15 (Rothman and Boice, 1979).

Table 9.15

Cohort Study Analysis Table for Incidence Density Relative Risk

	Exposed	Unexposed	Total
Cases	a	b	m_1
Person-time	n_1	n_0	t

$$IDR = \frac{a / n_1}{b / n_0} \qquad (9.15.1)$$

where *a* is the number of cases among the exposed, *b* is the number of cases among the unexposed, n_1 is the person-time in the exposed group, and n_0 is the person-time in the unexposed group. m_1 represents the person-time among the cases and *t* is the person-time for all study subjects.

(Rothman and Boice, 1979)

The confidence interval for the CIR and the IDR is given in formula 9.15.2.

(9.15.2)

$$\text{Confidence Interval} = RR^{(1 \pm z/\chi)}$$

where RR = CIR or IDR, z is a normal variate, $\chi = \sqrt{\chi^2_{RR}}$, and

(9.15.3)

$$\chi^2_{CIR} = \frac{(t-1) * [(a * d) - (b * c)]^2}{n_1 * n_0 * m_1 * m_0} \qquad \chi^2_{IDR} = \frac{(a - n_1 * m_1 / t)^2}{n_1 * n_0 * m_1 / t^2}$$

(Rothman and Boice, 1979). In Example 9.13, the CIR of 1.8 was not statistically significant since the 95% confidence interval (0.9 - 3.8) includes 1.0. Using the data in Example 9.13, the confidence interval for the CIR is computed by performing the following steps:

Step 1: $$\chi^2 = \frac{9{,}120 * (18 * 4{,}774 - 11 * 4{,}318)^2}{4{,}336 * 4{,}785 * 29 * 9{,}092}$$

$$= 2.46$$

$$\chi = \sqrt{\chi^2}$$

$$= 1.57$$

Step 2: For 95% confidence interval, z = 1.96

Step 3: Lower limit $= CIR^{(1 - z/\chi)}$

$$= e^{[\ln CIR * (1 - z/\chi)]}$$

$$= e^{[\ln 1.8 * (1 - 1.96/1.57)]}$$

$$= 0.9$$

Upper limit $= CIR^{(1 + z/\chi)}$

$$= e^{[\ln 1.8 * (1 + 1.96/1.57)]}$$

$$= 3.8$$

For Example 9.3, the unadjusted IDR for past users compared to never users was 0.8 with 95% confidence interval (0.7 - 0.9). This estimate suggests that women who have used OCs in the past have a statistically significant reduced risk of breast cancer than women who have never used OCs. The estimate is statistically significant since the confidence interval does not include 1.0.

Controlling for Confounding in the Analysis

Confounding is a form of bias that occurs when an extraneous factor related to both the exposure and outcome obscures the true relationship between exposure and outcome. To be a confounder, a variable must be associated with, but not a consequence of, the exposure and independent of its association with the exposure, be associated with the outcome. Stratified analysis is often used to correct for confounding. The relative risk estimate, which has been adjusted by the confounding variable, is computed and compared with the crude estimate. If the difference between the crude and adjusted estimates is more than some percentage specified before analysis,

Table 9.16

Neonatal Outcome by Maternity Clinic or District Hospital

Outcomes of Neonates at Seven days	Maternity Clinic	District Hospital
Died	94	50
Living	11,906	4,950
Total	12,000	5,000
Early neonatal mortality rate	7.8 / 1,000	10.0 / 1,000

CIR = 0.8

then the variable may be considered a confounder and should be controlled for in the analyses. Details of calculating the adjusted relative risk and confidence intervals are introduced in Chapter 4.

In a hypothetical historical cohort study, investigators designed a study to examine whether delivery at the maternity clinic in a certain district is associated with a higher risk of early neonatal mortality than at the district hospital. Records from the maternity clinic and the district hospital are abstracted; the results are shown in Table 9.16. Instead of finding a higher risk of early neonatal mortality associated with delivering in the maternity clinic, the investigator finds the risk is actually lower in the maternity clinic than in the district hospital.

The district has a referral system in which a score system has been developed to identify high-risk pregnancies. Women seen at the maternity clinic for prenatal care are referred to the district hospital for further care and delivery if they have a high maternal risk score. Thus, the investigator decides to examine whether maternal risk score may be confounding the comparison of early neonatal mortality rates at the clinic and the hospital (Table 9.17).

For both strata, the risk of early neonatal mortality is higher for the maternity clinic. The relative risk adjusted for maternal risk score also reflects the higher risk for the maternity clinic. In this example, maternal risk score is a confounder: it is associated with, but not a consequence of, the delivery sites, independent of its

Table 9.17

Neonatal Outcome by Risk Score for Maternity Clinic and District Hospital

Early neonatal outcome	Maternity Clinic	District Hospital
High maternal risk score:		
Died	30	38
Living	820	1,262
Total	850	1,300
Early neonatal mortality rate	35/1,000	29.2/1,000
CIR = 1.2		
Low maternal risk score:		
Died	64	12
Living	11,086	3,688
Total	11,150	3,700
Early neonatal mortality rate	5.7/1,000	3.2/1,000
CIR = 1.8		
Crude CIR = 0.8		
CIR adjusted for maternal risk score = 1.4		

association with delivery site, it is associated with poor early neonatal outcome.

There are essentially three ways to manage potentially confounding variables in a cohort study:

- Restrict admission to the study (e.g., admit only exposed and unexposed women with a maternal risk score lower than some specified level).
- Match on the potentially confounding variable.

- Control for the confounding variable in the analysis by calculating an adjusted measure of effect.

The risk estimate in Table 9.17 after adjusting for the confounding effect of maternal risk score is 1.4; the unadjusted risk is 0.8. Adjusted measures of relative risk can be computed using formula 4.12.1 or 4.12.4.

Effect Modification Analysis Table

Effect modification is a phenomenon that is frequently discussed in conjunction with confounding, although it is not a form of bias. Effect modification is present when the relationship between exposure and outcome is different for various subgroups in the population. Effect modification is detected by stratifying by the variable of interest, calculating a measure of association for each stratum, and looking for differences in the relative risks among strata. Differences may reflect biologic or other factors that can modify the relationship between exposure and outcome. For example, measles vaccine may result in a much lower risk of contracting measles when administered to children older than one year than to younger children. These differences are related to the interference of maternally transmitted antibodies with development of immunity in the infant and have important implications for vaccine policy.

Both confounding and effect modification can occur at the same time. When both are present, the effect modification should be noted, the confounding should be controlled for in the analysis, and the findings should be stratified by the variable that confounds and modifies the association under study.

Consider the analysis of the relationship between anemia and birthweight (Table 9.18). The overall crude CIR is 1.7 and the adjusted CIR is 1.6. These results indicate that maternal age is not a confounder. However, maternal age appears to modify the effect of anemia on low birthweight; young anemic women are twice as likely to have a low-birthweight baby than those without anemia (CIR = 2.2). Among older women, however, anemia is not associated with an increased risk of low birthweight (CIR = 1.0).

Table 9.18

Anemia and Low Birthweight According to Maternal Age

Maternal Age < 25 Years	Anemia	No Anemia
Low Birthweight	13	4
Normal Birthweight	18	17
Total	31	21

CIR = 2.2 (95% CI: 0.8 - 5.8)

Maternal Age ≥ 25 Years	Anemia	No Anemia
Low Birthweight	5	4
Normal Birthweight	37	31
Total	42	35

CIR = 1.0 (95% CI: 0.3 - 3.6)

Advantages and Disadvantages of Cohort Studies

Cohort studies have several advantages:

- They allow a complete description of the individuals' experiences subsequent to exposure, including the natural history of disease.
- They provide a clear temporal sequence of exposure and disease.
- They provide an excellent opportunity to study rare exposures.
- They permit the assessment of multiple outcomes (risks and benefits) that may be related to a specific exposure.
- They permit the direct estimation of the rate of the health problem and the relative risk associated with the exposure of interest.

- They present generally more understandable information to nonepidemiologists.
- They do not require withholding treatment as in a randomized clinical trial.

Disadvantages include the following:

- Large numbers of subjects are required to study rare diseases in cohort study designs.
- Long-term follow-up may be necessary when the latency period for the outcome of interest is long.
- Follow-up may be difficult and loss-to-follow-up may affect the study's results.
- The studies may be relatively expensive to conduct.
- The exposure status, which is present at enrollment into the study, may change during the conduct of the study.

Practice Exercises

1. Consider the following study of the risk of dehiscence among obstetric patients who were not seen at the hospital before delivery.

Example 9.19

The Risk of Dehiscence Among Women Who Were Not Seen at the Hospital Before Delivery

Problem: Is the risk of dehiscence (wound disruption) after cesarean section different for obstetric patients who were *booked* (seen at the hospital one or more times before delivery) or *unbooked* (seen for the first time when in labor)?

Research Hypothesis: Obstetric patients who were unbooked have a greater risk of dehiscence than patients who were booked.

Study Design: Cohort study

Exposed: Women who were not seen at the hospital before having a cesarean section.

Unexposed: Women who were seen at the hospital before having a cesarean section.

Outcome: Dehiscence (i.e., an abdominal wound breakdown involving all layers, including the peritoneum) at any time after cesarean section.

Follow-up: Information on dehiscence was collected at the 6-week follow-up visit.

Data Collection Methods: Data were abstracted from medical records for all cesarean deliveries at a single hospital in Nigeria for a 5-year period ending May 1977.

Results:

	Exposure (Prior Booking)	
	Unbooked	**Booked**
Number of women with dehiscence	39	12
Number of women without dehiscence	1,259	678
Total number of cesarean deliveries	1,298	690

CIR = 1.7 (95% CI: 0.9 - 3.3)

(Chukudebelu and Okafor, 1978)

(a) Considering the 95% confidence interval, is it possible that women who were unbooked have less risk of dehiscence than women who were booked?

(b) What differences between the booked and unbooked patients might account for the results?

2. Consider a hypothetical study of maternal education and the risk of neonatal death.

Example 9.20

Maternal Education and Neonatal Death

Problem: Is maternal education related to neonatal death?

Research Hypothesis: Women with less than six years of schooling have an increased risk of delivering an infant who will die during the neonatal period than do women with at least six years of schooling.

Study Design: Cohort study

Exposed: Women who had had less than six years of formal education.

Unexposed: Women who had had six or more years of formal education.

Outcome: Neonatal death (defined as death occurring between birth and 28 days following birth).

Follow-up: Information on neonatal deaths was collected at the 6-week follow-up visit.

Data Collection Methods: Personal interview data were collected from all women coming to the Maternity Clinic between January 1 and December 31, 1985.

(a) What characteristics might differ between the women in the exposed group and those in the unexposed group?

(b) Which of these characteristics might also be related to neonatal death?

(c) Eighty percent of the women who had had at least six years of schooling returned for follow-up, whereas only 65% of the less educated women returned for follow-up. How might this affect the results? What might be done about this problem?

(d) Construct a data analysis table for computing the cumulative incidence relative risk.

3. Consider an analysis from the Collaboration Review of Sterilization Study.

Example 9.21

Interval Sterilization Procedures by Electrocoagulation and Silastic Bands

Problem: Are interval tubal sterilization procedures (TSP) by electrocoagulation less safe than sterilization using silastic bands?

Research Hypothesis: TSP by electrocoagulation has a greater risk of surgical complications than TSP using silastic bands.

Study Design: Prospective cohort study

Exposed: Women having interval TSP by electrocoagulation.

Unexposed: Women having interval TSP by silastic bands.

Outcome: Unintended major surgery, hemorrhage requiring blood transfusion, febrile morbidity, cardiopulmonary crisis, hospitalization, and death occurring in relation to the tubal sterilization procedure.

Follow-up: Subjects were followed up with a standard questionnaire two to 12 weeks postoperatively.

Data Collection Methods: Data were collected at nine hospitals in the United States from 1978-1980. Information was obtained from standard questionnaires administered to all study subjects and from medical charts.

Data Analysis:

	Exposure	
Outcome	Electrocoagulation	Silastic Band
Surgical Complications	53	12
No Surgical Complications	2,344	1,097
Total Subjects:	2,397	1,109

(Adapted from DeStefano et al., 1983)

(a) Compute the CIR and the 95% confidence interval and interpret the results.

(b) Reverse exposed and unexposed groups in Table 9.21 to obtain Table 9.22; that is, let the exposed be the silastic band group and let the unexposed be the electrocoagulation group. Compute the CIR and the 95% confidence interval and interpret the results.

Table 9.22

Interval Sterilization and Silastic Bands Versus Electrocoagulation

	Exposure	
Outcome	**Silastic Band**	**Electrocoagulation**
Surgical complications	12	53
No complications	1,097	2,397
Total subjects	1,109	2,397

(Adapted from DeStefano et al., 1983)

4. Design a prospective cohort study in outline form based on the following problem situation. Briefly state the problem, define the exposed and unexposed groups, and describe how study subjects for each group will be identified. Specify the measurable outcome and the data collection methods. Develop an analysis table. Describe potential strengths and limitations of your study design.

   Background: Voluntary female sterilization is the most prevalent method of contraception in the world. Approximately 100 million couples use this method, and the demand is expected to increase during this century. One concern is that female sterilization may cause changes in menstrual patterns. For this exercise we will focus on an increase in the severity of dysmenorrhea as our outcome. The level of dysmenorrhea could be subjectively reported or measured in any way you choose. While designing this study, consider previous contraceptive use, the age of study subjects, the reasons for sterilization, potential losses to follow-up, and the appropriate length of follow-up for assessing a change in dysmenorrhea.

   Problem:

   Study design:

   Exposed:

   Unexposed:

   Outcome:

   Follow-up:

Data Collection Methods:

Potential Strengths and Weaknesses:

5. Circle true (T) or false (F).

   (a) T/F In a cohort design, the study groups are identified by exposure status before ascertaining disease status.

   (b) T/F In a historical cohort study, the researcher selects individuals based upon outcome and then determines exposure.

   (c) T/F In a prospective cohort study, the exposure has not happened before the study begins.

   (d) T/F Instead of using an unexposed group, investigators may compare the frequency of an outcome in the exposed study population with that of the general population.

   (e) T/F If a person's exposure status changes after initial assignment to an exposed or unexposed group, the person should remain in the group to which he or she was originally assigned.

   (f) T/F The process by which an outcome is defined needs to be standardized without making staff aware of the hypothesis under study.

   (g) T/F Cohort studies are well adapted for studying rare diseases.

   (h) T/F A 20% loss to follow-up may make study results questionable.

   (i) T/F Relative risk is the incidence rate for those who were exposed to some risk factor divided by the incidence rate for those who were not exposed.

(j) T/F A relative risk greater than 1 occurs when the risk in the unexposed group is greater than in the exposed group.

(k) T/F Information bias is greater in prospective cohort studies than in historical cohort studies.

(l) T/F Frequency matching involves enrolling an exposed individual and then selecting the next unexposed person with characteristics similar to the exposed individual.

6. Consider the hypothetical data presented below:

Table 9.23

Outcome Z and Exposure Q

	Exposure	
Outcome	Exposed to Q	Not exposed to Q
Health problem Z	225	75
Person-years	22,750	24,250

(a) Compute the IDR and the 95% confidence interval.

(b) Interpret the IDR and confidence interval.

Suggested Answers to Practice Exercises

1. The risk of dehiscence among obstetric patients.

   (a) Yes, it is possible that patients who were *unbooked* have less risk than patients who were *booked*, since the lower confidence limit for CIR is 0.9. However, it seems reasonable that patients who were *unbooked* would be at greater risk.

   (b) Possible problems with interpretation include the fact that, besides not receiving antenatal care, the unbooked patients are more likely to represent emergency cases with prolonged labor and have greater potential for preadmission complications. They are also more likely to be from a lower socioeconomic class. Additionally, because they may be more likely to be from rural areas, they may have begun physical labor sooner than women from urban areas and thus may have increased their risk of dehiscence. The same women who are not booked may be less likely to seek medical attention for problems such as wound infections and thus may have increased their risk of dehiscence. These differences could explain the study results.

2. Maternal education and the risk of neonatal death.

   (a) Income, nutritional status, access to health services (including prenatal care, maternity services, and care for the infant) differences in cultural practices affecting infant outcome (e.g., management of the umbilical stump in a way that may increase the risk of neonatal tetanus), intrapartum interval, infant feeding practices, etc.

   (b) Virtually all of the above are also associated with poor infant outcome.

   (c) It may underestimate the risk in the less educated women, if failure to return for follow-up is associated with adverse infant outcome. To correct the problem, arrange active follow-up on both the exposed and unexposed to ensure the highest possible response rate in both groups. Alternatively, select a random sample of those who do not return from each exposure group to ensure that they do not

differ substantially in their outcomes from those who did return for follow-up.

(d)

Data Analysis:

	Lower Education	Higher Education
Neonatal death	a	b
Survival to 28 days	c	d
Total	n_1	n_0

3. Interval sterilization procedures by electrocoagulation and silastic bands.

(a) CIR = 2.2/1.1 = 2 (95% CI: 1.1 - 3.8). In this study, TSP by electrocoagulation appears to be associated with an increased risk (two times) of surgical complications than silastic bands. Since the confidence interval does not include 1.0, the CIR is a statistically significant increased risk; that is, TSP by electrocoagulation appears to be associated with a statistically significant 2-fold increased risk of surgical complications compared with TSP by silastic bands. The data analysis confirmed the research hypothesis.

(b) If you reverse the "exposed" and "unexposed", the relative risk becomes less than one; the cumulative incidences would be 1.1% and 2.2%, and the CIR = 1.1/2.2 = 0.5 (95% CI: 0.3 - 0.9). Since the confidence interval does not include 1.0, this finding is statistically significant. TSP by silastic bands appears to be half as likely as TSP by electrocoagulation to be associated with surgical complications.

4. These are only suggested answers. The exposure groups and outcome you choose will vary, and many responses will be acceptable. The answer provided is based on the report of the actual study referenced.

Example 9.24

Female Sterilization and Dysmenorrhea

Problem: Does dysmenorrhea increase after female sterilization?

Research Hypothesis: More women who have had sterilization surgery report dysmenorrhea after surgery than women who did not undergo sterilization surgery.

Study Design: Cohort study

Exposed: Healthy women aged 25 to 34 years who are at least six months postpartum, are menstruating, and have not used hormonal or intrauterine contraception in the three months before seeking voluntary sterilization.

Unexposed: Healthy women aged 25 to 34 years who are at least six months postpartum, are menstruating, and have not used hormonal or intrauterine contraception in the three months before admission to the study and who are presently using barrier contraceptives.

Outcome: Dysmenorrhea is classified as none, mild, moderate, or severe as subjectively reported by the study subject in response to a standard set of questions.

Follow-up: Two years after the procedure, follow-up data were collected by telephone interviews with each woman.

Data Collection Methods: At admission to the study, subjects reporting dysmenorrhea will be given gynecologic examinations to determine the cause of the condition. If the cause is not idiopathic, the women will be dropped from the study. Each subject will be asked about the characteristics of her most recent menstrual period at admission; one and two years later, each women will be reinterviewed about the characteristics of her menstrual periods over the last 6 months. The interviewer will not know whether the subject is in the exposed or unexposed group.

Data Analysis: The data will be summarized in the following data table.

Example 9.24 (continued)

	Exposure	
	Voluntary	Barrier
Outcome	Sterilization	Methods
Dysmemorrhea		
None		
Mild		
Moderate		

Incidence Density:

Since we are able to establish accurately the time since sterilization for exposed groups and the time from entry into the study for barrier users, we will use incidence density as our incidence measure. The incidence density and confidence limits will be calculated from the data in the table. Finally, we will need to determine if the age and parity distribution differ in the two comparison groups.

Potential Strengths and Weaknesses:

Strengths—Voluntary sterilizations have been separated from medically indicated ones. The effects of prior contraceptive use have been controlled. The evaluator of dysmenorrhea will not know the exposure status of the subjects.

Weaknesses—Level of dysmenorrhea is a subjective outcome. Many subjects may be lost from the unexposed group because they change contraceptives or become pregnant.

(DeStefano et al., 1983)

5. True or false.

(a) T Introduction

(b) F In a historical cohort study, exposure and outcome have occurred before the study begins; as in prospective cohort studies, the subjects are selected according to exposure status rather than outcome.

(c) F In a prospective study, exposure status is determined when the study begins, and the exposed and unexposed groups are followed forward in time to observe whether they develop the study outcome. Often, the exposure has been going on for some time when the study begins.

(d) T Population Sources and Follow-up

(e) T Measuring Exposure

(f) T Determination of Outcome

(g) F Advantages and Disadvantages. Cohort studies are well adapted for studying rare exposures.

(h) T Population Sources and Follow-up

(i) T Relative Risk

(j) F Relative Risk. A relative risk greater than 1 occurs when the risk in the exposed group is greater than the risk in the unexposed group.

(k) F Validity and Bias. Information bias is likely to be larger in historical cohort studies than in prospective cohort studies.

(l) F Matching. Frequency matching involves enrolling exposed individuals and categorizing them into subgroups and then enrolling a specified number of unexposed individuals with the appropriate characteristics for the subgroup.

6. Outcome Z and exposure Q.

6.1 IDR = 3.2 (95% CI: 2.5 - 4.1)

6.2 Women exposed to Q have an approximately threefold increased risk of developing health problem Z as women who have never been exposed to Q. This finding is statistically significant because the 95% confidence interval does not include 1.0.

References

Alauddin M. Maternal mortality in rural Bangladesh: the Tangail District. Stud Fam Plan 1986;17:13-21.

Binkin N, Schulz KF, Grimes DA, Cates WJ. Urea-prostaglandin versus hypertonic saline for instillation abortion. Am J Obstet Gynecol 1983;146:947-952.

Cates WJ, Schulz KF, Grimes DA. The risks associated with teenage abortion. N Engl J Med 1983;309:621-624.

Chukudebelu WO, Okafor EI. Burst abdomen following cesarean section. Ant J Gynecol Obstet 1978;15:532-534.

DeStefano F, Greenspan JR, Dicker RC, Peterson HB, Strauss LT, Rubin GL. Complications of interval laparoscopic tubal sterilization. Obstet Gynecol 1983;61:153-158.

DeStefano F, Huezo CM, Peterson GB, Rubin GL, Layde PM, Ory HW. Menstrual changes after tubal sterilization. Obstet Gynecol 1983;62:673-681.

Dusitin N, Varakamin S, Ningsanon P, Chalapati S, Boonsiri B, Gray R. Post-partum tubal ligation by nurse-midwives and doctors in Thailand. Lancet 1980;1:638-639.

Last, JM (editor). A dictionary of epidemiology. New York: Oxford University Press, 1988.

Mumford SD, Bhiwandiwala PP, Cheng CI. Laparoscopic and minilaparotomy female sterilization compared in 15167 cases. Lancet 1980;2:1066-1070.

Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Bethesda, Maryland: National Institutes of Health; 1979; DHEW publication no. (NIH)79-1649.

Royal College of General Practitioners Oral Contraceptive Study. Mortality among oral contraceptive users. Lancet 1977;2:727-730.

Romieu I, Willett WC, Colditz, GA, et al. Prospective study of oral contraceptive use and risk of breast cancer in women. J Natl Cancer Inst 1989;81:1313-1321.

Chapter 10 - Learning Objectives

After completing this chapter, you should be able to:

1. Define the terms:

   case-control study
   case
   control

2. Describe how to select cases and controls.

3. Define the exposure variable in measurable terms.

4. Construct different measures of exposure.

5. Specify the data collection methods to be used in the study.

6. Describe types of bias.

7. Identify data analysis methods for case-control studies.

8. Describe why the measures of association used for cohort studies do not apply to case-control studies.

9. Construct analysis tables for a case-control study.

10. Calculate and interpret the odds ratio and the confidence interval for the odds ratio.

11. Interpret analysis tables for a case-control study.

12. Recognize advantages and disadvantages of case-control studies.

13. Design a hypothetical case-control study.

10 Case-Control Studies

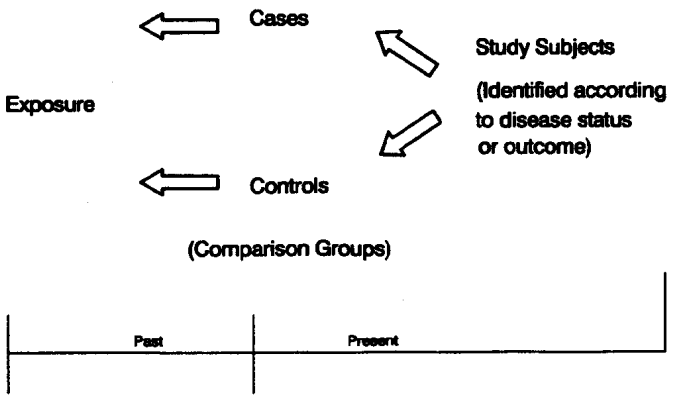
Introduction

Cases and controls defined

The case-control study is an analytic epidemiologic research design in which the study population consists of groups who either have or do not have a particular health problem or outcome. Comparison groups for case-control studies are formed on the basis of disease or health problem status. The study subjects with the health problem are called cases, and the study subjects without the health problem are called controls. In a case-control study we look back in time to measure exposure(s) of the study subjects (Figure 10.1). We compare the exposure among cases to the exposure among the controls to determine if the exposure could account for the health condition of the cases. The research hypotheses should clearly specify the expected relationship between the health problem and the exposure of interest.

Figure 10.1

Definition of Comparison Groups in Case-Control Studies



Comparison with cohort study

The case-control study differs from the cohort study in that individuals are categorized by their health problem or outcome; epidemiologists then look back in time to determine each individual's exposure status. In a cohort study, individuals are categorized by their exposure status; then epidemiologists look forward in time to determine the outcome. In Example 10.2 investigators examined the risk of myocardial infarction among women who were current users of oral contraceptives (OCs) and women who had discontinued using OCs (Slone et al., 1981). The cases were diagnosed with myocardial infarction during 1976 through 1979. The controls were admitted to the same hospitals during the same time period as the cases. The investigators looked back in time to measure exposure to contraceptives. Interviewers collected information about each woman's lifetime use of contraceptives, specifically OCs, before her diagnosis in 1976 through 1979.

Example 10.2

Oral Contraceptive Use and Myocardial Infarction

Problem: Is oral contraceptive (OC) use associated with myocardial infarction?

Research Hypothesis: Current OC use is positively associated with myocardial infarction.

Study Design: Multicenter hospital-based case-control study

Cases: Women who were hospitalized with a diagnosis of myocardial infarction (World Health Organization criteria).

Women 25 to 49 years old who were admitted to 155 hospitals in the northeastern United States between July 1, 1976, and June 30, 1979. Cases who had had a myocardial infarction, coronary-bypass surgery, cardiac-valve prosthesis, rheumatic heart disease, an occurrence of infarction after admission, or sickle-cell crisis preceding infarction were excluded from the study.

Controls: Women with no previous history of myocardial infarction who were admitted to the same hospitals and diagnosed during the same time period as the cases.

Four women, aged 25 to 49 years with diagnoses judged to be unrelated to contraceptive choice were selected at random for each case. The diagnoses

Example 10.2 (cont.)

considered unrelated to contraceptive choice included fractures, dislocations, soft-tissue injuries, gastroenteritis, appendicitis, disk disease, lower-back pain, renal calculus, hiatal hernia, reflux esophagitis, inguinal or femoral hernia, duodenal ulcer, gastric ulcer, and nontoxic thyroid nodules.

Exposure: Current OC use (OC use in the month before hospital admission).

Data Collection Methods: Hospital records were abstracted and cases and controls were interviewed by nurses in the hospital or at home after discharge.

(Slone et al., 1981)

Design and Data Collection Methods

Case Selection

Case defined

The cases are the study subjects who have the health problem or outcome to be investigated. The definition of a case requires two specifications:

- An unambiguous and objective description of the health problem, including how the health problem is to be diagnosed (i.e., diagnostic procedures, laboratory tests, and clinical signs and symptoms),

- The eligibility criteria that will be used to select cases for the study (i.e., ages, no history of selected diseases and conditions, etc., and how, when, and where cases are identified) (Schlesselman, 1982). Cases may be all persons who are newly diagnosed with disease during a specified period of time (also known as incident cases).

Incident cases

Example 10.3 presents the case definition from a case-control study of pelvic inflammatory disease (PID) and current use of an intrauterine device (IUD) (Burkman et al., 1981). Clinically, cases were defined as women who received a hospital discharge diagnosis of PID. For analysis, the case definition was refined to

Example 10.3

Use of Intrauterine Devices and Pelvic Inflammatory Disease

Problem: Is current intrauterine device (IUD) use associated with pelvic inflammatory disease (PID)?

Research Hypothesis: Current IUD use is positively associated with the development of PID.

Study Design: Multicenter hospital-based case-control study

Cases: Women with a hospital discharge diagnosis of PID. The diagnosis of PID was classified as *most certain*, *moderately certain*, and *least certain*. Women with anatomic findings of PID at laparoscopy, laparotomy, culdoscopy, or colpotomy (either the adnexa were visualized or an abscess cavity was drained) were classified as most certain. Women who were hospitalized with a diagnosis of PID for seven days or longer or who had a temperature greater than 100.4 degrees F on two or more days, and who did not undergo the operative procedures listed for the certain cases were classified as moderately certain. Women with a discharge diagnosis of PID who did not fit the criteria for the most certain and moderately certain cases were categorized as least certain.

Women 18 to 44 years old with a discharge diagnosis of PID between October 1976 and August 1978 were identified through a systematic review of admission lists, operative logs, ward census records, and discharge logs at one of 16 hospitals in nine cities in the United States. Women were excluded if they had not been sexually active during the previous year, had had sterilization surgery (or if partner had had sterilization surgery), had been pregnant during the previous six weeks, or had not menstruated in the past year.

Controls: Women who were admitted for acute conditions or elective surgery to the surgical and medical services of the same hospitals during the same time period as the cases.

Controls were selected using the same eligibility criteria as the cases, but they were not diagnosed with PID. Women with diagnoses of gallbladder disease; vascular, pelvic, or breast surgery; and other conditions that could influence contraceptive choice were excluded. Women with diagnoses of epilepsy, thromboembolism, cerebrovascular disease, or cancer were also excluded.

Exposure: Current IUD use (contraceptive method used during the 30 days before hospital admission). Women who had an IUD when admitted to the hospital or who had one removed within 30 days before being admitted to the hospital were classified as current IUD users.

Example 10.3 (Cont.)

Data Collection Methods: Information about the women enrolled in the study was abstracted from medical records, and a standardized interview was conducted during each woman's hospital stay. If a woman was discharged from the hospital before an interview was conducted, she was interviewed by telephone or in her home.

(Burkman et al., 1981)

assess the *certainty of diagnosis*. The investigators based their judgment on whether specific procedures necessary for confirmation of a diagnosis of PID had been performed and on the presence of selected signs and symptoms likely to indicate a diagnosis of PID. The criteria for defining cases of PID in this study were categorized as certain, moderately certain, and least certain.

Eligibility criteria for the cases in this study included a specific age range, sexual activity during the previous year, absence of sterilization surgeries (woman or her partner), menses during the past year, and no pregnancy during the six weeks before admission to the hospital. Study personnel described how, when, and where the cases were identified; admission logs were used to enroll women as potential cases. All women were interviewed before they were discharged from the hospital. Final eligibility was determined after discharge.

Sources of cases

The eligibility criteria identify groups of women who could have experienced the exposure under study (Schlesselman, 1982). Women with conditions that might have limited their choices of contraception were excluded. For example, women who are not sexually active or who have had a tubal ligation are not likely to have recently used any contraceptive method including IUDs.

Good sources of cases include admissions and discharge logs in physician offices, clinics, and hospitals; surgical logs; disease or tumor registries; and for deaths, vital statistics records (Schlesselman, 1982). Potential sources of cases should be evaluated to understand the case identification process, the potential for identifying all cases, and the likely referral patterns for cases. For example, seriously ill cases who live a long distance

from the medical care facility may be unable to travel to the facility, and thus may be missed as potential cases for study. Alternatively, cases from outside the defined study settings may be referred to the medical care facility for treatment and included for study. Failure to consider and manage these possibilities may bias risk estimates.

Control Selection

The controls are the study subjects who do not have the health problem or outcome under investigation. Controls should be similar to the cases with respect to the potential for exposure, since case-control analyses compare the rate of exposure among the cases to the rate of exposure among the controls (Schlesselman, 1982). The criteria used to select controls should be comparable in all ways with the criteria used to select cases, except the controls should not have the health problem under study. If controls are selected in this manner, differences in the rates of exposure are likely to reflect a true association between exposure and disease. If controls are not selected in this manner, then measures of the association between the health problem and the exposure may reflect the differences in the ways the cases and controls were selected and therefore may be biased.

Eligibility criteria

The eligibility criteria for both cases and controls in Example 10.3 were similar and the likelihood that both groups of women were potential users of contraception was increased. Importantly, controls chosen from hospital admission records did not have diagnoses that were related to contraceptive use.

Alternatively, consider a case-control study in which the cases are identified from the discharge logs of a public hospital located in a large urban area. Assume that the controls are selected at random from the general population in the same urban area. These controls would be inappropriate because the general population probably includes many persons who would always use a private instead of a public hospital. These controls are not likely to be comparable to the cases with respect to access to medical care, diseases and conditions, socioeconomic status, education, and other factors.

Sources of controls

The most common sources of control subjects are hospitals, clinics, physicians' offices, the community, or the general population. Selecting controls from the same hospital where cases are selected is often practical and cost efficient; these patients are referred to as hospital controls. However, hospital controls should be selected from

diagnostic categories that are unrelated to the exposure of interest. Controls may also be selected from the community or geographic area where the hospitalized cases reside; these controls are referred to as community or population controls. In some studies, neighbors, associates, friends, or relatives of cases have served as controls. Using the relatives of cases as controls will likely result in matching on variables such as diet, life-style behaviors, and family history of diseases.

Sampling and matching controls

Controls may be selected in several ways. Selection may involve sampling or may include the total population (excluding cases). Controls may be randomly or systematically selected from the total population. Finally, controls may be matched to cases on specific variables to make controlling for confounding in the analysis more efficient. Matching may be desirable if there is an unusual distribution of cases compared to controls with respect to a specific variable (e.g., age). Matching may be individual matching (also known as one-to-one matching) or frequency matching. If individual matching is used, one or more study subjects without disease is matched with a specific study subject who has disease. Alternatively, frequency matching involves matching several study subjects without disease in a given subgroup with study subjects who have disease.

Exposure Definition

Frequency measures

The intensity of the exposure is assessed through measures related to frequency and time. Measures of exposure that relate to frequency may be dichotomous, polytomous, or continuous:

- Dichotomous: categorized as exposed or unexposed (e.g., never and ever diaphragm use)
- Polytomous: measured on more than two levels (e.g., never, occasional, and frequent condom use; number of 8-ounce drinks of alcohol per week; days per week when calcium is taken)
- Continuous: measured on the continuum of a unit of measure (e.g., age in years, parity, birthweight).

Time measures

Exposure may be intermittent (i.e., exposure to a risk factor is interspersed with periods of nonexposure) or continuous (i.e., exposure to a risk factor is constant where a single episode of exposure is not interrupted by exposure to an alternative risk factor or to no risk factor). Measures that relate to time include duration of exposure (e.g., total months of IUD use), time since first exposure (e.g., number of months since first OC use), time since last exposure (e.g., number of months since last pregnancy), and ages at first or last exposure. For some health problems, exposure before, during, or after selected reproductive events or events that are related to the health problem under study is important. For example, McPherson and colleagues (1983) suggested that OC use before the first term pregnancy might increase the risk of breast cancer. Exposure to alcohol, smoking, and various drugs during pregnancy has been related to adverse fetal, neonatal, and infant outcomes. Whether the exposure is current or past may also be important for studying selected health problems or outcomes. In Example 10.3, the definition of current exposure is precisely specified with respect to number of days and includes criteria about recent discontinuations.

Minimum exposure

Since measures of risk are expressed in terms of the exposure, minimum exposure must be defined. The cases and controls with the minimum exposure are the comparison or referent group. Minimum exposure may be defined as no exposure or as some low level of exposure judged to be insufficient to be related to disease.

Sources of exposure data

The primary sources of exposure data are 1) hospital or clinic records, 2) vital statistics, 3) employment, insurance, or social service records, and 4) direct contact with study subjects. Exposure data should be collected from existing documents when possible. When collecting exposure data, the same data collection methods should be used for both cases and controls. If existing sources are unavailable or inadequate, exposure data must be collected by contacting the study subjects directly (in person, by telephone, or by mail).

Data Collection Methods

In-person and telephone interview questionnaires, self-administered questionnaires, and forms designed to abstract information from medical records are used to collect and organize data for study and analysis. Careful form design and wording of the questions affect data quality. Questions used in other studies of the same health

problem or exposure are useful for planning and constructing questions for new studies. Using questions that provided valid data in other studies minimizes the time required to test question wording and promotes comparability of data across studies.

Questionnaire construction

Questionnaire construction should begin with a list of variables needed for study. The list is best developed from a thorough review of the medical literature and should contain all the risk factors for the health problem and the exposure under study. A description of all possible ways to measure each variable should also be included. All data collection forms should be pretested, and data collection personnel (e.g., interviewers, abstractors) should be trained to use the forms correctly.

In Example 10.3, data collection involved medical record abstracts and personal interviews. The interview questionnaire contained questions to confirm the eligibility criteria, demographic questions, questions about factors that might be related to the occurrence of PID or to contraceptive choice, and histories of gynecologic conditions, chronic diseases, contraceptive use, pregnancies and menstrual events. (For additional sources on questionnaire design and data collection methods, see Bennett and Ritchie, 1975, Schlesselman, 1982, and Sudman and Bradburn, 1983.)

Bias

Types of bias defined

In the design of a case-control study, the methods by which study participants are selected and classified as diseased, not diseased, exposed, and unexposed affect validity (Schlesselman, 1982). Errors in design may cause the overrepresentation or underrepresentation of study participants into these categories and distort the measure of the association between the health problem and the exposure. Bias refers to the systematic errors that produce an inaccurate estimate of the association between the health problem and the exposure. Three types of bias can influence the results of the case-control investigation: selection bias, information bias, and confounding bias.

Selection bias. This type of bias refers to the process in which cases or controls are selected in a way that is related to exposure. Biased estimates of the association between the exposure and the

health problem can occur when the exposure is related to nonresponse, length of stay, survival, differential surveillance, diagnosis, referral, or selection of study participants.

Nonresponse bias

Nonresponse bias is a type of selection bias that refers to the respondent's refusal or inability to participate in the study or to the field personnel's inability to contact potential study participants. Nonresponse bias occurs when the rate of exposure among nonparticipating cases is different from the rate of exposure among nonparticipating controls. While the study is being conducted, every effort should be made to locate study participants, to enlist their participation, and to minimize refusals. Interviewers may need training on how to locate difficult-to-find study participants. When refusal rates are high, several actions are needed to identify ways to reduce nonparticipation. Introductions and information about the study that is provided to study participants may need to be revised. Interviewers with high refusal rates may need specialized training about how to persuade reluctant participants to cooperate. Convening regular meetings for interviewers permits exchange of ideas about locating difficult-to-find participants and eliciting cooperation.

Refusals

However, if nonresponse is high at the conclusion of the study, some measure of the possible effects of nonparticipation on the results is needed. The investigator may perform a worst-case analysis in which all nonparticipating cases are assumed to be exposed and all nonparticipating controls are assumed to be unexposed. If the conclusions about the association between the health problem under study are unchanged with the worst-case analysis, nonparticipants are unlikely to have affected the results. The investigator may also compare characteristics of the participating and nonparticipating cases and controls.

Length of hospital stay bias

Bias related to length of hospital stay, another type of selection bias, may occur if cases are selected from a registry of current hospital patients instead of from admission or discharge logs. If cases are selected from a registry of current hospital patients, then cases who have been hospitalized for the longest period of time have a higher probability of being selected than cases admitted for minor conditions or cases who died. Furthermore, these cases may have other diseases and conditions that may be related to the disease or exposure under study. Therefore, use of admission or discharge logs for identifying potential study participants for hospital-based studies is preferable.

Survival and surveillance bias

Two additional types of selection bias are survival bias and surveillance bias. Bias related to survival may occur if only the survivors of the outcome are selected as cases and if survival is related to the exposure of interest. Surveillance bias may occur when health problems that are mild or asymptomatic are diagnosed as a result of more frequent or thorough follow-up examinations and when more frequent or thorough examinations are conducted among study participants who have been subjected to suspected exposures. To assess whether surveillance bias is present in the data, cases and controls with frequent medical surveillance can be analyzed separately from cases and controls with infrequent surveillance.

Diagnostic and referral bias

Diagnostic bias and referral bias are also types of selection bias. Diagnostic bias may occur if knowledge of the exposure status inappropriately alters the diagnosis. Assuming that diagnostic bias decreases as the severity of disease or the certainty of diagnosis increases, analysis of cases according to the certainty of diagnosis permits some assessment of the presence of this type of bias. Referral bias may occur if knowledge of the exposure status or if a variable (e.g., socioeconomic status) related to exposure inappropriately affects referral patterns. Referral bias is most likely to occur when a study is based in the hospital, clinic, or physician's office.

Potential biases introduced in selecting cases and controls may be reduced by designing the study so that selections are not necessary. The ideal study design involves the enrollment of all cases of a disease occurring in a defined geographic region during a specified period of time and the random selection of controls from the general population of the same area. Exclusions that are applied to the cases must also be applied to the controls. Since the purpose of the control group is to determine the rate of exposure expected in the case group, if no association between the exposure and the health problem is present, the controls should be comparable with the cases in all relevant ways except that they do not have the health problem.

Example 10.4 describes a population-based case-control study. The cases were all women diagnosed with histologically confirmed primary breast cancer who resided in one of eight geographic regions in the United States (The Cancer and Steroid Hormone Study, 1986). These women were identified through population-

Example 10.4

Oral Contraceptive Use and Breast Cancer

Problem: Is OC use associated with the risk of breast cancer?

Research Hypothesis: There is an association between OC use and breast cancer.

Study Design: Multicenter population-based case-control study

Cases: All women 20 to 54 years old with newly diagnosed breast cancer who reside in one of eight geographic regions in the United States. The women were identified through population-based tumor registries.

Controls: Women of the same ages selected at random from the general population of the same eight geographic regions during the same time period when cases were diagnosed.

Exposure: OC use before cancer diagnosis for cases and before interview for controls.

(The Cancer and Steroid Hormone Study, 1986)

based tumor registries, which are agencies that collect information on all new cases of cancer diagnosed in the region. The controls were selected at random from the general population of the same geographic regions as the cases. Since the controls were selected from the general population, the data on their contraceptive history will likely provide an accurate measure of the expected rate of exposure to oral contraceptives in these geographic regions. Although population-based case-control studies are ideal, they may not always be feasible. Identifying all case patients may be difficult if they do not routinely seek medical attention for the disease under study. Identifying controls from the general population may be expensive or logistically impossible. Furthermore, not all hospitals in the specified geographic regions or metropolitan areas will necessarily consent to participate in the study.

When studying certain mild conditions or outcomes for which only selected patients, such as more affluent patients, would be hospitalized, use of neighborhood or general population controls may pose problems that relate to selection bias. (Under these

circumstances, general population controls may not be comparable to cases, since controls with middle to lower socioeconomic status may never be hospitalized for the mild condition.) Preferably, the controls should be restricted to individuals who would likely have been hospitalized if they developed the condition under study.

Hospital-based case-control studies maximize similarities between cases and controls. In the hospital-based case-control study, controls should be selected from patients admitted to the same hospitals for other minor conditions that are known not to be associated with the exposure. If such measures are taken, cases and controls are more likely to have similar socioeconomic status, life-style behaviors, religion, and other traits that can determine patterns of hospitalization. Although this selection procedure promotes comparability between cases and controls, the results may not be generalizable to the entire population.

If the control group is chosen from hospitalized patients, it seems best to include individuals with many different illnesses, none of which is known to be associated with the exposure being studied. In this way, even if one of the diseases is found to be associated with the exposure under study, it will likely have little effect on the study results since the number of patients with one specific illness is small (Example 10.5).

Example 10.5

Selection Bias in Hospital-Based Studies

One of the first case-control studies of OC use and cardiovascular disease included many women with gallbladder disease in the control group. Shortly after the study was concluded, several reports indicated that women using OCs were at increased risk of gallbladder disease. Thus the control group contained women who had developed gallbladder disease because they had used OCs. As a result, a spuriously high rate of exposure to oral contraceptives was noted among the controls, and the study artificially underestimated the relative risk of cardiovascular disease in oral contraceptive users.

(Boston Collaborative Drug Surveillance Program, 1973)

Information bias. This type of bias refers to the collection of incorrect information about exposure that results in an incorrect measurement of the exposure. Obtaining an accurate exposure history from the cases and controls is one of the major difficulties in conducting a valid case-control study. Participants' recall of past exposures may be inconsistent and inaccurate. Exposure information may be difficult to remember, particularly when the exposures occurred in the distant past. To improve recall, memory aids such as calendars, diaries, photographs, or other visual materials may be helpful. Exposure information should be validated whenever possible, using sources independent of the study subject's report (Example 10.6).

Example 10.6

The Use of Memory Aids to Reduce Information Bias

In studies of the association between oral contraceptive use and reproductive cancers, accurate information on use of specific formulations of oral contraceptives and use in the distant past was essential. Researchers in Great Britain designed a study to compare self-reported oral contraceptive histories collected during personal interview with data collected prospectively from the Oxford Family Planning Association cohort study. Two memory aides were used to assist recall during interview: a calendar of important life events to which the study participant might relate her use of oral contraceptives, and a book of color photographs of all oral contraceptive preparations marketed in Great Britain. Interview data collected with the calendar and photographs consistently had better agreement with the cohort records data than interview data collected without the aids. Agreement between self-reported data and cohort records within six months was 56.2% for total duration, 81.2% for date of first use, 66.7% for date of last use, and 79.2% for date of use prior to first term pregnancy. For interviews conducted without the memory aids, the rate for total duration was 39.2%, for date of first use, 80.4%, for date of last use, 52.9%, and for date of use prior to first term pregnancy, 70.6%.

(Coulter et al., 1986)

Recall bias

More important than forgetfulness, however, is whether cases remember exposures to risk factors in the past differently than controls. Recall bias refers to the effect of cases remembering exposures differently than controls. In some case-control studies,

recall bias may be of such concern that it influences the choice of a control group (Example 10.7).

Example 10.7

The Effect of Potential Recall Bias on the Choice of Controls

Investigators of congenital malformations seek information on exposures early in pregnancy. Because of the psychological trauma of giving birth to a child with defects, mothers of cases may be more likely than mothers of controls to recall exposures to potential risk factors during pregnancy; they may, in fact, search for an explanation of the malformation. This enhancement of recall could result in bias, and spuriously create a difference between the cases and the controls in their reported exposure histories. Recall bias in studies of malformations may be minimized by choosing control infants with different birth defects than the birth defect of interest.

In a case-control study of the relationship between birth defects and maternal consumption of caffeine, six groups of case patients (inguinal hernia, cleft lip, cardiac defect excluding murmur, pyloric stenosis, isolated cleft palate, and neural tube fusion defect) were compared to all other malformed infants. Mothers of malformed infants who did not have one of the six case defects would likely have similarly enhanced recall as mothers of the case patients.

(Rosenberg et al., 1982)

A similar rationale has been used in the defense of hospital-based case-control studies. Since the controls in hospital-based studies are also ill, they would be expected to have a similar enhancement of memory as the cases.

Confounding bias. This bias refers to the effect of an extraneous factor that distorts an apparent association between the health problem and the exposure under study or that obscures an underlying true association (Schlesselman, 1982). That is, the association between an exposure and the health problem under study may actually be due to another variable. Alternatively, the lack of an association could result from failure to control for the effect of some extraneous factor. To be a confounder, a variable must be associated with, but not a consequence of, the exposure, and be a risk factor for the health problem under study

(Schlesselman, 1982). Confounding may be controlled in the data analysis phase if the relevant information has been collected.

Data Analysis Methods

Analyses should be planned concurrently with designing the study protocol and data collection form or questionnaire to ensure that information is collected for all important variables in a manner that is appropriate for analysis. The first steps of the analysis begin with organizing the data into tables that are used to make comparisons between cases and controls. Typical tables presented for the analysis of case-control studies include characteristics (demographics and risk factors) of the cases and controls (Table 10.8), information about the magnitude of the association between the health problem and the exposure under study (Tables 10.14 and 10.16), and information about the risk of the health problem according to various subgroups of the cases and controls (Table 10.18).

Analysis Table of Characteristics of Cases and Controls

Careful examination of differences of various characteristics between cases and controls provides important information about the comparability between the case patients and the controls. First, these data permit comparisons with studies published in the literature. The risk factors present in the data under study can be compared with known risk factors in other studies. When known risk factors are not present in the data under study, the researcher should evaluate and question his data collection procedures and findings. Second, analysis of known risk factors may provide information about which variables will potentially confound an association between the health problem and the exposure under study. Third, evaluation of similarities or dissimilarities between the cases and controls may provide information that may indicate if cases and controls were selected or interviewed in a comparable manner. For example, in multicenter studies, are the geographic, hospital, or clinic distributions of the cases and controls similar?

Table 10.8 presents selected percentage distributions of demographic characteristics and risk factors for cases and controls enrolled in the case-control study of oral contraceptive use and

Table 10.8

Characteristics Women with Epithelial Ovarian Cancer and Controls

Characteristics	Cases (n=492) %	Controls (n=4,228) %
Age (years)		
20-29	11.0	6.5
30-39	19.9	23.4
40-49	35.8	42.0
50-54	33.3	28.1
Parity		
Nulliparous	25.2	14.8
1	14.4	11.0
2	26.2	24.2
3	19.1	21.8
4+	14.6	27.7
Unknown	0.4	0.5
Obesity [(Weight (g)/Height^2(cm)]		
<2.25	48.8	56.0
2.25-2.49	26.2	21.5
2.50+	24.8	22.1
Unknown	0.2	0.4

(The Cancer and Steroid Hormone Study, 1987)

epithelial ovarian cancer (The Cancer and Steroid Hormone Study, 1987). Cases were more likely than controls to be younger, have lower parity, and be obese.

The Odds Ratio as an Estimate of the Relative Risk

Rare disease assumption

Data collected from case-control studies do not permit a direct calculation of the relative risk. Because the proportion of the population that the cases and the controls represent is usually not known, we cannot derive incidence rates in the exposed and unexposed populations. Instead, the odds ratio is used to estimate

the relative risk. When the health problem is rare among the exposed and unexposed persons in the general population, the odds ratio closely approximates the relative risk. Historically, case-control studies have been the preferred study design when the incidence of the study outcome is rare, such as in cancer studies. Example 10.9 and Table 10.10 permit the comparison between the use of the relative risk in a hypothetical cohort study and the use of the odds ratio in a case-control study with similar hypotheses.

Example 10.9

Oral Contraceptive Use and Cervical Intraepithelial Neoplasia Cohort Study

Problem: Is OC use related to the risk of cervical intraepithelial neoplasia (CIN)?

Research Hypothesis: OCs are associated with the risk of CIN.

Study Design: Multicenter cohort study

Exposed: Women 18 to 44 years old who were using OCs when they visited any of 20 family planning clinics.

Not Exposed: Women of the same ages who were not using OCs when they visited the same family planning clinics.

Outcome: Exposed and unexposed women were followed for one year to observe occurrences of biopsy-proven CIN.

Data Collection Methods: Clinic records for the current visit were abstracted for each visit during one-year follow-up.

Clinic records indicated that the one-year incidence of cervical dysplasia among the exposed women would be 5% and unexposed women would be 2%. The cumulative incidence relative risk (Table 10.10) was:

$$CIR = \frac{a/n_1}{b/n_0} = \frac{100/2{,}000}{40/2{,}000} = 2.5$$

Table 10.10

Analysis Table

Cohort Study

	OC Use	
	Yes	No
Cervical Dysplasia		
Yes	a 100	b 40
No	c 1,900	d 1,960
Total	n_1 2,000	n_0 2,000

Case-Control Study

	OC Use		
	Yes	No	Total
Cervical Dysplasia			
Yes (Case)	100	40	140
No (Control)	69	71	140

If this problem had been investigated in a case-control study (Table 10.10), the 140 cases of cervical dysplasia would constitute the group of cases. An equal number of controls would have been selected at random from the 1,900 + 1,960=3,860 women who did not have cervical dysplasia (69 = 1,900 * 140/3,860 controls who had used OCs, and 71 = 1,960 * 140/3,860 controls who had never used OCs). Information about exposure to oral contraceptives would have been collected through in-person interviews with the study participants.

The odds ratio was:

$$OR = \frac{a/b}{c/d} = \frac{100/40}{69/71} = \frac{100 * 71}{69 * 40} = 2.6$$

Because the incidence of cervical dysplasia was low in the general population, the case-control study yielded an odds ratio of 2.6. This ratio is approximately equal to the cohort study cumulative incidence relative risk of 2.5. However, the case-control design required a substantially smaller study size (280) than the cohort study (4,000).

Alternatively, if the incidence of the health problem under study is high, the odds ratio will be an inadequate estimate of the relative risk (Example 10.11). In the cohort design, the proportion of the exposed with the outcome (1,000/2,000) is the same as the proportion of cases exposed in the case-control study (1,000/2,000). A similar relationship exists among the unexposed women in the cohort study (1,000/10,000) and the controls who were exposed in the case-control study (200/2,000). The unbiased estimate of the cumulative incidence relative risk from the cohort study is CIR=5.0. Because the incidence of the outcome was high in the general population the estimate of relative risk from the odds ratio, OR=9.0, is a poor estimate.

Case-Control Study Analysis Table for the Odds Ratio and Confidence Interval

Odds ratio

In case-control studies, the odds ratio (OR) estimates the relative risk as a measure of the magnitude of the association between the exposure and the health problem under study. Table 10.12 presents the analysis table for the odds ratio. Table 10.12 is similar to the tables used to estimate the relative risk (RR) and the incidence density ratio in cohort studies and randomized clinical trials. However, the risk measures applicable for cohort studies cannot be calculated from data gathered by the case-control design because, in a case-control study, the health problem or outcome is sampled, not the exposure. The ratio of risks in a case-control study may be estimated *by dividing the odds of disease among the exposed (a/c) by the odds of disease among the unexposed (b/d)* (See 10.12.1). An OR larger than 1.0 implies that the risk of the health problem is increased if the study participant was exposed. An OR smaller than 1.0 implies protection as a result of the exposure (i.e., the risk of the health problem is decreased if the study participant was exposed).

Example 10.11

High Incidence Rates
Analysis Table

Cohort Study

Outcome	Exposure: Yes	Exposure: No
Yes	1,000	1,000
No	1,000	9,000
Total	2,000	10,000

$$CIR = \frac{1{,}000/2{,}000}{1{,}000/10{,}000} = \frac{.5}{.1} = 5.0$$

Case-Control Study

Outcome	Exposure: Yes	Exposure: No	Total
Yes (Case)	1,000	1,000	2,000
No (Control)	200	1,800	2,000
Total	1,200	2,800	

$$OR = \frac{1{,}000 * 1{,}800}{1{,}000 * 200} = 9.0$$

Table 10.12

Case-Control Study Analysis Table for Estimating the Odds Ratio

	Exposure		
Outcome	Number Exposed	Number Not Exposed	Total
Outcome Present (Cases)	a	b	m_1
Outcome Absent (Controls)	c	d	m_0
Total	n_1	n_0	t

$$\text{Odds Ratio (OR)} = \frac{\text{The Odds of Disease Among the Exposed}}{\text{The Odds of Disease Among the Unexposed}}$$

(10.12.1)

$$= \frac{a/c}{b/d} = \frac{a*d}{b*c}$$

Confidence interval for odds ratio

The confidence interval (CI) for the odds ratio (Mantel and Haenszel, 1959, Miettinen, 1976) is given by the formula 10.12.2.

(10.12.2)

$$\text{Confidence Interval for OR} = OR^{(1 \pm z/\chi)}$$

where z is a normal variate, and

(10.12.3)

$$\chi = \sqrt{\chi^2_{MH}}, \text{ and } \chi^2_{MH} = \frac{(t-1) * (a * d - b * c)^2}{n_1 * n_0 * m_1 * m_0}$$

This confidence interval provides a good estimate when the OR is close to 1.0 but becomes less stable for ORs greater than, say 3. Table 10.13 presents the risk of epithelial ovarian cancer among

women who have ever used oral contraceptives (The Cancer and Steroid Hormone Study, 1987). Women who have ever used OCs have a 40 percent [(1 − 0.6)*100] decreased risk of developing epithelial ovarian cancer than women who have never used OCs. Oral contraceptive use protects against ovarian cancer.

Using the data in Table 10.13, the confidence interval for the odds ratio is computed by performing the following steps:

Step 1:

$$\chi^2_{MH} = \frac{(t-1) * (a*d - b*c)^2}{n_1 * n_0 * m_1 * m_0}$$

$$= 4{,}719 * \frac{(250 * 1{,}532 - 242 * 2{,}696)^2}{2{,}946 * 1{,}774 * 492 * 4{,}228}$$

$$= 31.51$$

$$\chi_{MH} = \sqrt{\chi^2_{MH}}$$

$$= 5.61$$

Step 2: For 95% confidence interval, z = 1.96

Step 3:

$$\text{Lower limit} = OR^{(1 - z/\chi)}$$

$$= e^{[\ln OR * (1 - z/\chi)]}$$

$$= e^{[\ln 0.6 * (1 - 1.96/5.61)]}$$

$$= 0.5$$

$$\text{Upper limit} = OR^{(1 + z/\chi)}$$

$$= e^{[\ln OR * 1 + z/\chi)]}$$

$$= e^{[\ln 0.6 * (1 + 1.96/5.61)]}$$

$$= 0.7$$

Because the confidence interval does not include 1.0 in Table 10.13, the odds ratio is considered statistically significant. That is, OC use protects against ovarian cancer, and the estimate of the relative risk is significantly less than 1.0.

Table 10.13

Oral Contraceptive Use and Epithelial Ovarian Cancer

Epithelial Ovarian Cancer	Ever OC Use		
	Yes	No	Total
Yes	250	242	492
No	2,696	1,532	4,228
Total	2,946	1,774	

OR = 0.6 (95% CI: 0.5 - 0.7)

(The Cancer and Steroid Hormone Study, 1987)

In Example 10.14, compared with women who have no history of legal abortion, women with a previous legal abortion have a relative risk of 1.4 of having placenta previa in a later pregnancy (Grimes and Techman, 1984). In this analysis, because the confidence interval includes 1.0, the investigators concluded that legal abortion appeared to have little effect on the development of placenta previa in a later pregnancy.

Evaluating some frequency or time-related measure of the exposure may be important for some health problems. That is, does the risk of the health problem differ for different levels of exposure? Does the risk of the health problem increase (decrease) with increasing (decreasing) levels of the exposure? Table 10.15 presents the risk of ovarian cancer according to years duration of OC use (The Cancer and Steroid Hormone Study, 1987). An odds ratio is computed for each duration grouping of years of OC use, with never users as the referent group. The risk of ovarian cancer decreased from 0.6 among women who use OCs for less than one year to 0.3 among women who use OCs for at least 10 years.

Other possible measures for evaluating the association between ovarian cancer and OC use might include years since first (last) use of OCs, micrograms of estrogen (milligrams of progestin) in the oral contraceptive used, etc.

Example 10.14

Legal Abortion and the Risk of Placenta Previa

Problem: Is legal abortion associated with placenta previa in a later pregnancy?

Research Hypothesis: Legal abortion increases the risk of placenta previa in a later pregnancy?

Study Design: Hospital-based case-control study

Cases: Women who gave birth to infants weighing >499 g at Grady Hospital (Atlanta, Georgia) between January 1, 1975 and December 31, 1979. The women were diagnosed either with complete placenta previa (documented by ultrasonography) that necessitated cesarean delivery of the infant or with placenta previa (documented by ultrasonography) that led to bleeding and required hospitalization during pregnancy.

Controls: Women without placenta previa who delivered infants weighing >499 g at Grady Hospital during the study period; they were randomly selected.

Exposure: A history of one or more previously induced abortions.

Data Collection Methods: At the first prenatal visit, information about the cases and controls was collected through personal interview; and abstracts of computerized medical records and charts were used to collect information for the remainder of the pregnancy. Data obtained at each interview was validated, when possible, through linkage of family planning clinic records.

Results:

	Exposure	
Outcome	Previous Legal Abortion	No Previous Legal Abortion
Placenta Previa		
Yes	12	56
No	9	59

OR = 1.4 (95% CI: 0.5 - 3.6)

(Grimes and Techman, 1984)

Table 10.15

The Risk of Epithelial Ovarian Cancer by Duration of Oral Contraceptive Use

Duration of Use (Years)	Cases	Controls	Crude Odds Ratio	95% Confidence Interval
Never	242	1,532	1.0	Referent
< 1	40	414	0.6	(0.4 - 0.9)
1-2	65	602	0.7	(0.5 - 0.9)
3-4	40	397	0.6	(0.4 - 0.9)
5-9	39	594	0.4	(0.3 - 0.6)
≥ 10	13	328	0.3	(0.1 - 0.4)

(The Cancer and Steroid Hormone Study, 1987)

Controlling for Confounding in the Analysis

Unlike selection and information bias, we can adjust for confounding in the analysis phase. During data collection, we should collect information on potentially confounding variables so that they can be addressed during analysis. To control for confounding, we arrange the data in strata according to levels of the potentially confounding variable. Then, the Mantel-Haenszel method may be used to compute a summary (adjusted) relative risk and summary χ^2 (Mantel and Haenszel, 1959). The adjusted relative risk is a weighted relative risk that takes into account the relative risk estimates for each stratum.

To determine whether or not a variable confounds the association between the outcome and the exposure, the crude (unadjusted) relative risk (Formula 10.12.1) is compared with the adjusted relative risk. If the adjusted relative risk estimate differs from the crude relative risk by more than some previously specified percentage, the variable may be considered a confounder. Under these circumstances, the adjusted relative risk estimate should be presented instead of the crude estimate.

In Table 10.16, the relative risk adjusted for age (3.1) differs from the crude relative risk (2.4) by 29%. This result suggests that

Table 10.16

Current Oral Contraceptive Use and Myocardial Infarction
Adjustment for Confounding by Age

Age		OC Use		OR	95% CI
		Yes	No		
<40	Cases	21	26		
	Controls	17	59	2.8	(1.3 - 6.1)
40-44	Cases	8	44		
	Controls	2	50	4.5	(1.0 - 20.3)
Total	Cases	29	70		
	Controls	19	109		

Crude RR = 2.4 (95% CI: 1.2 - 4.5)

RR_{MH} = 3.1 (95% CI: 1.6 - 6.3)

(Hogue et al., 1985)

age is a confounding variable in this study. Because age confounds the association between current OC use and the risk of myocardial infarction, the adjusted relative risk (3.1) should be reported. Current OC use is associated with a statistically significant increased risk of myocardial infarction; overall, women who currently use OCs have three times the risk of myocardial infarction as women who are not currently using OCs.

Effect Modification Analysis Table

Some health problems or outcomes require analyses of subgroups of study participants. That is, does the risk of the health problem for the exposure of interest vary for different ages or is the risk greater for some racial or ethnic groups? Do women with a particular characteristic have a different risk than women without the characteristic? When the association between a health problem and an exposure varies according to categorizations of a particular variable, that variable is considered an effect modifier. In Table 10.16, age is not only a confounder but also an effect modifier. The relative risks differ according to age. Among women who are 40 to 44 years old, current OC users have 4.5 times the risk of a myocardial infarction as women of the same ages who were not currently using OCs. Among younger women the risk is 2.8.

Table 10.17 presents the risk of developing epithelial ovarian cancer among women who have used OCs and women who have never used OCs, by age and parity (Cancer and Steroid Hormone Study, 1987). The risk of epithelial ovarian cancer does not appear to vary by age among women who had ever used OCs and women who had never used OCs. However, the risk of epithelial ovarian cancer varied slightly for different levels of parity. For example, nulliparous women had a risk of 0.7, women with parity 1 to 4 had a statistically significant reduced relative risks of 0.4 or 0.5, and women with parity 5 or more had a relative risk of 1.2.

Although effect modification seems similar to confounding, the two are very different. A variable can be an effect modifier of the association between the outcome and an exposure, a confounder of the association, both, or neither. When effect modification appears to be present, Woolf's χ^2 test for heterogeneity may be used to test whether the variations in stratum-specific relative risks are real or due to chance (Woolf, 1955).

Advantages and Disadvantages

Advantages:

- Case-control studies are useful for studying health problems that occur infrequently.

Table 10.17

Oral Contraceptive Use and the Risk of Epithelial Ovarian Cancer by Age and Parity

Characteristic	Cases	Controls	Odds Ratio*	95% Confidence Interval
Age (years)				
<30	15/ 32†	53/206†	0.5	(0.3 - 1.1)
30 - 34	114/ 40	695/414	0.6	(0.4 - 0.9)
35 - 39	61/ 29	454/515	0.4	(0.3 - 0.7)
40 - 44	33/ 33	199/474	0.4	(0.3 - 0.7)
45 - 49	10/ 33	81/462	0.6	(0.3 - 1.2)
50 - 54	9/ 38	50/351	0.6	(0.3 - 1.3)
Parity				
Nulliparous	63/ 54	231/302	0.7	(0.4 - 1.0)
1	30/ 34	135/292	0.5	(0.3 - 0.9)
2	60/ 53	324/672	0.4	(0.3 - 0.6)
3	52/ 37	305/578	0.4	(0.2 - 0.6)
4	22/ 10	245/294	0.4	(0.2 - 0.8)
5+	15/ 17	292/284	1.2	(0.6 - 2.4)

*In each stratum, ever users are compared to never users.
†Did not use/used oral contraceptives.

(The Cancer and Steroid Hormone Study, 1987)

- Case-control studies are useful for studying health problems with a long latent interval.
- The relatively short study period required usually makes case-control studies less time consuming and less expensive than cohort studies.
- Case-control studies are useful for characterizing the effects of a variety of potential risk factors on the health problem under study.

Disadvantages:

- Because cases and controls may be selected from two separate populations, it is difficult to ensure they are comparable with respect to extraneous risk factors and other sources of bias.
- Exposure data are collected from records or by recall after the disease has occurred. Records may be incomplete, and recall of past events is subject to human error and the possibility of selective recall.
- Case-control studies cannot be used to determine incidence rates.
- If the health problem is relatively common in the population (i.e., > 5%-10%), the odds ratio is not a reliable estimate of the relative risk.
- Case-control studies cannot be used to determine the other possible health effects of an exposure. By definition, case-control studies are concerned with only one outcome.

Practice Exercises

1. In Example 10.3, investigators examined the association between IUDs and PID. Using the information presented in Table 10.18, analyze and interpret the data.

Table 10.18

The Risk of Pelvic Inflammatory Disease Among Women Who Use Intrauterine Devices and Women Who Use No Method of Contraception

	IUD Use	
PID	Current IUD User	No Method Currently
Yes	841	724
No	518	967

(Burkman et al., 1981)

(a) Compute the odds ratio and the 95% confidence interval for this study.

(b) Interpret the odds ratio and the confidence interval.

(c) What are some questions the investigators should consider regarding the choice of the control group?

(d) What are some other potential risk factors for PID that need to be considered in the analysis?

2. Example 10.19 presents a case-control study designed to investigate whether early age at first coitus is a risk factor for cervical cancer.

Example 10.19

Early Age at First Coitus and Cervical Cancer

Problem: Is age at first coitus associated with cervical cancer?

Research Hypothesis: Age ≤ 15 years at first coitus is positively associated with subsequent development of cervical cancer.

Study Design: Hospital-based case-control study

Cases: Women of any age with a histologic diagnosis of invasive cancer of cervix at one hospital.

Controls: Women of any age who were healthy and attended the hospital's family planning clinic.

Exposure: The woman's report that her age at first coitus was ≤ 15 years.

Data Collection Methods: Questionnaires were administered to cases and controls.

Results:

	Age at First Coitus	
Cervical Cancer	**≤ 15 years**	**>15 years**
Yes (Cases)	36	78
No (Controls)	11	95

Crude OR = 4.0 (95% CI: 2.0 - 8.1)

(Andolusi, 1977)

(a) Interpret the odds ratio and confidence interval.

3. Example 10.20 presents a study of OC use and the risk of PID.

Example 10.20

Current Oral Contraceptive Use and Pelvic Inflammatory Disease

Problem: Is current use of OCs associated with pelvic inflammatory disease (PID)?

Research Hypothesis: Current OC use is associated with PID.

Study Design: Hospital-based case-control study

Cases: Women 18 to 44 years old who were admitted to nine hospitals in the United States with an initial episode of PID.

Controls: Women 18 to 44 years old with no history of PID who were admitted to the same hospitals as the case patients but had acute conditions or elective procedures not related to PID.

Women in either group who reported sterility, recent pregnancy, or lack of sexual activity, as well as women with conditions that might contraindicate OC use, were excluded from the study.

Exposure: The woman's report of the contraceptive method used in the three months prior to interview.

Data Collection Methods: Standard questionnaires were administered to both cases and controls.

Results:

	Exposure		
Outcome	Current OC User	No Method	Total
PID			
Yes	139	170	309
No	831	558	1,389
Total	970	728	

(Rubin et al., 1982)

(a) Compute the odds ratio and the 95 % confidence interval.

(b) Interpret the odds ratio and confidence interval.

4. In Example 10.21, the description of a study of depo-medroxyprogesterone acetate (DMPA) and the risk of breast cancer is presented:

Example 10.21

DMPA Use and Breast Cancer

Problem: Is DMPA use associated with the risk of breast cancer?

Research Hypothesis: DMPA is associated with the risk of breast cancer.

Study Design: Case-control study

Cases: Women 25 to 58 years old who were diagnosed with breast cancer between January 1, 1982 through March 31, 1984. The women were retrospectively enrolled using the National Tumor Registry records in Costa Rica.

Controls: Women 25 to 58 years old who were selected at random from the general population with a multistage probability household survey.

Exposure: The woman's report of using an injectable contraceptive.

Data Collection Methods: Personal interviews conducted in the home and tumor registry records were abstracted.

Results:

	DMPA Use	
Breast Cancer	Yes	No
Cases	19	129
Controls	49	724

Crude OR = 2.2 (95% CI: 1.3 - 3.8)

(Lee et al., 1987)

(a) Interpret the odds ratio and confidence interval.

5. Circle true (T) or false (F).

   (a) T/F In the case-control study, study subjects from the study population are assigned to treatment or comparison groups.

   (b) T/F Cases are those who have the health problem under study.

   (c) T/F The purpose of the case-control study is to decide whether the exposure under study could account for the health condition of the cases.

   (d) T/F If there is an association between the exposure and the health problem, then the control group provides an estimate of the rate of exposure in the cases.

   (e) T/F To be enrolled in a case-control study, a case must clearly have the health problem under study.

   (f) T/F Controls may be matched to cases to adjust for potentially confounding variables.

   (g) T/F The effects of variables on which subjects have been matched cannot be evaluated.

   (h) T/F Cases and controls should be matched on variables related to exposure.

   (i) T/F Variables used for matching during data collection need not be considered as matched variables for analysis.

   (j) T/F Relatives of cases may not be used for matching.

   (k) T/F The exposure variable must be defined in clear, measurable terms.

6. Multiple Choice. Select one response.

   6.1 All of the following are frequency measures of exposure except?

   (a) Days per week vitamins taken
   (b) Total months of condom use
   (c) Parity
   (d) Ever use of natural methods of birth control
   (e) History of previous abortion

   6.2 All of the following are measures of exposure related to time except?

   (a) Age at first intercourse
   (b) Time since IUD was inserted
   (c) Current IUD use
   (d) Duration of OC use
   (e) Ever breast-fed

7. Which of the following are appropriate (True) methods to help reduce bias? Circle true (T) or false (F).

   (a) T/F Select all cases in a defined region.

   (b) T/F Choose controls at random from an appropriate population.

   (c) T/F If you use a hospital-based design, then the control group should include individuals with different illnesses not associated with the exposure.

   (d) T/F Use memory aids to improve subject recall.

   (e) T/F Select controls whose recall bias is likely to be similar to the recall bias of the cases.

   (f) T/F Validate recall with medical or other institutional records whenever possible.

   (g) T/F Omit from the study any participants with poor memory.

8. Multiple Choice. Select one response.

8.1 All the data sources listed below are useful in case-control studies except:

(a) Medical events that occur after diagnosis of the health problem of interest
(b) Abstracts of clinical records
(c) Vital statistics
(d) Physical examination
(e) Questionnaire data

8.2 Why do the measures of association used for cohort studies not apply to case-control studies?

(a) Because you select controls from a different population than cases
(b) Because of selection bias
(c) Because you cannot determine rates without sampling exposure
(d) Because you sample exposed and unexposed individuals in a case-control design

8.3 All of the following are types of bias except:

(a) Selection
(b) Surveillance
(c) Effect modification
(d) Confounding
(e) Diagnostic

9. Circle true (T) or false (F).

(a) T/F Case-control studies are appropriate for infrequently occurring health problems.

(b) T/F Case-control studies are appropriate for determining incidence rates.

(c) T/F Case-control studies are inappropriate for studying problems with long latency.

(d) T/F Case-control studies are generally less expensive than cohort studies.

(e) T/F In case-control studies it is often difficult to ensure that cases and controls are comparable.

(f) T/F Data collected for case-control studies permit calculation of relative risk.

(g) T/F The odds ratio is an appropriate estimate of the relative risk when the incidence rates of the health problem under study are high.

(h) T/F Case-control studies usually require smaller study sizes than cohort studies.

(i) T/F A variable may be an effect modifier or a confounder, but not both.

(j) T/F Summary relative risk measures may mask increases or decreases in risk that are present in subgroups.

(k) T/F An OR greater or smaller than 1.0 implies that the risk of the health problem is associated with the exposure.

(l) T/F Confounding bias may be minimized or eliminated by exclusions, matching, or analytic methods.

10. In outline form, design a case-control study based on the following problem situation. State the problem and the hypothesis. Describe the cases and the controls and how both groups of women will be enrolled in the study. Define the exposure in measurable terms. Specify the data collection methods. Develop the analysis table for measuring any possible association between the outcome and the exposure.

    Background: As a researcher you have a collaborative relationship with several health care providers that allows you to conduct scientific studies. In particular, you want to investigate the risk of myocardial infarction among women who used OCs in the past.

Problem:

Research
Hypothesis:

Study Design:

Cases:

Controls:

Exposure:

Data Collection Methods:

Data
Analysis:

11. Table 10.22 presents results from a study of ectopic pregnancy and current contraceptive methods (Ory et al., 1981).

Contraception and Ectopic Pregnancy

Current Method	Women With Ectopic Pregnancy	Non-Pregnant Controls	Odds Ratio	95% CI
IUD	67	497		
Barrier or natural methods	57	573		
OC	32	775		
No method	319	1078		

(a) Compute the odds ratios and confidence intervals for current IUD, barrier or natural methods, and oral contraceptive use. Consider women who used no method as the referent category.

(b) Interpret the odds ratios and confidence intervals.

(c) Compute the odds ratio and the 95% confidence interval for the risk of ectopic pregnancy among IUD users compared to OC users.

(d) Interpret the odds ratio and confidence intervals.

Suggested Answers to Practice Exercises

1. Using the information in Table 10.18, analyze and interpret the data.

   (a)

$$OR = \frac{841/724}{518/967} = 2.168$$

$$n = 3050$$

$$\chi^2 = \frac{3049 * (841 * 967 - 724 * 518)^2}{(841 + 518) * (724 + 967) * (841 + 724) * (518 * 967)}$$

$$= 109.63$$

$$\chi = 10.47$$

$$\text{Lower Limit} = 2.2^{(1 - 1.96/10.47)}$$

$$= 1.876$$

$$\text{Upper Limit} = 2.2^{(1 + 1.96/10.47)}$$

$$= 2.507$$

(b) Women who are current IUD users have a 2.2 times greater risk of PID than women who are not using any method of contraception. The 95% confidence interval means that the researcher was 95% confident that the true odds ratio is between 1.9 and 2.5. Another way of saying this is, we know that the OR is statistically larger than 1.0, since the lower confidence limit is greater than 1.0. Current IUD users are at greater risk of PID than women who are not currently using any method of contraception.

(c) Other questions the investigators might ask: Were cases and controls comparable? For example, perhaps controls should have been selected from a minor surgery clinic in the same hospital. Were other possible risk factors similar for cases and controls (e.g., number of sexual partners)?

(d) Age, marital status, education, history of sexually transmitted diseases, frequency of intercourse, previous episodes of PID, contraceptive history, previous episodes of IUD use, number of sexual partners.

2. Using Example 10.19, interpret the odds ratio.

(a) The risk of developing cervical cancer is approximately four times higher among women who were sexually active at age 15 years or younger than among women who were sexually active later in life. Since the confidence interval does not include 1.0, then the hypothesis that there is no difference between women with cervical dysplasia and controls is rejected.

3. Using Example 10.20, analyze and interpret the data.

(a) OR = 0.5 (95% CI: 0.4 - 0.7)

(b) The risk of developing PID for women currently using OCs is 0.5 (half) that of women not currently using OCs.

4. Interpret the odds ratio and confidence interval.

   (a) The risk of breast cancer among women who use DMPA is twice the risk among women who have never used DMPA. Since the confidence interval does not include 1.0, then the hypothesis that there is no difference in breast cancer risk between women who have used DMPA and women who have not used DMPA is rejected; the odds ratio is statistically significant.

5. True or false.

   (a) F Introduction. No, groups are identified according to having or not having a particular health problem.

   (b) T Introduction

   (c) T Introduction

   (d) F Control Selection. No, the control group provides an estimate of the rate of exposure if there was *no* association between the exposure and the health problem.

   (e) T Case Selection

   (f) T Control Selection

   (g) T Control Selection

   (h) F Control Selection. No, cases and controls should be matched on variables related to the exposure and the health problem.

   (i) F Control Selection. No, matching during data collection must be maintained for analysis.

   (j) F Control Selection. No, matching on relatives may be important for some investigations and will likely result in matching on other variables, such as diet and behaviors.

   (k) T Exposure Definition

6. Multiple choice.

6.1 b Exposure Definition. Total months of condom use or duration of condom use is a time-related measure of exposure.

6.2 e Exposure Definition. Ever breast-fed is a frequency measure of exposure.

(a) T Selection Bias

(b) T Selection Bias

(c) T Information Bias

(d) T Information Bias

(e) T Information Bias

(f) T Information Bias

(g) F Information Bias. No, excluding study participants with a poor memory is not a method for reducing bias. In fact, such an exclusion might introduce bias.

7. True or false.

(a) T Selection Bias

(b) T Selection Bias

(c) T Information Bias

(d) T Information Bias

(e) T Information Bias

(f) T Information Bias

(g) F Information Bias. No, excluding study participants with a poor memory is not a method for reducing bias. In fact, such an exclusion might introduce bias.

8. Multiple choice.

8.1 a Data Collection Methods. We are interested in medical events that occurred prior to the diagnosis or enrollment in a case-control study.

8.2 c Data Analysis Methods

8.3 c Bias

9. True or false.

(a) T Advantages #1

(b) F Disadvantages #3. No, incidence rates cannot be computed from case-control data.

(c) F Advantages #2. No, case-control designs are ideal for studying health problems with long latency.

(d) T Advantages #3

(e) T Disadvantages #2

(f) F Disadvantages. No, incidence rates cannot be computed from case-control data.

(g) F Data Analysis Methods. No, the odds ratio is an appropriate estimate of the relative risk when the incidence of the health problem is low.

(h) T Data Analysis Methods

(i) F Data Analysis Methods. A variable may be both a confounder and an effect modifier.

(j) T Data Analysis Methods

(k) T Data Analysis Methods

(l) T Data Analysis Methods

10. The answer given here is a suggested answer and is based on the report of the actual study referenced.

Example 10.23

Oral Contraceptive Use in the Past and Myocardial Infarction

Problem: Is long-term OC use in the past associated with the occurrence of myocardial infarction?

Research Hypothesis: Long-term OC use after discontinuation is associated with a myocardial infarction.

Study Design: Multicenter hospital-based case-control study

Cases: Women younger than 65 years old with an admission diagnosis of first myocardial infarction, defined according to World Health Organization criteria. Cases with a history of rheumatic valvular disease, cardiomyopathy, or cardiac surgery (including coronary artery bypass surgery) were excluded.

Controls: Women younger than 65 years old without a history of myocardial infarction, rheumatic valvular disease, cardiomyopathy, or cardiac surgery. Control women were admitted for nonmalignant and nongynecologic symptoms that were not related to OC use.

Exposure: The woman's report that she used medicines including oral contraceptives and noncontraceptive estrogens. The timing and brands for all medicines taken were recorded.

Data Collection Methods: Personal interview and medical record review.

Data Analysis:

	OC Use		
Myocardial Infarction	Past Use	Never Use	Total
Yes	291	613	904
No	673	1,047	1,720
Total	964	1,660	2,624

(Rosenberg et al., 1990)

11. Compute the odds ratios and confidence intervals.

(a) IUD OR = 0.4 (95% CI: 0.3 - 0.6)

Barrier or natural OR = 0.3 (95% CI: 0.25 - 0.45)

OC OR = 0.1 (95% CI: 0.1 - 0.19)

(b) Women who are currently using IUDs, traditional methods of contraception, and OCs have a lower risk of an ectopic pregnancy than women who are currently not using any method. Since the confidence intervals do not include 1.0, the hypotheses of no association between ectopic pregnancy risk for women who use these methods compared with women who do not practice contraception are rejected.

(c) OR = 3.2 (95% CI: 2.1 - 5.0)

(d) Women who are currently using an IUD have a greater risk of an ectopic pregnancy than women currently using OCs. Since the confidence interval does not include 1.0, the hypothesis of no association between ectopic pregnancy risk for women currently wearing an IUD compared with women currently using OCs is rejected.

References

Adolusi B. Carcinoma of the cervix uteri in Ibadan: coital characteristics. Int J Gynaecol Obstet 1977;15:5-11.

Bennett AE, Ritchie K. Questionnaires in medicine--a guide to their design and use. London: Oxford University Press, 1975.

The Boston Collaborative Drug Surveillance Program. Oral contraceptives and venous thromboembolic disease, surgically confirmed gallbladder disease and breast tumors. Lancet 1973;1:1399-1404.

Burkman RT, the Women's Health Study. Association between intrauterine devices and pelvic inflammatory disease. Obstet Gynecol 1981;57:269-276.

The Cancer and Steroid Hormone Study of the Centers for Disease Control, National Institute of Child Health and Human Development. Oral contraceptive use and the risk of breast cancer. N Engl J Med 1986;315:405-411.

The Cancer and Steroid Hormone Study of the Centers for Disease Control, National Institute of Child Health and Human Development. The reduction in risk of ovarian cancer associated with oral contraceptive use. N Engl J Med 1987;316:650-655.

Coulter A, Vessey M, McPherson K. The ability of women to recall their oral contraceptive histories. Contraception 1986;33:127-137.

Grimes DA, Techman T. Legal abortion and placenta previa. Am J Obstet Gynecol 1984;149:501-504.

Hogue CJR, Rubin GL, Schulz K. [Unpublished report]. Reproductive epidemiology in a nutshell. [1985]. Located at: Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, Georgia.

Lee NC, Rosero-Bixby L, Oberle MW, Grimaldo C, Whatley AS, Rovira EZ. A case-control study of breast cancer and hormonal contraception in Costa Rica. J Natl Cancer Inst 1987;79:1247-1254.

McPherson K, Neil A, Vessey MP, Doll R. Oral contraception and breast cancer. Lancet 1983;2:1414-1415.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-748.

Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976;103:226-235.

Ory HW, the Women's Health Study. Ectopic pregnancy and intrauterine contraceptive devices: new perspectives. Obstet Gynecol 1981;57:137-144.

Rosenberg L, Mitchell AA, Shapiro S, Slone D. Selected birth defects in relation to caffeine-containing beverages. JAMA 1982;247:1429-1432.

Rosenberg L, Palmer JR, Lesko SM, Shapiro S. Oral contraceptive use and the risk of myocardial infarction. Am J Epidemiol 1990;131:1009-1016.

Rubin GL, Ory HW, Layde PM. Oral contraceptives and pelvic inflammatory disease. Am J Obstet Gynecol 1982;144:630-635.

Schlesselman JJ. Case-control studies. New York: Oxford University Press, 1982.

Slone D, Shapiro S, Kaufman DW, Rosenberg L, Miettinen OS, Stolley PD. Risk of myocardial infarction in relation to current and discontinuous use of oral contraceptives. N Engl J Med 1981;305:420-424.

Sudman S, Bradburn NM. Asking questions. San Francisco: Jossey-Bass Publishers, 1983.

Woolf B. On estimating the relation between blood group and disease. Ann Hum Genet 1955;19:251-253.

Appendix 1
List of Figures, Examples, and Tables

Chapter 1 Reproductive Health Epidemiology

Chapter 2 Developing a Research Proposal

Chapter 3 Measures of Disease Frequency in Reproductive Health

Chapter 4 Epidemiologic Study Design

Chapter 5 Sample Size and Power

Chapter 6 Descriptive Statistics

Chapter 8 Randomized Clinical Trials

Chapter 9 Cohort Studies

Appendix 3 Life-Table Analysis

Appendix 2

Recommended Definitions, Standards, and Reporting Requirements for ICD-10 Related to Reproduction

Fetal, Perinatal, Neonatal, and Infant Mortality Definitions[1]

Live birth. Live birth is the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after such separation, breathes or shows any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles, whether or not the umbilical cord has been cut or the placenta is attached; each product of such a birth is considered liveborn.

Fetal death (deadborn fetus). Fetal death is death prior to the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy; the death is indicated by the fact that after such separation the fetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord, or definite movement of voluntary muscles.

Birthweight. The first weight of the fetus or newborn obtained after birth.

Low birthweight. Less than 2,500 g (up to, and including 2,499 g).

Very low birthweight. Less than 1,500 g (up to, and including 1,499 g).

Extremely low birthweight. Less than 1,000 g (up to, and including 999 g).

Gestational age. The duration of gestation is measured from the first day of the last normal menstrual period. Gestational age is expressed in completed days or completed weeks (e.g., events occurring 280 to 286 completed days after the onset of the last normal menstrual period are considered to have occurred at 40 weeks of gestation).

Preterm. Less than 37 completed weeks (less than 259 days) of gestation.

Term. From 37 completed weeks to less than 42 completed weeks (259 to 293 days) of gestation.

[1] World Health Organization. International Conference on the Tenth Revision of the International Classification of Diseases, Geneva: World Health Organization, In Press 1991.

Postterm. 42 completed weeks or more (294 days or more) of gestation.

Perinatal period. The perinatal period commences at 22 completed weeks (154 days) of gestation (the time when birth weight is normally 500 g), and ends seven completed days after birth.

Neonatal period. The neonatal period commences at birth and ends 28 completed days after birth. Neonatal deaths (deaths among live births during the first 28 completed days of life) may be subdivided into early neonatal deaths, occurring during the first seven days of life and late neonatal deaths, occurring after the seventh day but before 28 completed days of life.

Notes on Definitions

For live births, birth weight should preferably be measured within the first hour of life before significant postnatal weight loss has occurred. Whilst statistical tabulations include 500 g groupings for birth weight, weights should not be recorded in those groupings. The actual weight should be recorded to the degree of accuracy that it is measured.

The definitions of *low*, *very low*, and *extremely low* birthweight do not constitute mutually exclusive categories. Below the set limits they are all-inclusive and therefore overlap (i.e., *low* includes *very low* and *extremely low*, while *very low* includes *extremely low*).

Gestational age is frequently a source of confusion when calculations are based on menstrual dates. For the purposes of calculation of gestational age from the date of the first day of the last normal menstrual period and the date of delivery, it should be borne in mind that the first day is day zero and not day one; days 0-6 therefore correspond to *completed week zero*, days 7-13 to *completed week one*, and the 40th week of actual gestation is synonymous with *completed week 39*. Where the date of the last normal menstrual period is not available, gestational age should be based on the best clinical estimate. In order to avoid misunderstanding, tabulations should indicate both weeks and days.

Age at death during the first day of life (day zero) should be expressed in units of completed minutes or hours of life. For the second (day 1), third (day 2), and subsequent days of life, age at death should be expressed in days.

Reporting Requirements

It is recognized that legal requirements for the registration of fetal deaths and live births still vary from country to country and even within countries. However, it is recommended that, wherever possible, all fetuses and infants delivered weighing at least 500 g, whether alive or dead, be included in the statistical tabulations. When birth weight is unavailable, the corresponding criteria for gestational age (22 completed weeks), or body length (25 cm crown-heel) should be used. The criteria for deciding whether an event has taken place within the perinatal period should be applied in the order 1) birth weight, 2) gestational age, 3) crown-heel length. The inclusion of fetuses and infants weighing between 500 g and 1,000 g in national statistics is recommended both because of its inherent value and because this inclusion improves the completeness of reporting at 1,000 g and over.

In statistics for international comparison, inclusion of this group of births of extremely low birth weight disrupts the validity of comparisons and is not recommended. Countries should therefore arrange registration and reporting procedures so that the events and the criteria for their inclusion in the statistics can be easily identified. Less mature fetuses and infants not corresponding to these criteria should be excluded from perinatal statistics unless there are legal or other valid reasons to the contrary, in which case this inclusion must be explicitly stated. Where these characteristics are unknown, the event should be included in, rather than excluded from, mortality statistics of the perinatal period. Countries should also present standard statistics in which both the numerator and the denominator of all ratios and rates are restricted to fetuses and infants weighing 1,000 g or more (weight-specific ratios and rates); where birth weight is unavailable, the corresponding gestational age (28 completed weeks) or body length (35 cm crown-heel) should be used.

In reporting fetal, perinatal, neonatal and infant mortality statistics the number of deaths due to malformations should whenever possible be identified for live births and fetal deaths and in relation to birth weight of 500-999 g and 1,000 g or more. Neonatal deaths due to malformations should be subdivided into early and late neonatal deaths. The availability of this information enables perinatal and neonatal mortality statistics to be reported with or without the deaths from malformations. A malformation is defined as a congenital morphological anomaly, regarded as the underlying cause of death during the fetal and neonatal period.

Ratios and Rates

Published ratios and rates should always specify the denominator that has been used, i.e., live births or total births (live births plus fetal deaths). Countries are encouraged to provide the ratios and rates listed below, or as many of them as their data collection systems permit:

Fetal death ratio: $\frac{\text{Fetal deaths}}{\text{Live births}} * 1{,}000$

Fetal death rate: $\frac{\text{Fetal deaths}}{\text{Total births}} * 1{,}000$

Fetal death rate, weight-specific: $\frac{\text{Fetal deaths weighing 1,000 g and over}}{\text{Total births weighing 1,000 g and over}} * 1{,}000$

Early neonatal mortality rate: $\frac{\text{Early neonatal deaths}}{\text{Live births}} * 1{,}000$

Early neonatal mortality rate, weight-specific:

$$\frac{\text{Early neonatal deaths of infants weighing 1,000 g and over at birth}}{\text{Live births weighing 1,000 g and over}} * 1{,}000$$

Perinatal mortality ratio: $\frac{\text{Fetal deaths and early neonatal deaths}}{\text{Live births}} * 1{,}000$

Perinatal mortality rate[1]: $\frac{\text{Fetal deaths and early neonatal deaths}}{\text{Total births}} * 1{,}000$

[1] The perinatal mortality rate is the number of fetal deaths weighing at least 500 g (or, when birth weight is unavailable, after 22 completed weeks of gestation or with a crown-heel length of 25 cm or more), plus the number of early neonatal deaths, per 1,000 total births. Because of the different denominators in each component, this is not necessarily equal to the sum of the fetal death rate and the early neonatal mortality rate.

Perinatal mortality rate, weight-specific:

$$\frac{\text{Fetal deaths weighing 1,000 g and over, plus early neonatal deaths of infants weighing 1,000 g and over at birth}}{\text{Total births weighing 1,000 g and over}} * 1{,}000$$

Neonatal mortality rate: $\frac{\text{Neonatal deaths}}{\text{Live births}} * 1{,}000$

Neonatal mortality rate, weight-specific:

$$\frac{\text{Neonatal deaths of infants weighing 1,000 g and over at birth}}{\text{Live births weighing 1,000 g and over}} * 1{,}000$$

Infant mortality rate: $\frac{\text{Number of deaths under one year of age}}{\text{Live births}} * 1{,}000$

Infant mortality rate, weight-specific:

$$\frac{\text{Infant deaths among live births weighing 1,000 g and over at birth}}{\text{Live births weighing 1,000 g and over}} * 1{,}000$$

Maternal Mortality Definitions

Maternal death. The death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes.

Late maternal death. The death of a woman from direct or indirect obstetric causes more than 42 days but less than one year after termination of pregnancy.

Pregnancy-related death. The death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the cause of death. This definition is provided to permit calculation of an alternative to the *maternal death* rate in countries that wish to identify deaths occurring in pregnancy, childbirth and the puerperium, but cannot distinguish direct and indirect maternal deaths as defined.

Maternal deaths should be subdivided into two groups:

Direct obstetric deaths: those resulting from obstetric complications of the pregnant state (pregnancy, labor and puerperium), from interventions, omissions, incorrect treatment, or from a chain of events resulting from any of the above.

Indirect obstetric deaths: those resulting from previous existing disease or disease that developed during pregnancy and which was not due to direct obstetric causes, but which was aggravated by physiologic effects of pregnancy.

Reporting Requirements

For the purposes of the international reporting of maternal mortality only those maternal deaths occurring before the end of the 42-day reference period should be included in the calculation of the various ratios and rates, though the recording of later deaths is useful for national analytical purposes.

Late maternal deaths should not be included in the calculation of the maternal mortality rate.

Published maternal mortality rates should always specify the numerator (number of recorded maternal deaths), which can be given as:

- the number of recorded direct obstetric deaths, or
- the number of recorded obstetric deaths (direct plus indirect).

The denominator used for calculation should likewise be specified as either the number of live births or the total number of births (live births plus fetal deaths). Where both denominators are available, a calculation should be published for each.

Ratios and Rates

These should be expressed as a ratio of the numerator to the denominator, multiplied by k (where k may be 1,000, 10,000, or 100,000 as preferred and indicated by the country). Maternal mortality ratios and rates can thus be expressed as:

Maternal mortality rate[1]: $\frac{\text{Maternal deaths (direct and indirect)}}{\text{Live births}} * k$

Direct obstetric mortality ratio: $\frac{\text{Direct obstetric deaths only}}{\text{Live births}} * k$

Pregnancy-related mortality ratio: $\frac{\text{Pregnancy - related deaths}}{\text{Live births}} * k$

[1] The use of the term *rate*, although inexact in this context, is maintained for the sake of continuity.

Appendix 3

Life-Table Analysis

Introduction

Life-table analysis is a method of summarizing the results of a study by time, that is, by grouping into smaller intervals the time between admission to a study and the end of the designated follow-up period. For each time interval, life-table analysis records the number of study subjects who are still in the study at the start of the interval, the number who experience the outcome of interest during the interval (e.g., pregnancy, intrauterine device (IUD) expulsion, continuation of contraceptive use), and the number of study subjects who discontinue participation during the interval without having had the outcome of interest (i.e., because they are lost to follow-up or no longer participate in the study for a reason other than the outcome of interest). From these numbers, the probability of the outcome event occurring in each interval is estimated.

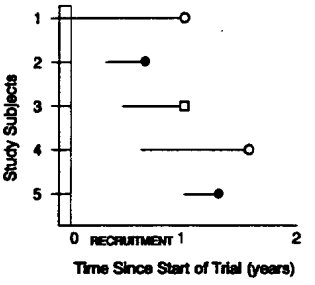
The major advantage of the life-table approach is that it takes into account how long each participant is in the study. Life tables take advantage of the experience of individuals even though they do not complete the study. This establishes the period of time within the study when each participant was at risk of exhibiting the outcome of interest.

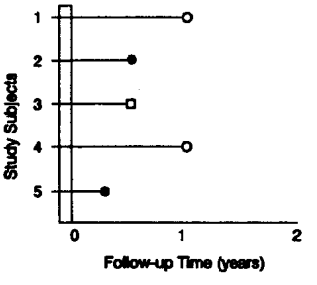
Grouping observations into fixed time periods of follow-up. In most clinical trials, the subjects do not enter the study at the same time. Their entry is staggered. In life-table analysis, however, all entry times are considered time zero. In Figure 1, study subjects were recruited during a one-year time period and followed up for one year. Their staggered entry is shown in the upper panel of the figure. Study subjects 1 and 4 completed the trial without experiencing the event of interest (i.e., pregnancy). Study subject 3 was lost to follow-up and study subjects 2 and 5 experienced the event of interest during the follow-up. For life-table analysis, the time of follow-up observation for the five study subjects are shown in the lower panel of Figure 1. All entry times are at time zero.

Study subjects who do not experience the event of interest while they are being followed up are said to survive (without the event). They could complete the study without the event or drop out (be withdrawn because they have an event that disqualifies them from the study or be lost to follow-up). Some study subjects enter the study near its conclusion and do not complete the full time period before the conclusion of the study. In these situations, the follow-up experience of all study subjects who survive is said to be censored before we can observe the event of interest. Note, however, that the term *censored* as used for any particular time period includes only those lost or discontinued during the interval for some reason other than the event of interest, not those who complete the interval without the event.

Figure 1

Life-Table Analysis
Time Since Start of Trial Versus Follow-up Time





o = Completed trial with no event
• = Had the event of interest during follow-up
□ = Withdrawn or lost to follow-up

The steps and formulas required for life-table calculations are:

Step 1. Group the observations into fixed time periods of follow-up.

Step 2. Compute the probability of surviving any time interval, the ith interval, given survival up to the beginning of the interval as:

(8.33.1)

$$q_i = \frac{n_i - d_i - (c_i/2)}{n_i - (c_i/2)}$$

where n_i = number of subjects entering the ith interval; d_i = number of events of interest in the ith interval; c_i = number of subjects censored in the ith interval (lost-to-follow-up or discontinuing for a reason other than the event of interest); q_i is also called the survival rate for the ith interval; and $n_i - (c_i/2)$ is the group at risk for the event of interest during the ith interval.

Step 3. Compute the probability of not surviving the ith interval as:

(8.33.2)

$$p_i = 1 - q_i$$

$$= \frac{d_i}{n_i - (c_i/2)}$$

p_i is also called the event rate for the ith interval.

Step 4. Compute the probability of surviving from the beginning of the study until the end of each time interval. The probability of surviving from the beginning of the study until the end of the ith interval is:

(8.33.3)

$$S_i = q_1 * q_2 * \cdots * q_i$$

S_i is also called the cumulative survival rate.

Step 5. Compute the probability of experiencing the event of interest sometime from admission to the study until the end of the ith interval, $1 - S_i$. This is also called the cumulative event rate.

Step 6. Compute the variance of the cumulative event rate for the ith interval as:

(8.33.4)

$$\text{Variance}\ (1 - S_i) = \text{Variance}(S_i)$$

$$= S_i^2 * \sum_{k=1}^{i} \left[\frac{1}{n_k - d_k - (c_k/2)} - \frac{1}{n_k - (c_k/2)} \right]$$

Step 7. Compute the Standard Error (SE) for $1 - S_i$ as:

(8.33.5)

$$SE\ (1 - S_i) = SE\ (S_i)$$

$$= \sqrt{\text{Variance}\ (S_i)}$$

Recall that the SE of the cumulative event rate is also known as the standard deviation. Steps 3, 5, and 7 are required for the life tables in Table 2. The other steps, however, are needed to calculate these items.

Table 2

Life Table
Sponge Users

Follow-up Month i	Number Entering n_i	Number Pregnant d_i	Number Censored c_i	Step 3 Monthly Pregnancy Probability p_i	Step 5 Cumulative Pregnancy Probability $1 - S_i$	Step 7 SE of Cumulative Pregnancy Probability SE $(1 - S_i)$
1	723	15	104	0.0224	0.0224	0.0057
2	604	8	35	0.0136	0.0357	0.0073
3	561	12	26	0.0219	0.0568	0.0094
4	523	10	41	0.0199	0.0756	0.0109
5	472	11	31	0.0241	0.0979	0.0125
6	430	10	31	0.0241	0.1196	0.0140
7	389	6	23	0.0159	0.1336	0.0149
8	360	6	8	0.0169	0.1482	0.0158
9	346	5	11	0.0147	0.1607	0.0165
10	330	4	24	0.0126	0.1713	0.0171
11	302	1	12	0.0034	0.1741	0.0173
12	289	0	85	0.0000	0.1741	0.0173
		88				
			Diaphragm Users			
1	717	5	102	0.0075	0.0075	0.0033
2	610	9	14	0.0149	0.0223	0.0059
3	587	12	39	0.0211	0.0430	0.0083
4	536	4	49	0.0078	0.0505	0.0090
5	483	6	36	0.0129	0.0627	0.0102
6	441	7	31	0.0165	0.0781	0.0116
7	403	3	24	0.0077	0.0852	0.0122
8	376	5	16	0.0136	0.0977	0.0132
9	355	3	23	0.0087	0.1055	0.0139
10	329	3	25	0.0095	0.1140	0.0146
11	301	0	24	0.0000	0.1140	0.0146
12	277	4	63	0.0163	0.1284	0.0160
		61				

For month 1:

$$q_1 = \frac{n_1 - d_1 - (c_1/2)}{n_1 - (c_1/2)}$$

$$= \frac{723 - 15 - (104/2)}{723 - (104/2)}$$

$$= \frac{656}{671}$$

$$= 0.9776$$

$$p_1 = 1 - q_1 = 1 - 0.9776 = 0.0224$$

$$S_1 = q_1 = 0.9776$$

$$1 - S_1 = 0.0224$$

$$\text{Variance}(1 - S_1) = S_1^2 * \left[\frac{1}{n_1 - d_1 - (c_1/2)} - \frac{1}{n_1 - (c_1/2)}\right]$$

$$= 0.9776^2 * \left[\frac{1}{723 - 15 - (104/2)} - \frac{1}{723 - (104/2)}\right]$$

$$= (0.9557) * \left[\frac{1}{656} - \frac{1}{671}\right]$$

$$= 0.00033$$

$$\text{SE}(1 - S_1) = \sqrt{\text{Variance}(1 - S_1)} = 0.0057$$

For month 2:

$$q_2 = \frac{n_2 - d_2 - (c_2/2)}{n_2 - (c_2/2)}$$

$$= \frac{604 - 8 - (35/2)}{604 - (35/2)}$$

$$= \frac{578.5}{586.5}$$

$$= 0.9864$$

$$p_2 = 1 - q_2 = 0.0136$$

$$S_2 = q_1 * q_2 = 0.9776 * 0.9864$$

$$= 0.9643$$

$$1 - S_2 = 0.0357$$

$$\text{Variance}(1 - S_2) = 0.9643^2 * \left[\frac{1}{656 - 671} - \frac{1}{578.5 - 586.5}\right] = 0.000054$$

$$\text{SE}(1 - S_2) = \sqrt{\text{Variance}(1 - S_2)} = 0.0073$$

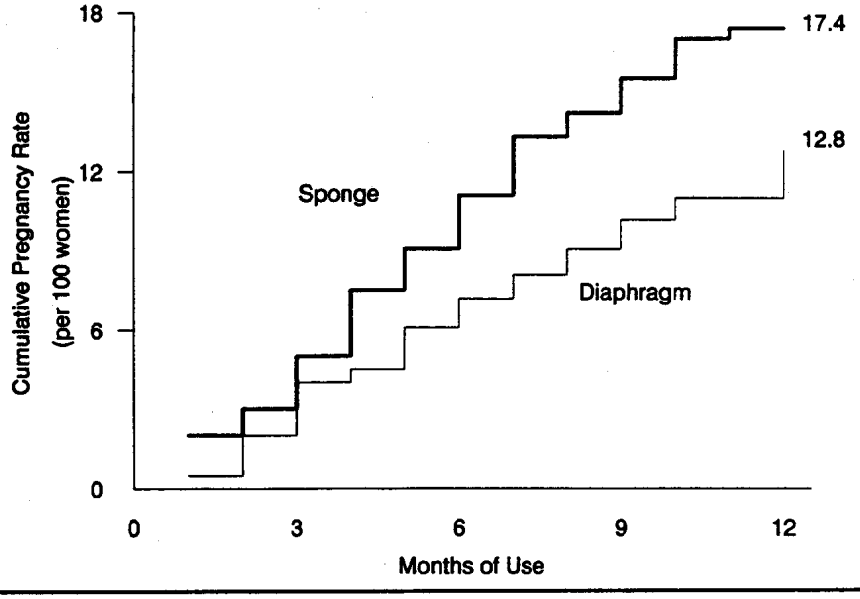
Table 2 contains follow-up data from the randomized parallel clinical trial of the contraceptive sponge and the diaphragm in Example 8.32 (Edelman et al., 1984). The format for Table 2 is a good working format. To illustrate the calculations, consider the first two months of follow-up for sponge users (Use formulas 8.33.1 through 8.33.5).

In reporting the results of a life-table analysis, the emphasis is on the cumulative event rate (cumulative pregnancy probability $(1-S_i)$ in Table 2). With the cumulative event rate, the probability of having the event of interest within the first k months of use of a treatment can be estimated. For example, the probability of becoming pregnant in the first six months is 0.1196 for sponge users and 0.0781 for diaphragm users. The SE for the cumulative event rate measures the precision of the estimate.

For comparative trials, the interest is usually in comparing cumulative event rates or probabilities between treatments. This is the probability calculated in Step 5 as shown in Table 2. A useful way to compare cumulative event rates is to plot them. The method of plotting is illustrated in Figure 3. Since the estimates of cumulative event rates are made at the end of each month, the plotted points are joined by steps rather than by straight lines. The joined, plotted points are called cumulative event rate curves.

Figure 3

Cumulative Pregnancy Rates
for Sponge and Diaphragm Users



The rates are expressed as pregnancies per 100 women. For the contraceptive sponge, the annual pregnancy rate is estimated at about 17 per 100 women. For the diaphragm, approximately 13 of 100 users will become pregnant within one year. The plots show that throughout the follow-up, sponge users have a higher pregnancy rate than diaphragm users.

Point Estimates and Confidence Intervals

The primary point estimate from the life table is the cumulative event rate at different times during the follow-up. By combining the point estimates of the cumulative event rate with the corresponding standard error of the estimate of the cumulative event rate, a confidence interval for the cumulative event rate at a specific time during the follow-up can be calculated. The formula for a 95% confidence interval is:

(8.34.1)

$$\text{Lower Limit} = (1 - S_i) - 1.96 * SE_i(1 - S_i)$$
$$\text{Upper Limit} = (1 - S_i) + 1.96 * SE_i(1 - S_i)$$

For the contraceptive sponge data in Table 2, the 95% confidence interval for the 12-month cumulative pregnancy probability is:

$$\text{Lower Limit} = 0.1741 - 1.96 * 0.0173 = 0.1402$$
$$\text{Upper Limit} = 0.1741 + 1.96 * 0.0173 = 0.2080$$

Note that the values for $(1 - S_i)$ and $SE_i(1 - S_i)$ to be used in formula 8.34.1 are from the last line of the upper panel in Table 2.

Table 4

Condensed Life Table
Sponge Users

Follow-up Month	Cumulative Pregnancy %	95% CI for the Cumulative Probability
1	2.24	(1.12, 3.36)
3	5.68	(3.84, 9.52)
6	11.96	(9.22, 14.70)
9	16.07	(12.84, 19.30)
12	17.41	(14.02, 20.80)

In reports and publications, it is useful to present an abbreviated version of the working life table that includes a column for the confidence interval for the cumulative event rate. The cumulative event rate and confidence intervals may be expressed as percentages or proportions. Table 4 presents the cumulative pregnancy probability and confidence intervals expressed as percentages.

Statistical Comparison. The relative effectiveness of two treatments is best evaluated statistically by comparing the cumulative rate curves for the total study period. The logrank method uses a chi-square (χ^2) statistic to compare the number of observed events for each treatment with the number of events expected if the treatments were equally effective. For purposes of illustration, assume that we are comparing treatment A with treatment B. The calculations are made easier by using a table that requires the following entries:

i = the number of the follow-up interval (e.g., month i where i = 1, 2, . . ., 12).

d_i = total number of events observed in interval i (both treatments combined).

n_{Ai} = number at risk in treatment group A in interval i.

= number in group A entering the interval − 1/2 the number in group A censored during the interval.

n_{Bi} = number at risk in treatment group B in interval i.

= number in group B entering the interval − 1/2 the number in group B censored during the interval.

E_{Ai} = expected number of events in group A in interval i.

(8.36.1)

$$= \frac{n_{Ai}}{n_{Ai} + n_{Bi}} * d_i$$

E_{Bi} = expected number of events in group B in interval i.

(8.36.2)

$$\frac{n_{Bi}}{n_{Ai} + n_{Bi}} * d_i = d_i - E_{Ai}$$

In the sponge and diaphragm example, consider the first month in Table 2:
d_1 = 15 sponge pregnancies + 5 diaphragm pregnancies = 20 pregnancies.

During the first month of follow-up, 723 sponge users entered the trial, 104 quit the trial for a reason other than pregnancy or were lost to follow-up (censored) during the month.

$n_{S1} = 723 - (104/2) = 671$ where S refers to sponge usage.

During the first month of follow-up, 717 diaphragm users entered the trial, 102 quit the trial or were lost to follow-up (censored) during the month.

$n_{D1} = 717 - (102/2) = 666$ where D refers to diaphragm usage.

$$E_{S1} = \frac{n_{S1}}{n_{S1} + n_{D1}} * d_1 = \frac{671}{671 + 666} * 20 = 10.04$$

$$E_{D1} = d_1 - E_{S1} = 20 - 10.04 = 9.96$$

The calculations in Table 5 are used to determine E_S and E_D, the expected number of pregnancies among the sponge and the diaphragm users, respectively. The expected values are compared to the observed values using the following chi-square (χ^2) statistic with 1 degree of freedom (df):

(8.37.1)

$$\chi^2_{1df} = \frac{(O_S - E_S)^2}{E_S} + \frac{(O_D - E_D)^2}{E_D}$$

$$= \frac{(88 - 73.79)^2}{73.79} + \frac{(61 - 75.21)^2}{75.21}$$

$$= 5.42$$

where O_S and O_D are the observed number of pregnancies among the sponge and the diaphragm users, respectively (see Table 2). The example $\chi^2_{1df} = 5.42$ has a p-value slightly less than 0.025. This is a small p-value and indicates a statistically significant probability that the two treatments are not equally effective. The small p-value taken together with the life-table displays and the plot of the cumulative pregnancy curves suggests that the sponge is less effective than the diaphragm in preventing pregnancy.

Table 5

Logrank Test

Month i	Total Pregnancies d_i	n_{Si}	n_{Di}	E_{Si}	E_{Di}
1	20	671	666	10.04	9.96
2	17	586.5	603	8.38	8.62
3	24	548	567.5	11.79	12.21
4	14	502.5	511.5	6.94	7.06
5	17	456.5	465	8.42	8.58
6	17	414.5	425.5	8.39	8.61
7	9	377.5	391	4.42	4.58
8	11	356	368	5.41	5.59
9	8	340.5	343.5	3.98	4.02
10	7	318	316.5	3.51	3.49
11	1	296	289	0.51	0.49
12	4	246.5	245.5	2.00	2.00
				E_S = 73.79	E_D = 75.21

Practice Exercises

1. Circle true (T) or false (F).

   (a) T/F In life table analysis, we are interested in how long persons in the trial do not have the outcome.

   (b) T/F Life tables summarize results of the study according to time intervals.

   (c) T/F Life tables take advantage of the experience of individuals even though those persons do not complete the study.

   (d) T/F Life tables provide a method of recording the actual time when study subjects enter the trial.

   (e) T/F Life-table analysis permits estimates of the probability that an individual who follows a particular treatment regimen will have the outcome within 3 months.

   (f) T/F The results of life tables are usually analyzed by visual methods alone.

   (g) T/F The logrank method is used to compare cumulative event rate curves.

Suggested Answers to Practice Exercises

1. True or false.

   (a) T

   (b) T

   (c) T Any length of time in the study is recorded in the life table.

   (d) F No. In the analysis, all participants are treated as if they enter the study at the same time.

   (e) T

   (f) F No. Visual methods and formal statistical tests are used to analyze life tables.

   (g) T