Basic Methodology of

Scientific Research

for Clinicians

LECTURE NOTES FROM
THE EPIDEMIOLOGICAL RESEARCH TRAINING COURSE
AT UNIVERSITY OF MEDICINE AND PHARMACY,
HO CHI MINH CITY

Uploaded in June, 2014

Japan International Cooperation Agency

This textbook is made available on the course website, enabling it to be constantly upgraded based on comments from course participants and any interested readers. If you wish to share any comments, please e-mail the project manager.

Contact

Goto, Aya
Project Manager
Department of Public Health, Fukushima Medical University School of Medicine, Japan
E-mail: agoto@fmu.ac.jp

Forewords

Clinicians face the constant challenge of providing the best care to their patients. Although a substantial amount of knowledge on clinical management has been inculcated in doctors' medical training, rapid changes of the environmental and societal factors along with the advancement of our knowledge in molecular medicine, immunology, pathology and many other related disciplines have demanded doctors to update their knowledge continuously in order to provide their patients with the best care.

Furthermore, new information is being generated almost everyday in the realm of medicine today, making it a challenging task for clinicians to select the most accurate and valuable data at any point in time. As most clinical studies are carried out in developed countries, the application of these findings in a developing country like Vietnam will further require critical analysis and judgment. As clinical observations are abundant and diverse in nature, it is necessary to design proper clinical studies to resolve the issues encountered from clinical practice. In this vein, epidemiology can be considered as one of the most appropriate tools for clinicians to exchange their clinical experiences with local and international colleagues.

The skills of critical appraisal of studies, conducting clinical research, and effective communication do not come naturally and spontaneously. These skills can only be acquired through deliberate research and learning in epidemiology. The chapters included in this book were based from the series of courses "Epidemiological Research Training Courses" resultant of the collaboration between Ho Chi Minh City (HCMC) University of Medicine and Pharmacy and the Department of Public Health of Fukushima Medical University School of Medicine, Japan from 2004 to 2009. These courses were originally intended for clinicians working in HCMC. Since 2011, this project has an expanded focus on capacity building towards evidence-based medicine and clinical services with a simultaneous aim to outreach a wider range of medical doctors working in and out of HCMC. Funded by the Japan International Cooperation Agency (JICA), this training was only made possible by further support from HCMC University of Medicine and Pharmacy, HCMC Medical Association, Fukushima Medical University, and Fukushima prefecture. The teaching team has now evolved to

i

include faculty from other universities in Japan. It is worth mentioning that the first few batches of clinicians who have graduated from these series have also now become teaching assistants on their own

The objectives of these courses are to provide the basic foundational knowledge in epidemiology and biostatistics to clinicians, building their capacities to process information and doing clinical research. The courses placed less emphasis on the theoretical aspects, and focused more on the practical domain. As a result of these trainings, some doctors have completed their research projects and published their findings in peer-reviewed medical journals. Such accomplishments represent an encouraging start despite the modest number of publications. A number of hospitals in Vietnam are now also regularly organizing Journal Clubs to update their clinicians' knowledge and to disseminate know-hows in the critical appraisal of information reported in medical papers. Additional benefits derived from the series of courses include the participants' recognition of the benefits of teamwork, and the satisfaction from conducting their own clinical research.

This book is the result of the collaboration between faculty members and teaching assistants, and above all represents an effective and fruitful collaboration between Vietnamese and Japanese clinicians. We would like to express our sincere thanks to the Department of Public Health, Fukushima Medical University School of Medicine and Fukushima prefecture, Japan. We also like to thank the Executive Boards of HCMC University of Medicine and Pharmacy, and the HCMC Medical Association for their support to make this project a successful one. We are very grateful to JICA for its funding to sustain the running of this project. Our special thanks to Associate Professor Aya Goto, who has been the main driver of this endeavor, and Dr. Nguyen Quang Vinh, who has introduced Associate Professor Aya Goto to us and his active participation as part of the teaching team. We also thank the faculty of Hanoi School of Public Health, who has provided us with useful comments on the manuscript. We hope to receive feedback from the readers about the contents of this book for subsequent editions in the future.

Nguyen Thy Khue, MD, PhD
Vice Chair, Ho Chi Minh City Medical Association
Associate Professor Emeritus, University of Medicine and Pharmacy, Ho Chi Minh City

Research offers clinicians an opportunity to confirm, clarify, or discover new aspects of a subject or topic of interest. Research plays a key role in solving practical problems and advancing medical knowledge, thus improving patient care. This book is a short introduction to the research and research methods that have an emphasis on its application in clinical areas. It covers all key concepts of epidemiology and research methods, from the introduction of basic steps in research process, specific guidelines on literature search and critical appraisal, different study designs, key concepts in biostatistics, to basic tests used commonly in clinical research.

Despite the proliferation of textbooks in epidemiology and research methods in the past two decades, I have encountered, surprisingly, some difficulties in finding one that suits the needs of medical doctors who intend to conduct their own clinical research, particularly for those in Vietnam. Now I have an opportunity to read this book and I strongly believe that this book will provide robust guidelines for Vietnamese researchers working in clinical areas..

Bui Thi Thu Ha, MD, PhD

Associate Professor & Dean, Hanoi School of Public Health

Contents

Dedication to Dr. Pham Nghiem Minh

This book is dedicated in memory of Dr. Pham Nghiem Minh, who was one of the most successful participants of the inaugural course and passed away in 2013. Without the active involvement of course participants such as Dr. Pham's, our course could not have continued for over a decade.

Contributors

Editors

Vietnamese version:

Nguyen Thy Khue, MD, PhD (Vice Chair, Ho Chi Minh City Medical Association; Associate Professor Emeritus, University of Medicine and Pharmacy, Ho Chi Minh City)

English version:

Goto, Aya, MD, MPH, PhD (Associate Professor, Department of Public Health, Fukushima Medical University School of Medicine, Japan)

Reviewer

Le Cu Linh, MD, PhD. (Associate Professor, Department of Demography, Hanoi School of Public Health)

Authors

In alphabetical order

Koriyama, Chihaya, MD, MPH, PhD (Associate Professor, Department of Epidemiology and Preventive Medicine, Kagoshima University)

Nguyen Quang Vinh, MD, MS (Unit Chief, Clinical Epidemiology Unit, Nguyen Tri Phuong Hospital)

Nguyen Thi Boi Ngoc, MD (Attending physicians, Department of Endocrinology, People's Hospital 115, Ho Chi Minh City)

Nguyen Thi Tu Van, MD, PhD (Senior Lecturer, Department of Obstetrics and Gynecology, University of Medicine and Pharmacy, HCMC)

Nguyen Thu Tinh, MD (Lecturer, Department of Pediatrics, University of Medicine and Pharmacy, HCMC)

Suzuki, Yuriko, MD, MPH, PhD (Division Chief, Department of Adult Mental Health, National Institute of Mental Health)

Tran Quang Nam, MD (Lecturer, Department of Endocrinology, University of Medicine and Pharmacy, HCMC)

Tran The Trung, MD (Lecturer, Department of Endocrinology, University of Medicine and Pharmacy, HCMC)

Tran Viet Thang, MD (Lecturer, Department of Endocrinology, University of Medicine and Pharmacy, HCMC)

Vo Tuan Khoa, MD (Attending physician, Department of Endocrinology, People's Hospital 115, Ho Chi Minh City)

Yokokawa, Hirohide, MD, PhD (Associate Professor, Department of General Medicine, Juntendo University School of Medicine)

English editor

Alden Y Lai, MPH (Research Fellow, National University of Singapore)

Course Information

Aya Goto, Nguyen Quang Vinh

1. Course Background

Health research is one of the driving forces behind the improvement of health system performance. Research can help countries identify their own needs and communicate findings for informed policy-making and implementation. In this vein, capacity in health research is therefore an essential first step into quality healthcare.

In Vietnam, continuing medical education has gained increasing attention through the recent establishment of legal and policy frameworks. Having recognized the need to develop institutional capacity in research, Ho Chi Minh City University of Medicine and Pharmacy (UMP) started seeking external technical support in 2000. The authors of this textbook have thereafter taken leading roles in the support of such continuing medical education, conducting training courses. The very first project was initiated by Population Council, which was a 9-month part-time course targeting obstetricians and gynecologists. The training format was then revised to short-term full-time course and co-organized by UMP and Fukushima Medical University. The schedule and content of the course is summarized in Table 1.1.

Table 1.1 Course organization

	Courses I-IV	**Courses V 1-3**	**Courses VI 1&2**
Time	2004-2009	2010-2012	2013-2015
Grant	Japanese government research grant	Japan International-al Cooperation Agency	Japan International-al Cooperation Agency, local company
Accreditation	University	University, Medical Association	University, Medical Association, Ministry of Health (tentative)
Target	Physicians at university	Physicians in Ho Chi Minh City	Physicians in the south
Lecturer	Japanese	Japanese, Vietnamese, third country	Japanese, Vietnamese, third country
Course structure	· Lectures on basic epidemiology and biostatistics · Groups works	· Lectures on basic epidemiology and biostatistics · In-class exercises · Groups works	· Lectures on basic epidemiology and biostatistics · In-class exercises · Groups works

2. Learning Objectives

Our courses target practicing physicians and offer basic training in practical survey methods that can be applied to improve daily clinical practice. The course deals with various health topics at all stages of life, addressing physical, mental and social aspects of people's health.

The primary objective of the course is to enable participants to understand epidemiological evidence, design and conduct epidemiological studies, and disseminate and use findings. Other objectives include:

1. To reinforce basic knowledge of epidemiology and biostatistics;

2. To provide technical competencies required for research: literature search, critical appraisal of published medical evidence, study design, questionnaire development, data handling, data analysis, and results presentation.

This textbook introduces and explains basic terms and concepts that have been taught in the courses, although the training courses include more actual examples of clinical research.

Basic Steps of Research

Aya Goto, Nguyen T. Tinh

1. Research questions

Why am I seeing this type of patients more and more these days? How can we better diagnose this disease? What is the most updated treatment for this disease? How can we prevent the recurrence of this disease in patients? These are all clinical research questions that come into your mind every day. These questions will lead you to constructive research aims.

The topic of your research does not have to be a specific disease or treatment outcome. It can be any health-related events. It can be health behaviors, physical or psychological symptoms, duration of hospital admission, or cost of treatment. In this textbook, the health event is expressed as the outcome throughout. Formulating a research question is similar to the way you define a question for literature search, which is explained in Chapter 2.

2. Tools

The tools you need to answer these questions are epidemiology and biostatistics. You must be able to understand epidemiology and biostatistics terminologies when you read scientific articles. You must learn how to use

these terminologies when you want to conduct an epidemiological study. Terms and concepts in <u>epidemiology</u> help you to collect data, and those in of <u>biostatistics</u> help you to analyze the data; both help you to understand evidence. You can learn the basic concepts in later chapters of this book, while issues that warrant special attention in epidemiology are explained in Chapter 3, followed by major techniques to collect data in Chapters 4 to 8. Basic concepts of biostatistics are outlined in Chapters 9 and 10. **Important keywords are <u>underlined</u> and listed in the end of this book. You should be able to explain and use the underlined terms after reading this book.**

---

3. Where to start?

You need to start with learning how to use <u>literature search</u> engines like PubMed and Medline. In addition, there are many e-journals that provide information free-of-charge at current. These will open the door for you to the ocean of scientific evidence. Then, you will need to learn how to swim. You will need to learn how to critically assess the articles you have collected. Articles are written in epidemiology and biostatistics terminologies. The main goal of literature searches is to know what is known and what is unknown about the topic you are searching, based on where you decide to start your research.

If you are not even familiar with English terms, the MeSH function of the PubMed can guide you. Just enter the word you know and the MeSH will list appropriate search terms. When you are using the search for the first time, look for reviews or articles published by Cochrane. These will give your overview of existing evidence. If your time is limited, focus on checking articles from the past ten years. These techniques are explained more in detail in Chapter 2.

4. Preparations

The success of your research, no matter how small or big it is, depends on how well you design it. The major study designs, which you will learn later on (Chapters 4 to 8), are descriptive studies, case series, cross-sectional studies, case-control studies, cohort studies, and intervention studies. This is a list according to levels of difficulty in terms of planning and executing. You should look around to find appropriate collaborators and select a design that is feasible for your team to manage.

After forming a feasible team that you feel comfortable to work with, you will need to develop a research protocol and a questionnaire. An example document is uploaded with this text book. Box 1.1 shows the basic items to be included in your protocol. If this is the first research methodology textbook you are reading, please read the box after reading of the main text of this book.

Box 1.1. Research protocol

1. Research Title
2. Research Team: Name, affiliation, and responsibilities of each member
3. Grant: If available.
4. Introduction
   1) Brief explanation of the study topic that you are trying to address. International and domestic situation of the issue.
   2) Description of previous studies (international and domestic). If you have conducted a pilot or another study on the same topic, briefly describe here. What is new in this study?
   3) State specific objectives.
5. Study Design and Method
   1) Study design: Cross-sectional, case-control, cohort, or intervention? Why is this study design being chosen?

2)   Study period and place
3)   Subjects: Who will be included or excluded? In the case of a case-control study, define cases and controls, including whether matching will be performed. Calculate a sample size by using realistic assumptions and appropriate tests.
4)   Main outcome measure(s) and other survey items: Describe how you develop the questionnaire. Attach a draft of questionnaire with a face sheet explaining the study. In a clinical setting, if you are utilizing existing data, describe how you select the survey items from medical files or other data sources. If you are collecting information through observing clinical practice, describe how you develop the observation sheet (a list of observation points).
5)   Survey procedures: Is it an interview or self-administered questionnaire survey? For questionnaire surveys, describe how you distribute and collect the questionnaires. For interview surveys, describe who will, and where and when to do the interviews. If you are using clinical tests or any data collection involving clinical procedures, explain the measurement techniques. If you are conducting an intervention study, explain the intervention thoroughly. Do you give any rewards/incentives to subjects? Who monitors the survey or intervention?
6)   Ethical consideration and procedures: Explain methods of recruitment and obtaining informed consent (oral or written). Are there any potential risks involved in this study, such as risks for your study subjects or the research team? How will you control these risks? Who will review and approve your study protocol?
7)   Data management: Who codes and enters the data? Who keeps the data, and where?
8)   Data analysis: Select a statistical package and statistical tests to be used. Preferably, prepare the data analysis plan with dummy tables/graphs (blank tables/graphs) to be shown in the final report.
6.   Limitations: Do you have any potential bias in your study (recall bias, observer bias, selection bias, losses to follow-up, or misclassification)? If you have any, how do you try to minimize the bias?
7.   Dissemination of study results: Are you going to present the findings

> at any conference, submit a paper to any journal, or report the results to your study population?
> 8. Expected benefits: How can the study population, you, and your institute/hospital benefit from the study? How can the study contribute to the progress of further research in the field? Can you make any recommendations for the development of health policies concerning your study topic?
> 9. References: List the important references used to develop your research protocol.
> 10. Time table of major activities
> 11. Budget: Personnel, equipment, supplies, travels, patient care, communication, and other expenses

It is often not realistic to target the entire population. Sampling and recruiting of your study participants is one of the important key points when conducting an epidemiology study. You might be tempted to ask any of your patients to volunteer. However, the volunteers tend to be different from others; most of the time they show better compliance or certain health behaviors. You must select a scientifically sound recruitment method. A simple random sampling is easy to understand theoretically, but may not be useful in practice for clinicians. It is often difficult to randomize, especially when your outpatients come without prior appointments. If that is the case, you may consider systematic/convenience sampling. For example, you can recruit patients coming to your office consecutively or recruit every third other patient. In any case, do not forget to count how many you have recruited in total, and how many have refused to participate. Otherwise, you will not be able to calculate a response rate. This issue is further explained in Chapter 5.

In addition, always do not forget to do a pilot testing of your tools in a setting similar to the main survey. You can test the feasibility of your plans written in your protocol; whether all equipment are prepared, staff have followed the protocol, recruited subjects joined the study and answered your

questions. Check the <u>response rate</u> and proportion of missing answers. The pilot study can also give you data to calculate a sample size.

5. Returning results

Study results should always be returned to where the data originally belonged to. They should be utilized to improve your services and ultimately, to improve health of your patients and communities. You can also contribute to domestic and international academia by publishing the study findings.

6. A historical perspective

Many epidemiology textbooks start with the case study of John Snow. I did not give much attention to this matter in the beginning of my career, but the longer I worked in this field, the more I came to realize about the important notions of this famous case. He was a practicing anesthesiologist. At the time when etiology of Cholera was not known, he found and assessed the epidemic in a place where he lived, and contributed to resuming it. His research question came from his daily clinical practice; he mapped the incident cases, and found the source of the disease. He was a clinician with a public health perspective. With knowledge of and techniques in epidemiology and biostatistics, a clinician can contribute to the health promotion of a wider population, not just those who visit you.

Literature Search and Critical Appraisal

Chihaya Koriyama, Vo Tuan Khoa

1. Literature search

Searching for literature is an absolutely essential process at the beginning (before) of any research, in the middle of your research, and in writing your paper. You should spend more time in literature searching, than the writing of the manuscript. Otherwise, you will get a wrong direction, or your efforts to conduct the study will end up in vain. Thus, sufficient and unsparing efforts of this process are recommended.

1) What you should get in literature searching

You need to know what has already been done, or known before starting your study. To do so you need to review literature to improve your knowledge of the subject which you are interested in. This is also true during the research and when you write the manuscript. You need to always stay updated of your knowledge in the subject, and check new findings that may deeply impact on your study. It is also recommended to know who (which group) is a leading person (research group) in that area through your

literature review. By doing this, you will not miss important papers published in the domain of the topic.

During your literature search, you may also find unanswered research questions, come up with new research hypotheses, or get ideas on how to conduct your study.

2) How to do a literature search

If you belong to a university or academic institute, your librarians are good advisors for your literature search. However, I would like to mention some general tips for literature search through the Internet.

STEP 1. Defining the question; Literature search is different from browsing journals and websites. You should have specific research question(s) before searching. Otherwise, you will waste your time.

STEP 2. Searching; Once you define your research question(s), pick up proper keywords to use for the search. Attempt to use a wide range of relevant sources. Recommended online information sources are listed below.

· PubMed is the biggest database of medicine and life sciences, including MEDLINE, provided by National Library of Medicine (USA)
http://www.ncbi.nlm.nih.gov/sites/entrez

· Embase holds millions of indexed records from thousands of journals, including all of MEDLINE as well as over 5 million records and 2,000 biomedical journals not currently covered by MEDLINE. It is provided by Elsevier.
http://www.embase.com/

· Free Medical Journals was created to promote the free availability of full text medical journals (3780 journals) on the Internet.
http://www.freemedicaljournals.com/

· Popline provides access to 350,000 selected publications and resources related to family planning and reproductive health. It is supported by USAID (the United States Agency for International Development).

    http://www.popline.org/

·   HINARI is supported by the WHO and major publishers. It enables developing countries to gain access to one of the world's largest collections of biomedical and health literature. Up to 11,400 journals (in 30 different languages), 18,500 e-books, 70 other information resources are now available to health institutions in more than 100 countries (including Vietnam), areas and territories.

    http://www.who.int/hinari/en/

3) PubMed at a glance

When you open the PubMed page, you see interface like the Figure 2.1. This could change over time, but its basic structure is maintained. After opening the PubMed page, enter key words into a box on the top, click "go", and you will see a total number of hits (Figure 2.1). In this case, it is 147. If you find an attractive title, click the title, and you will see an abstract (Figure 2.2). In some cases, you can retrieve a full paper by clicking an icon on the upper right hand corner of the abstract page.

Figure 2.1. PubMed: Entering keywords

Figure 2.2. PubMed: Reading an abstract

Click an icon like this to access to a full paper.

Click "See all" to obtain a list of related citations.

4) Refining the search

You may feel overwhelmed at the long list of search hits. To refine searched items, you can apply the following methods; use more specific keyword(s), use a combination of keywords, or refine by article type, study subjects and publication year (Figure 2.3). You may also use journal title, publication year and authors' name as keywords. Supposedly, you would like to find Dr. Matsumura's papers on cancer published in "Lancet" in 2010. You can type "matsumura cancer 2010" in the keyword box. There are more resources, such as "Clinical Queries", which is a good tool for clinicians to find references related to clinical medicine (Figure 2.4).

Here are a few practical tips:

· Combine key words using operators (AND/OR): For example, if you type "diabetes AND hypertension", you will get a list of articles that discuss about both diabetes and hypertension. The list of hits will be shorter than only typing "diabetes" (or "hypertension"). If you type "diabetes OR hypertension", you will get a list of articles about diabetes or

hypertension. The list will be longer than only typing "diabetes" (or "hypertension").

· Utilize limit functions: For examples, limit your search to;

- Past 10 years.
- Review articles if you are new in the field or want a quick over-view of the topic.
- "Free full text available" articles. Then, you can read the full text of listed articles.

Figure 2.3. PubMed: Refining search
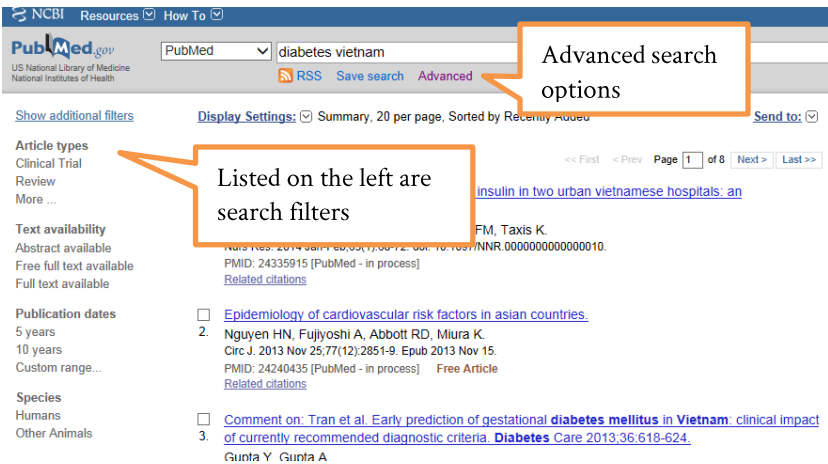
Figure 2.4. PubMed: Using clinical queries
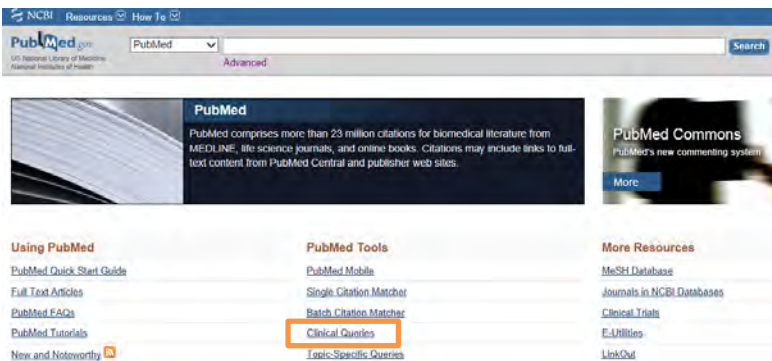
5) Keeping records

After searching for references, keep your records to avoid repetitive searches. It is recommended to develop an <u>annotated bibliography</u>. It is a reference list with your notes about the paper. You can organize and compile your search hits into your own 'library'. There are number of useful software/tools that are free and commercialized.

· PubMed "My NCBI": http://www.ncbi.nlm.nih.gov/pubmed
  (Need to register. Click an icon in the right upper corner of PubMed screen.)
· Zotero (Free software): http://www.zotero.org
· Endnote (Commercialized): http://endnote.com

These tools help you to: 1) extract/export/import articles; 2) organize and store them in a 'library'; 3) and most importantly: properly formatting the references when writing a manuscript, report or protocol. Commonly used reference styles are Vancouver or Harvard styles. Tools like EndNote will automatically format references in an assigned style.

· Example of Vancouver Style
  Text: A study has been completed to determine the prevalence of probable depressive state among mothers in Vietnam.1
  Reference:
  1. Suzuki Y, Goto A, Nguyen QV, Nguyen TTV, Pham NM, Chung TMT, Trinh HP, Pham VT, Yasumura S. Postnatal depression and associated parenting indicators among women. Asia-Pacific Psychiatry. 2011; 3: 219–227.
· Example of Harvard Style
  Text: A study has been completed to determine the prevalence of probable depressive state among mothers in Vietnam (Suzuki et al. 2011).
  Reference:
  Suzuki, Y., Goto, A., Nguyen, Q.V., Nguyen, T.T.V., Pham, N.M., Chung, T.M.T., Trinh, H.P., Pham, V.T. & Yasumura, S. (2011)

Postnatal depression and associated parenting indicators among women. *Asia-Pacific Psychiatry*, vol. 3, pp. 219–227.

6) Further information sources of EBM

There are other useful sources for evidence-based medicine providing you with results of systematic reviews. The Cochrane library is extremely helpful for clinicians to gain up-to-date overview of topics (prevention, diagnosis, treatment) of your interest due to its systematic reviews.

- The Cochrane Library
  http://www.thecochranelibrary.com/view/0/index.html
- BMJ Clinical Evidence
  http://clinicalevidence.bmj.com/x/index.html
- TRIP Database
  http://www.tripdatabase.com/

2. Critical appraisal of scientific articles

Critical appraisal is a systematic approach to read, understand, interpret, assess study validity, identify study limitations, and decide upon the usefulness of results (applicability) of scientific papers. General key points of reviewing a paper are listed in Box 2.1, and guidelines of major study designs are listed in Table 2.1. If this is the first research methodology textbook you are reading, please read the box and table after reading of the main text of this book. For more information, critical appraisal sheets are available at various websites as listed below. It is recommended to carry out a regular (weekly, bi-weekly or monthly) journal club at your work place with your colleagues to share updated medical evidence in order to improve your clinical practice.

- Center for Evidence Based Medicine, University of Oxford
  http://www.cebm.net/index.aspx?o=1157

· McMaster University
http://fhswedge.csu.mcmaster.ca/cepftp/qasite/CriticalAppraisal.html
· University of South Australia
http://www.unisa.edu.au/Research/Sansom-Institute-for-Health-Research/Research-at-the-Sansom/Research-Concentrations/Allied-Health-Evidence/Resources/CAT/
· The College of Family Physicians of Canada
http://www.cfpc.ca/uploadedFiles/CPD/Pearls/Pearls%20for%20Residents%20Critical%20Appraisal%20Sheet.pdf
· Royal College of Psychiatrists, UK
http://www.rcpsych.ac.uk/pdf/app2.pdf
· Stanford School of Medicine
http://peds.stanford.edu/Tools/documents/Critical_Appraisal_Form_CGP.pdf.

Box 2.1. How to do a critical appraisal

1. Research question
The research question is what the researchers attempted to answer in the paper, and it should clearly appear in the introduction.
  1) How does the question relate to findings in previous studies?
  2) Is the question original, or has it been already raised in other studies?
  3) Is the question rational?
2. Study design and population
  1) What type of study design was applied?
  2) Is it appropriate to answer the research question? Please see the section on study designs in epidemiological studies to learn the advantages and disadvantages of each design in Chapter 4 to 8.
  3) Which population was selected?
  4) Is the population relevant to the research question(s)?
  5) Is there a possibility of selection bias? (Also see the section of selection bias in Chapter 3) Were all people in the target population invited to participate, or was it a random sample? If not, is selection

process clearly explained?

6) What was the participation rate?

7) Are the inclusion and exclusion criteria clearly mentioned?

8) (In a case-control study) What are the definitions of cases and controls?

9) Are the results obtained in this study generalizable to other populations?

10) How large is the sample size? (How many subjects were recruited?)

11) Has a power calculation been conducted? (in the case of clinical intervention)

12) (In an intervention study) Is the intervention thoroughly explained?

13) Is the development or construct of a questionnaire thoroughly explained?

3. Outcome and study factors

The main purpose of epidemiological studies is to examine a causal association between the exposure(s) (study factors) and the outcome. The outcome and study factors should be relevant to answer the research question.

1) What are the outcome variables and its definition (how, when, and by whom)?

2) What is the estimate of the outcome? (Mean, median, prevalence (%), incidence rate, odds ratio, hazard ratio, etc.)

3) What are biases and confounding factors? (See the section of bias and confounding in Chapter 3)

4. Biases and confounding factors

Biases are potential systematic errors in any processes of the study, such as recruitment of study subjects, data collection, analyses, publication and the combination of these factors, which leads to a conclusion apart from the truth. (Also see Chapter 3)

1) Is there a systematic error in each process (study design, data collection, analyses, publication and the combination of these processes)?

2) Did the author(s) carefully evaluate and discuss biases (the magnitude and direction of the deviation) if any?

3) In clinical interventions, has/have the author(s) assessed the completeness of follow-up (the number of dropouts and their reason)? If, for example, the dropout rate in treatment A group was higher than

  that in treatment B because of severe side effects of treatment A, the conclusion of this study is biased.

 4) In Randomized Controlled Trials (RCT), were the participants and assessors, who evaluate clinical outcome(s), blinded? It is preferable for both to not know whether the participant has received the trial medicine or placebo.

5. Statistical analysis

 1) Is the framework of the analysis clear-cut and theoretical?

 2) Are statistical methods adequately described in the methods and results?

 3) Are statistical tests appropriate to evaluate the association between the outcome and exposure?

 4) Are the effects of confounding factors adjusted properly?

6. Ethics

There might be ethical concerns even after peer reviews.

 1) Has the study protocol been reviewed and approved by an independent ethics committee? If the review is waived, is it clearly explained?

 2) Have the authors obtained appropriate informed consent from the participants if necessary?

 3) Are there any other ethical concerns?

7. Statistical results and interpretation

 1) Did the authors present and interpret their results in accordance with their research question?

 2) Do you see the appropriate risk estimates, their corresponding confidence intervals and p values?

 3) Are there any other possible interpretations?

 4) Have the author(s) made a causal inference?

8. Limitations

Every study has limitation(s) in varying degrees. The important point is whether or not the authors were aware of their limitation(s), and have interpreted the findings taking them into account.

 1) Did the authors discuss the limitation(s) in their study?

 2) Are there any over or under estimations in their interpretations?

3) Are there any other possible limitations?

9. Conclusion and applicability

The authors should make a conclusion based on the appropriate interpretation of their findings. Applicability may differ among clinical situations. We have to ask ourselves: can I generalize this finding in my practice?

Table 2.1. Studies design guidelines

Acronym	Study design	Website
CONSORT	Randomized control trial	http://www.consort-statement.org/
STROBE	Observational studies	http://www.strobe-statement.org/
MOOSE	Meta-analysis of observational studies	http://www.equator-network.org/?o=1073
PRISMA	Systematic reviews and meta-analysis	
STARD	Diagnostic test	http://www.stard-statement.org/
STREGA	Genetic associations (Extent of STROBE)	http://www.medicine.uottawa.ca/public-health-genomics/web/eng/strega.html

Bias and Confounding

Chihaya Koriyama, Tran The Trung

1. Error and misclassification

According to the Dictionary of Epidemiology (edited by Last), a bias is a "deviation of results or inferences from the truth or processes leading to such deviation." When you carry out epidemiological studies, it is important to spare no effort in the prevention of bias at any stage of the study design. In order to understand biases, you need to understand error. Here are two types of errors. Comparison of these two concepts is presented in Figure 3.1.

- Random error; When we say "error", it means a random error. Since an "error" occurs randomly, the mean measured value will come close to the true value with the increasing of the number of measurement.

- Systematic error; A systematic error is called a "bias". Since a "systematic error" has a tendency (large/higher or small/lower) in the measurement, the mean measured value does not come close to the true value even if you measure a thousand times..

Figure 3.1. Comparison of random and systematic error

Random error Measured value (mm)	Systematic error Measured value (mm)
53	48
47	48
48	48
49	48
51	48
52	48
51	48
Mean=50	48

God knows that the true value is 50mm.

Similar terms that are commonly used are non-differential and differential misclassification. The definitions are as below. Calculations in Box 3.1 explain these two concepts in detail. If this is the first research methodology textbook you are reading, please read the box after reading of the main text of this book.

- Non-differential misclassification; Within a comparison between groups (populations), a systematic error may not be a critical problem as long as it occurs in all comparison groups (populations) in a same manner, which is called "non-differential misclassification". In this case, however, the risk estimate will approach null.

- Differential misclassification; If the error occurs only in one specific group due to bias, the risk estimate deviate from null. When you are concerned about potential bias(es) in your study, careful evaluation is required to see if there is an under or over estimation.

Box 3.1. Example calculations of non-differential and differential misclassifications

Suppose that 50 out of 60 in an exposed group and 50 out of 140 in an un-exposed group were diagnosed as cases. (Note that the sensitivity and specificity of the diagnosis are 100% for both.) Distributions of cases and controls are as follows:

	Exposed	Un-exposed
Case	50	50
Control	10	90

The relative risk in this setting is

$$\frac{(50 / 60)}{(50 / 140)} = 2.3$$

If the sensitivity and specificity of case diagnosis are 80% and 90% respectively, 10 out of 50 cases will be misclassified as controls (80% sensitivity), and 1 out of 10 controls will be misclassified as cases (90% specificity) in the exposed group. This is also true in the unexposed group because of non-differential misclassification. Thus, the case-control distribution will be as follows;

	Exposed	Un-exposed
Case	41	49
Control	19	91

Thus, the relative risk in this setting is

$$\frac{(41 / 60)}{(49 / 140)} = 1.95$$

The observed risk estimate comes close to 1 (null) if non-differential misclassification exists, which means that the observed estimate is an underestimated value.

On the other hand, the deviating direction of an observed risk estimate will vary depend on the setting of differential misclassification. Consider two different settings.

a) Suppose that the sensitivity and specificity of case diagnosis are 80% and 90% respectively, in the exposed group ONLY. (Cases are perfectly diagnosed in unexposed group) The case-control distributions will be as follows:

	Exposed	Un-exposed
Case	41	50
Control	19	90

Therefore, the relative risk in this setting is.

$$\frac{(41 / 60)}{(50 / 140)} = 1.91 < 2.3 \text{ (true estimate)}$$

b) On the other hand, suppose that the sensitivity and specificity of case diagnosis are 80% and 90% respectively, in unexposed group ONLY. (Cases were perfectly diagnosed in exposed group) The case-control distributions will be as follows:

	Exposed	Un-exposed
Case	50	49
Control	10	91

The relative risk in this setting is.

$$\frac{(50 / 60)}{(49 / 140)} = 2.38 > 2.3 \text{ (true estimate)}$$

As you can see, the relative risk in setting A is underestimated, but is over-estimated in setting B. Thus, you need to evaluate the direction of your results as much as possible.

2. Type of bias

There are different types of bias in research procedures.

1) Selection bias

Selection bias is introduced by recruiting study subjects based on some characteristics which are likely to be related to the exposure (your study factor) and/or outcome. For example, in estimating the risk of lung cancer by cigarette smoking, you need to recruit appropriate control subjects (without cancer history). If you recruit controls from participants who attended annual health checkup, the proportion of smokers in controls will be lower than that in the population since participants of health checkup are health-conscious in general. The result will be likely be an overestimate.

Selection bias is also likely to occur in a study with a low participation rate. At the health checkup in the above, we asked them to participate voluntarily, and only health-conscious people tend to take part in the survey. Thus, a low participation rate may enhance the overestimation.

2) Detection bias

Detection bias occurs in case ascertainment or diagnosis. For example, a doctor may examine the patient's chest X-ray more carefully if he knew the patient was a smoker but not for the non-smoking patient. To avoid or minimize the detection bias, a common protocol of making a diagnosis and case ascertainment is required. For laboratory examinations, it is preferable for the assessment to be blinded.

3) Recall bias

Recall bias is a systematic error because of differences in accuracy or completeness of the recall to memory of past experiences. This bias tends to occur especially in case control studies. For example, a mother of child with atopic dermatitis is more likely than the mother of a healthy child to remember the details of diets during the pregnancy, breast-feeding, weaning age, and family history of atopic diseases. Consequently, the association between these factors and child's atopic dermatitis will be exaggerated. One of

the strategies to minimize recall bias is to recruit a hospital control (or out-patient) with other diseases.

3. Confounding

In some textbooks, confounding is explained as one of the biases since it leads to systematic errors. In others, however, it is distinguished from bias because the effect of confounding can be adjusted to some extent by adequate statistical analyses, though this effect of bias cannot be corrected after the completion of a survey. Although there is a conflicting classification of confounding (in terms of bias), the definition of confounding is the same.

1) Definition of confounding

Confounding is a distortion of the association between an exposure and the outcome by other factor(s). These factors are called "confounders" or "confounding factors". To be concrete, a confounding factor meets the following three requirements.

i) A confounder is a risk factor of the outcome.
ii) A confounder is related to the exposure (It could be by chance)
iii) A confounder is NOT in the process of the association between the exposure and the outcome

2) Examples of confounding

In the scenario presented in Figure 3.2, we wonder whether exposure to radiation in uterus causes high infant death. Living in high background radiation area (HBRA) is a confounder since i) the socio-economic status of this area is low, which leads a high infant death, ii) living in HBRA causes an exposure to radiation in uterus among women, and iii) living in HBRA is

not in the process of causal association between radiation exposure in uterus and infant death.

Figure 3.2. Example of confounder: Living in HBRA

HBRA=High back ground radiation area
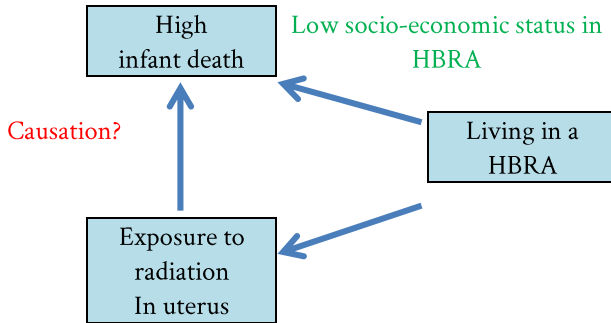
Let's see another example of confounding. In the scenario described in Figure 3.3, we wonder whether or not radiation exposure causes myocardial infarction (MI) (suppose you observe a statistically significant association in your data). Smoking is a confounder since i) smoking is a well-known risk factor of MI, ii) smoking was related to the radiation exposure by chance (you do not have any good reason to explain this relation), and iii) smoking is not in the process of causal association between radiation exposure and MI.

Figure 3.3. Example of confounder: Smoking
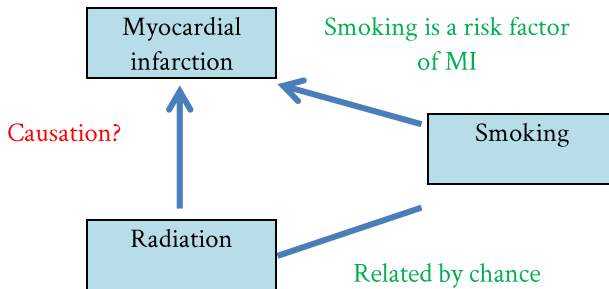
3) Prevention of confounding

There are three ways to prevent the effect of confounding at the stage of study design.

· Limitation: For example, sex is a common confounder in epidemiological studies. If you limit the study subjects to female, the effect of confounding by sex is completely prevented.

· Randomization: This method is applicable for intervention studies, and you do not need to specify confounders.

· Matching: Although this method is applied in both case-control studies and cohort studies, the effect of confounding can be prevented only in cohort studies.

4) Diagnosis of confounding

There are two ways to check confounders, theoretically or mathematically.

· Theoretically, the confounder candidate should meet the criteria of as above.

· Mathematically, you may compare the risk estimates between pre- and post-stratification by the confounder candidate. If you see a discrepancy in terms of risk estimates, it is likely to be a confounder. Let's see an

example of a lung cancer case-control study. If this is the first research methodology textbook you are reading, please check the calculation after reading this book thoroughly.

	Lung cancer	Control
Alcohol volume		
High	33	1,667
Low	27	2,273

Odds ratio = $(33{*}2273) / (1667{*}27) = 1.67$

Alcohol consumption seems to be a modest risk factor of lung cancer. After stratification by smoking status,

	Smokers		Non-smoker	
	Lung ca.	Control	Lung ca.	Control
Alcohol volume				
High	24	776	9	891
Low	6	194	21	2,079
Odds ratio	$24{*}194 / 776{*}6$		$9{*}2079 / 891{*}21$	
	$= 1$		$= 1$	

The association between alcohol consumption and lung cancer risk is gone after stratification by smoking status. Thus, smoking is a confounder in this example.

5) Treatment of confounding (methods to obtain adjusted risk estimates)

The effects of confounding can be removed to some extent using adequate statistical methods. These advanced methods are beyond the scope of this textbook, and will be explained in the course.

- Stratification: After stratification, a common odds ratio among strata is calculated (Mantel-Haenszel odds ratio).
- Statistical adjustment: Multiple regression models are used to adjust the effect of confounding by including confounders as covariates.

6) Effect modification

It is not always to combine multiple strata together. When the strength of association (relative risk, odds ratio) between exposure and outcome greatly differs between each stratum, you should not combine them. In other words, when you see effect modification, it is better to present stratified data.

Descriptive Studies

Hirohide Yokokawa, Tran Viet Thang

1. Introduction of descriptive studies

A <u>descriptive study</u> is an observational study, as it observes and describes the patterns of disease (outcome) occurrence in a population. When describing the outcome occurrence, the three "major key points" to be considered are person, place and time. It is an important tool to diagnose a community and a group of people, and to investigate an emerging health event. In recent years, computerized geographical analysis (geographic information system, GIS) is becoming rapidly popular as a powerful to tool to graphically capture outcome occurrences.

We are often tempted to do an intervention study just after reading articles from high impact journals. However, we must recognize a way to go through the "stairs of epidemiology". We have to study characteristics of our target population and the prevalence of our target outcome prior to estimating an association between exposure and outcome. Just like John Snow's case in Chapter 1, a detailed descriptive analysis of the population and outcome may provide important information to carry out a public health action toward the target population. Therefore, a descriptive study is an important initial step that should not be skipped. We can then carry out (analytical)

cross-sectional study, case-control study or cohort study to estimate a true association between exposure and outcome. It is only after identifying risk or preventive factors of the outcome, we can design an intervention study to estimate the effectiveness of prevention.

Figure 1. Overview of study designs and their step-up

Descriptive studies
To investigate characteristics and prevalence

Cross-sectional studies
To investigate a possible association between exposure and outcome

Case-control studies, Cohort studies
To estimate an association between exposure and outcome

Intervention studies
To estimate an effectiveness of intervention
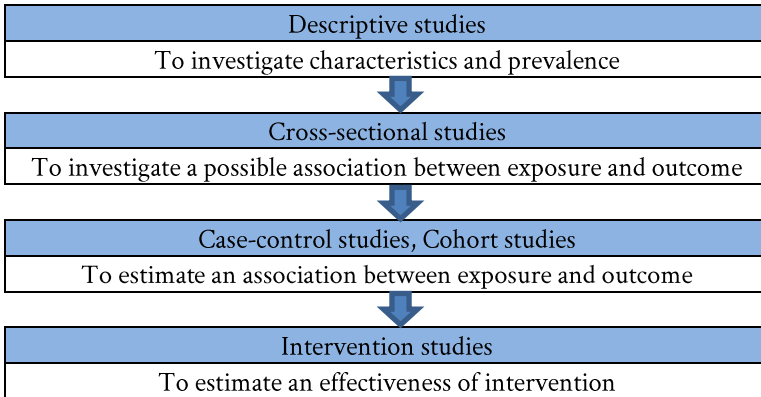
2. Types of descriptive studies

Descriptive studies include case reports, case series, a cross-sectional studies and surveillance studies that analyze individual-level data. Another type of descriptive study is ecological study that examines population-level data.

One major limitation of the descriptive study is not being able to estimate a causal relationship between exposure and outcome. However, the study may provide a hypothesis which can be tested by analytical observational studies.

1) Case report

A case report describes the experience of a single or several patients with a similar diagnosis. These subjects present unusual diseases or unusual features of a disease. They can be expected as "the first clues" in the identification of a new diseases or adverse effects of an exposure. Case report is quite common in clinical journals. They are useful to public health as they can provide an interface between clinical medicine and epidemiology.

2) Case series

A case series describes a series of patients with an outcome of interest, and does not involve a "control group". It can be considered as a collection of individual case reports. This study design has historical importance in epidemiology. It was often used as an early means to identify the beginning or presence of an epidemic. It can also potentially serve as a case group for a case-control study.

3) Cross-sectional (prevalence) study

A cross-sectional study is an observation of a defined population at a single specific point in time or time interval. Exposure and outcome are measured at the same time. Advantages of the study include minimal costs, and it is a quick way to estimate prevalence (an indicator of frequency of existing cases). The major disadvantage is, as mentioned above, a causal relationship cannot be estimated as well as incidence (an indicator of occurrence of new cases). Note that a cross-sectional study can either be descriptive without analyzing exposure-outcome associations or analytical with the analysis.

4) Ecological study (or ecological correlational study)

Ecological studies can help you to look for a potential association between exposure and outcome at the population-level rather than at the individual level. It usually is a secondary analysis of existing data.

3. Prevalence and incidence

The prevalence is the proportion of a population with the health event of your interest. It is the number of people with the health event over the total number of people studied and expressed as a percentage or as the number of cases per 1,000, 10,000 or 100,000 people. The incidence is an occurrence measurement of a new health event within a specified period of time. It is the number of new cases over the number of population in a given time period. While prevalence can be estimated from a cross-sectional study, estimating incidence requires a follow-up of a cohort.

4. Study example

As a diabetes specialist working in Vietnam, I heard about an increase in metabolic syndrome and diabetes in the country from Vietnamese colleagues. Through literature searching, I came to know that not much evidence on the issue was available. I therefore formed a research team, planned and conducted a study to investigate characteristics of metabolic syndrome and its clinical components among diabetic Vietnamese patients in Ho Chi Minh City.[1]

Information was collected from 652 outpatients who were recruited from one public general hospital (People's Hospital 115) and one private clinic (Medic Center) in Ho Chi Minh City, Vietnam. We evaluated the collected information descriptively and diagnosed metabolic syndrome according to the International Diabetes Federation (IDF) Worldwide Definition of the Metabolic Syndrome. Metabolic syndrome was observed in 39.4% of

---

[1] Yokokawa H, Goto A, Watanabe K, Yasumura S. Internal Medicine Journal. 2007; 237: 161-167.

men and 70.5% of women. Our study revealed a high prevalence of metabolic syndrome in diabetic Vietnamese patients, especially among women, and revealed central obesity as a key feature.

Cross-sectional Studies

Yuriko Suzuki, Nguyen Thi Boi Ngoc

1. Basic concept and study purposes

Cross-sectional studies examine the prevalence of disease or problems (outcomes) in a defined population, and associations between variables and the outcome at a single time point.

1) To know the prevalence of outcome

The purpose of a cross sectional study is to calculate the prevalence of an outcome in a defined population. There are point prevalence and period prevalence. With a cross-sectional study, the frequency of the outcome is described, and it can be further stratified by factors, such as person (gender, age, work, lifestyle, etc.), place (region, country, etc.), time (season, year, etc.). By knowing its prevalence, it is possible to examine the burden of an outcome or a disease and compare it with other groups of people. Information of the prevalence of a disease is particularly useful for a chronic condition with a long disease course. By examining the prevalence over the year with serial cross-sectional studies, the trend of the health problem can be examined. These are very simple information, but it is an important first step in developing a public health strategy.

2) To generate a hypothesis on outcome and exposure and foundation for further studies

Another purpose of cross-sectional studies is to identify the association between outcome and exposure that may have effects on the outcome. Traditionally, the association of outcome and exposure was simply examined; these days however, the association useful for forming public health policies is examined. Examples include:

- Association between an outcome and risk factors (e.g., postnatal depression and pregnancy related factors)
- Association between a health problem and service utilization, and its international comparison
- Association between service utilization and knowledge, attitude and belief of health problems

2. Designing

To develop a protocol for a cross-sectional study, it is important to carefully consider sampling of study population and explicitly define the outcome. Sampling requires the consideration of a denominator, while defining an outcome requires the consideration of the numerator of the prevalence. Inappropriate sampling may cause sampling bias, and a rudimentary definition of an outcome may cause a measurement bias.

1) Sampling

It is not feasible to conduct studies on the entire target population given limited time and resources. In practice, a study population is drawn from the target population, and study participants consist of a part of the study population because of those who cannot be reached, declined to participate in the

study, or due to other reasons. Whether the statistical results drawn from the study participants can be inferred to the target population (<u>internal validity</u>) depends on the appropriateness of the sampling. The premise of whether the study results can be generalized to a wider population outside the target population is defined as <u>external validity</u>. Other aspects, such as the place of recruitment (community or clinic setting), and duration of the study are critically important in examining the generalizability of the study.

Random sampling aims each person in the target population to have the same probability of being chosen, and it is preferred to minimize differences between the sampled and non-sampled. In the real world, you can have access only to a reachable group, as per convenience samples as explained in Chapter 1. In this case, you should pay attention to characteristics of your study site (primary care clinic, tertiary hospital, specialized care unit, etc.) and scrutinize the generalizability of the results from the convenience sample.

The effort to minimize non-response rate is essential since those who meet eligibility criteria and do not participate may differ in terms of important characteristics from those who choose to participate. You should work hard to reduce the percentage of refusals through among participants. In reporting the study results, it is the best if you could present a data to show the comparability of those who participated and refused, so that the readers can examine a potential bias.

2) Defining variables

In writing a study protocol, you should define and operationalize three types of survey items (variables); outcome, exposure, and confounder. These variables should be clearly defined before data collection.

To begin, defining the <u>outcome</u> is a difficult task. In the context of mental disorders for example, there are gold standards of diagnostic guidelines such as the International Classification of Diseases (ICD), and the Diagnostic

and Statistical Manual of Mental Disorders (DSM, America Psychiatric Association), yet it is still difficult to describe what the disease or disorder is, because there is no clear line between what constitutes normal and abnormal, and there is usually some form of gradation between the two extremes. However, in research, it is usually required to set an arbitrary line to differentiate 'normal' and 'abnormal'. Researchers should therefore explicitly define the outcome of the interest.

Exposure is a factor which may have effects on the outcome. Based on clinical observations and a review of existing literature on the related topic, researchers should select variables that seem to have a relationship with outcome to be examined. Examples of exposure variables include person (gender, age, work, etc.), place (region, country, etc.), time (seasonality, year, etc.), lifestyle and genetic factors, to name a few.

Confounding factor is a third factor that is independently associated with both the outcome and exposure. Taking an example from cervical cancer of uterus and sexual activity, these two factors seem to have an association. However, behind these two factors, there is a third factor, the human papilloma virus that is independently associated with both cervical cancer and sexual activity. Without controlling for this factor, one can wrongly conclude that there is a direct association between cervical cancer and sexual activities. In many epidemiological studies, age, gender, socioeconomic status, and smoking habit are treated as confounders during study design.

3. Collecting data

Methods of data collection include mail-based survey of self-administered questionnaires, face- to-face interviews, chart reviews, disease registries etc. Depending on the variables described above, an appropriate data source or measurement should then be selected. For example, if the

primary outcome of the study is the prevalence of a disease, a disease registry or chart review in which diagnoses of disease confirmed by professionals is more appropriate than self-administered questionnaires by patient. If the main outcome is a sensitive or subjective issue, such as erectile function among men, face-to-face interviews are more instrumental than self-administered questionnaires.

In terms of measurement, it is important to confirm and present information on the reliability and validity of the chosen mode of measurement. Reliability or repeatability is the degree to which the same results can be reproduced by the same measuring method or the same observer. Different measuring procedures or its implementation can result in low reliability. Validity is the degree to which how accurately a selected mode of measurement can assess the object. This includes construct validity, content validity, and criterion validity. For an example of content validity, in the case of measuring subjective assessment of quality of life with a measurement that was developed in different culture and language, it is required to use a validated scale in the culture and language in which the research is being conducted. Especially in the case of using a scale that has been originally developed in Western culture, forward and backward translations of the original questionnaire and examination of reliability and validity should be completed before using the scale in the study. Often issues of copyright and cost will arise, and this should also be addressed when preparing a protocol.

4. Presenting results

Below is a list of indicators that you can calculate from a cross-sectional study.

1) Prevalence

- <u>Point prevalence</u> = number of cases of disease or health problem (outcome) in a defined population at a certain point / number of people in a defined population at a certain point
- <u>Period prevalence</u> = number of cases of disease or health problem (outcome) in a defined population at a certain point / number of people in a defined population during a defined time period

2) Association indicators

A tabulation (cross table) of outcome and exposure helps to examine its association.

	With outcome	Without outcome
Exposed	a	b
Unexposed	c	d

- Overall prevalence of outcome = a+c / a+b+c+d
- Prevalence of outcome in exposure group = a / a+b
- Prevalence of outcome in non-exposure group = c / c+d
- <u>Risk ratio</u> (RR) = Prevalence of outcome in exposure group / prevalence of outcome in non-exposure group = (a/a+b) / (c/c+d)
- <u>Odds ratio</u> (OR) = (a/b) / (c/d) = ad / bc

If OR is 1, the exposure is not related to the outcome. If OR is greater than 1, the exposure is positively related to the outcome occurrence, that means having the exposure increased the risk of having the outcome. If OR is less than 1, the exposure is negatively related to the outcome occurrence, that means having the exposure decreased the risk of having the outcome.

When prevalence is very low, the odds ratio is close to risk ratio. To calculate statistical significance of the association, you can use Chi-square test or Fisher's exact test. These tests are explained in Chapter 10.

5. Strengths and limitations

In cross-sectional studies, the frequency of outcome is captured at a single point of time. This is relatively cheap, feasible and may contribute to immediate responses to the given health problem. Generally, evidence level of cross-sectional study is ranked as low, but it acts as foundation for other types of studies such as cohort study and intervention study. Because cross-sectional studies give information on disease burden of the population, this can be a description of the baseline status for further studies.

The major limitation of cross-sectional studies is how causal relationships cannot be determined due to one-time nature. Thus, always consider a possibility of reverse causality. You also need to be cautious if this study design captures existing cases, and not newly onset cases. A high prevalence may imply high incidence of the disease, but there is also possibility that the condition has a long duration due to the nature of the disease or ineffective treatment.

6. Study example

An example of a cross sectional study to derive the prevalence of a disease and associations between the variables is as follows. A study has been completed to determine the prevalence of probable depressive state among mothers in Vietnam, and to examine its risk factors with respect to social support and maternal childrearing attitude.[2] In this study, study participants were 299 mothers who visited a tertiary hospital in Vietnam for regular

---

[2] Suzuki Y, Goto A, Nguyen QV, Nguyen TTV, Pham NM, Chung TMT, Trinh HP, Pham VT, Yasumura S. Postnatal depression and associated parenting indicators among Vietnamese women. Asia-Pacific Psychiatry. 2011; 3: 219–227.

check-ups between 1 and 3 months postpartum, and they were consecutively invited for the study. The main outcome was depression assessed by a two-question case-finding instrument for depression.[3] The questionnaire was administered by a trained research pediatrician in October through December 2007. In terms of results, the prevalence of mothers with probable depressive state was 23.1% (69 of the 294 mothers). Risk factors for probable depressive state were parental and familial conflict and recent moving. In terms of childrearing attitude, the following factors increased the risk of mothers being in a probable depressive state - a lack of confidence (adjusted odds ratio=2.74, 95% confidence interval: 1.40-5.38), and less relaxed feeing (adjusted odds ratio=2.85, 95% confidence interval: 1.21-6.71) after controlling for subjective health. From the findings, mothers with low confidence levels and those feeling less relaxed in childrearing are at increased risks of being in a probable depressive state. Thus we concluded that social support factors may have played a role in depression. For practical implications, a brief screening tool for depression is useful to identify those in need of additional support when rearing a child in Vietnam.

---

[3] Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. Journal of General Internal Medicine. 1997; 12: 439-45.

Cohort Studies

Nguyen Quang Vinh, Nguyen Thi Tu Van

1. Basic concept and purpose

A cohort is a group of people sharing a common condition. A <u>cohort study</u> is a follow-up study of group(s) that is initially free from the studied outcome at the start of an investigation to assess the incidence of the outcome(s) of interest over time. The required length for follow up should be longer than the length of the latent period of the outcome of interest. In cohort studies, groups are established by the presence or absence (2 groups), or different levels of selected exposure (>2 groups). The goals of cohort studies are to measure (always), and compare (usually) the incidence of outcome in one or more study cohorts.

2. Designing

1) Prospective or retrospective

A cohort study is classified as either prospective or retrospective depending on the temporal relationship between the initiation of a study and occurrence of the outcome(s) of interest.

·   <u>Retrospective</u>: Occurrence of outcome → Study initiation
·   <u>Prospective</u>: Study initiation → Occurrence of outcome

The choice of retrospective or prospective is a trade-off between the scientific value of the study and its feasibility. For prospective cohort studies, biases can be minimized. However, this is time consuming, and also expensive. In the case of retrospective cohort studies, they can be carried out more quickly with less budget, although this depends on the availability of adequate records.

Example 1. An analytic prospective cohort study: Assess pregnant women's intention to get pregnant (exposure), follow-up till postpartum, and ask about maternal confidence (outcome).

Example 2. An analytic retrospective cohort study: Assess graduation status of medical school graduates (outcome) and review their entrance examination scores (exposure).

2) Selection of a cohort

In a cohort study, you can specifically select exposed and non-exposed cohorts when sampling, or you can select one cohort and categorize it into exposed and non-exposed during analysis. For common variables of interest such as smoking and alcohol consumption, a cohort can be easily sampled from a general population. For rare variables of interest including occupational exposure, disaster, dioxin and ionizing radiation, sampling should target a specific cohort. In this case, you should be cautious about the characteristics of the selected cohort. For occupational exposure, workers are generally healthier than the general population (<u>healthy worker's bias</u>).

3) Variables

The exposed and non-exposed groups should be comparable except for the exposure status under investigation, but it is not always the case. Information on confounders (e.g. socioeconomic, demographic, geographic char-

acteristics) and other co-risk factors (e.g. alcohol, smoking, nutrition) should be collected and taken into account during the analyses.

4) Follow-up

Securing a high <u>follow-up rate</u> is very important in a cohort study. The longer the observation period is required, the more difficult it will be to achieve a satisfactory follow-up rate. Failure to follow-up is a major issue leading to bias. Subjects with better compliance to the survey are often healthier, and those who drop out often end in negative outcomes without being recorded.

3. Collecting data

As mentioned in the previous chapter, a variety of data sources can be considered depending on the study objectives - interviews, questionnaires, medical examinations, laboratory tests, environmental measurements, medical charts or other information database/records. You can design the methods for data collection in a prospective study, while existing data are used in a retrospective study.

4. Presenting results

1) Risk calculation

Calculations are based on the 2x2 table.

Note: E+= positive exposure; E-=negative exposure

	With outcome	Without outcome
Exposed	a	b
Unexposed	c	d

- Incidence E+ (IE+) = a/(a+b).

  IE+ is the risk of the outcome in E+ group.

- Incidence E- (IE-) = c/(c+d).

  IE- is the risk of the outcome in E- group.

- Total Incidence (I) = (a+c)/(a+b+c+d).

  I is the risk of the outcome in study population.

- <u>Relative Risk</u> (RR) = Risk Ratio (RR) = IE+/IE- = [a/(a+b)]/[c/(c+d)].

  RR is the relative difference of risk between E+ and E- groups.

- <u>Excess Risk</u> (ER) = IE+ - IE- = a/(a+b) - c/(c+d).

  ER is the absolute difference of risk between E+ and E- groups.

- <u>Attributable Risk</u> (AR) = ER/E+ = [a/(a+b) - c/(c+d)] / [a/(a+b)]

  = (RR-1)/RR

- <u>Population Risk</u> (PR) = (IE+) x P + (IE-) x (1-P).

  PR is the risk of the outcome in a population.

- <u>Population Attributable Risk</u> (PAR)

  = Excess Risk x Prevalence of exposure in a population (P)

  = (IE+ - IE-) x P

  PAR is the difference of risk between E+ and E- in a population.

- <u>Population Attributable Fraction</u> (PAF)

  = PAR/Outcome incidence in a population.

  PAF is the proportion of outcome among study population is due to the risk.

2) Person-time concept

In some studies, an exposure is continuous and repeats over a time unit (e.g. year, month, week, day). However, subjects are followed for different lengths of time. <u>Person-time</u> is a measurement of the actual time-at-risk that subjects were exposed to. The incidence density rate, the number of new cases divided by the person-time at risk during the observation period, is an

estimation to reveal how quickly subjects are exhibiting the outcome of interest.

5. Strengths and limitations

Since you follow a cohort, the incidence of an outcome can be measured and the temporal relationship between an exposure and outcome can be clearly established. This design is well suited for assessing multiple outcomes of a single and/or rare exposure, such as factors related to high-risk occupations or environments. The study cohort can be easily sampled and followed in a specific population compared to a general population, with a caution of prevalent/survival cases.

Major limitations are time, cost, possible failure to follow-up, and difficulty of assessing rare outcomes. This design should not be used for studying rare outcomes, unless its attributable risk (AR) and/or attributable risk fraction (PAR) is high.

6. Study example

A typical example of a cohort study is one that was implemented after the nuclear power plant accident in Fukushima, Japan.[4] This cohort study enrolled all people living in Fukushima Prefecture after the nuclear accident. All data have been entered into a database and will be used to support the residents and analyze physical and mental health effects of a long-term low-dose radiation exposure. A cohort study design is suitable in this case be-

---

[4] Yasumura S, Hosoya M, Yamashita S, Kamiya K, Abe M, Akashi M, Kodama K, Ozasa K; Fukushima Health Management Survey Group. Study protocol for the Fukushima Health Management Survey. Journal of Epidemiology. 2012; 22: 375-83.

cause the exposure is obviously rare and multiple outcomes are expected. However, the low response rate (<30%) complicates the estimation of health effects in this example study.

Case-control Studies

Hirohide Yokokawa, Tran The Trung

1. Basic concept and study purposes

In a case control study, individuals with an outcome (termed as 'cases') are compared with those without an outcome (termed as 'controls') in regard to an exposure of interest. This study design can estimate the possible association of an exposure to a certain outcome. If cases have been more exposed than controls, the exposure is defined as a risk factor. On the other hand, if cases have been less exposed than controls, the exposure is defined as a protective factor. A case-control study can identify a new factor associated with an outcome, and can also further delineate the association of a factor previously associated with this particular outcome in another population.

2. Designing

Eligibility of cases and controls must be clearly defined; their exposure statuses are then studied retrospectively. Major steps of the procedure are shown below.

Figure 8.1. Process of a case-control study

Exposed
NOT exposed

WITH
outcome

Exposed
NOT exposed

WITHOUT
outcome
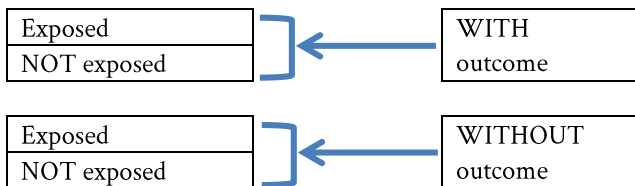
1. Define and sample cases and controls
2. Collect information on exposure
3. Develop 2x2 table
4. Calculate association

1) Selection of cases

The inclusion and exclusion criteria should be clearly stated before any form of recruitment. Questions worth asking at this stage include: how do you define the disease status? Do you exclude anyone with specific conditions/characteristics? In a multicenter study, a uniform protocol that defines inclusion and exclusion criteria is required to minimize methodological differences during recruitment.

2) Selection of controls

Controls should be representative of individuals without an outcome and ideally recruited from the same population as the cases. However, it is often difficult to implement random selection so as to recruit controls in an ideal fashion. There are several practical ways of recruiting controls. The first is a population control. We select eligible controls using community-based registries or records including residence registry, primary care records, neighborhood networks, and school directories. The second is a hospital control, which allows better access to medical information. For clinicians, this method is highly feasible. When cases and controls are selected from the same hospital, it will increase the comparability between two groups. However,

patients in the hospital control group might have certain illnesses, and their characteristics may differ from individuals in the community. Thus, its generalizability will be limited. The third method is a neighborhood control, which has an advantage of matching socioeconomic factors. However, the difficulty in identifying and accessing cases' matching neighbors is a critical disadvantage for physician-researchers working at hospitals.

Table 8.1. Different types of controls

	Advantage	Disadvantage
Population controls	Representative of a target population	Low feasibility
Hospital controls	Convenient for clinicians and may increase internal validity	Different from people in community leading to limited external validity
Neighborhood controls	Matching of socioeconomic factors	Access might be difficult due to security issues

3) Number of controls

Often, the number of cases is small and cannot be increased. In that case, increasing the number of controls may improve the statistical power of a study. The ratio of cases to control may be up to 1:4 approximately, as a cost-effective way to improve the power.

4) Matching of cases and controls

It is a major concern in a case-control study as to how cases and controls differ. It is recommended to evaluate the distribution of their basic characteristics in the initial stages of the research. When an obvious difference is expected, one effective approach is matching of cases and controls for the important factors that are expected to differ. In practice, you select a control that is similar in particular characteristics to match a case. These characteristics often include age and sex. For group matching, you select a control

group with the same proportion of certain characteristics as in a case group. If 30% of cases are men, controls are selected in a way that the proportion of males will be also be 30%, as in the cases. In individual matching, for example sex and age, you select a control of the same sex and age that ranges within two years of difference with the matched case. Individual matching is often used for hospital controls.

3. Collecting data

After defining cases and controls, the same data must be collected in the same way from both groups. Information on exposure are collected retrospectively from past records, including medical, employee, pharmacy, health checkup, or public survey records. You can also interview participants about potential associated factors in their past behaviors (such as history of smoking, diet, medication, or sexual behavior). An advantage of data collection in case-control study when compared to cohort study is its inexpensive cost and the short time period required for data collection. On the other hand, study variables are limited (it is difficult to add new variables), and recall bias is a major concern in such study design. The recall bias is explained in Chapter 3.

4. Presenting results

In a case-control study, the relative risk (RR) cannot be estimated directly because an incidence of outcome cannot be obtained. Instead, we can obtain a prevalence of exposure in cases and controls, and calculate the odds ratio (OR) as a measure of association between exposure and outcome. Odds are a probability defined as a ratio of the number of ways that an outcome

can occur to the number of ways that the outcome cannot occur. In case-control studies, OR is defined as the ratio of the odds of exposed (to unexposed) in case to the same odds in controls.

2x2 contingency table in a case-control study

	Cases	Controls
Exposed	a	b
Unexposed	c	d

Odds ratio (OR) = (a/c)/(b/d) = ad/bc

If OR is 1, the exposure is not related to the outcome. If OR is greater than 1, the exposure is positively related to the outcome occurrence (= risk factor). If OR is less than 1, the exposure is negatively related to the outcome occurrence (= protective factor).

OR = 1	No association
OR $\geqq$ 1	Risk factor
OR $\leqq$ 1	Protective factor

When calculating the OR, it is recommended to compute 95 % confidence interval (CI), which is an estimated range of OR. The 95 % CI means that there is a 95 % probability which the interval contains the true OR. If the 95 % CI range includes "1", the association is not statistically significant since there are possibility that the exposure could be both risk (OR>1) or protective (OR<1). If a lower limit of the 95 % CI range is greater than 1, the exposure poses a risk with more than 95 % probability. If an upper limit of the 95 % CI range is less than 1, the exposure is protective with more than 95 % probability. These concepts are further explained in Chapter 9.

Of note, when cases and controls are matched, the formulation of a 2x2 table will be more complex. The numbers in cells are now indicated as pairs. For example, the number "a" in the first cell indicates the number of

matched pairs, in which both cases and controls are exposed. OR is calculated with numbers of discordant pairs.

2x2 contingency table in a matched case-control study

Cases	Controls	
	Exposed	Unexposed
Exposed	*a*	***b***
Unexposed	***c***	*d*

Odds ratio (OR) = b/c

5. Strengths and limitations

This study design has several strengths, such as 1) inexpensive study cost, 2) no long follow up period, 3) efficient for the study of rare diseases and 4) for examining multiple exposures, compared to cohort study. However, the several limitations include: 1) prone to bias (especially selection, recall and observer biases), 2) limited to examining one outcome, 3) unable to estimate the disease incidence rate, 4) difficult to investigate rare exposures, 4) difficult to assess a causal relationship, and 5) limited generalizability.

6. Study example

One typical example of a case-control study is my thesis work.[5] The mortality rate for cerebrovascular diseases is much higher in Japan than in North American and West European countries. I thus investigated associations between cerebral infarction and demographic factors, medical history and other clinical measurements including pulse wave velocity (PWV), a

---

[5] Yokokawa H, Goto A, Watanabe K, Yasumura S. Evaluation of atherosclerosis-associated factors and pulse wave velocity for predicting cerebral infarction: a hospital-based, case-control study in Japan. Internal Medicine Journal. 2007; 237: 161-167.

newly introduced noninvasive measurement procedure used to assess aortic stiffness. This was a hospital-based, matched case-control study in northern Japan where a high incidence of cerebrovascular diseases is present. The subjects consisted of 92 matched pairs of cerebral infarction patients (cases) and healthy individuals admitted for a thorough health checkup (controls) at the Southern Tohoku General Hospital in Fukushima Prefecture, Japan. I found that a high PWV over 1600 cm/sec, in addition to other traditional risk factors (family history of hypertension and cerebrovascular disease, high density lipoprotein cholesterol of lower than 40mg/dl) were associated with a higher risk of cerebral infarction.

The study design was suitable because the outcome was rare (although mortality rate is higher than in Western countries, the disease is not as common as, for example, diabetes), multiple exposures were expected, and my time was limited to complete the thesis work. The design was also convenient for a physician working at a hospital.

Intervention Studies: Basic Theories and Reality

Aya Goto, Tran Quang Nam

1. Defining an intervention study

An intervention study is a follow-up study in which an exposure status is assigned by a researcher. When the assignment is at random, it is called a randomized control trial (RCT). According to the hierarchy of research designs, the intervention study is ranked as the study that provides top-level scientific evidence, provided that it is well designed and implemented. But this also suggests difficulties in terms of its implementation in the real world. High quality scientific evidence can only be achieved by an appropriate baseline randomization, blinding and usage of placebo.

2. Randomization and compliance

The classic envelop method of randomization gives you a concrete picture of what it is about. Imagine you have decided that your sample size is 100, out of which 50 will be assigned to an intervention group and another 50 to a control (non-intervention) group. You prepare 50 sealed envelops

with a card inside stating "assign this person to an intervention group" and another 50 stating "assign this person to a control group", and mix these envelopes randomly. You pick one when seeing a patient. You are not allowed to try more than once. Nowadays, this can be done by a computer. Through this randomization, characteristic of two groups (the intervention and control groups) will be comparable. Otherwise, a physician tends to assign a patient to a treatment that he/she believes to be the best for the patient, which is their job. Therefore, this is a way to control physician's good intentions.

On the other hand, there are patients' own will to decide whether or not to follow the assigned treatment. This is called <u>compliance</u>. An important implication drawn from the Coronary Drug Project[6] is that those with poor compliance hold distinct characteristics from those with good compliance. An RCT was conducted to evaluate the efficacy and safety of several lipid-influencing drugs including Clofibrate. They found lower five-year mortality rates among patients with good compliance than those with poor compliance, even within the control group. Interestingly, the mortality of the patients with good compliance in the control group was much lower than that of the patients with poor compliance in the intervention group.

Even if patients have given informed consent to participate in your RCT, some may follow the assigned treatment and some may not. At the analysis stage, you may be tempted to move a patient in an intervention group who did not follow the treatment regimen to a control group or vice versa. However, you are advised to go against your temptation. This is called an <u>intention-to-treat</u> analysis not to violate the random assignment you have done at the baseline. This is to make sure two groups are comparable; in other words, to consider potential confounders.

---

[6] The Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. New England Journal of Medicine. 1980; 303: 1038-1041.

3. Double-blinding and placebo

In order to further mask intentions of all people involved in the study, you need to do <u>double blinding</u>. They include patients who are participating in the study, physicians who are assigning treatment to patients and assessing the outcomes, and researchers who are managing the study and analyzing the data. Once a patient knows in which group he/she belongs to, their compliance to the treatment and perception of the outcomes might be influenced. Once a physician knows in which group the patient belong to, his/her assessment and observation of the outcome might be influenced. Once a researcher knows which group is the treatment group, his analysis might be biased towards getting the intended results.

In order to mask patients, a <u>placebo</u> is used. When Karlowski and colleagues[7] conducted an RCT aimed at evaluating the efficacy of Ascorbic acid to prevent common cold, their double blinding procedure failed. Many of the participants could tell which pill they were taking (Ascorbic acid or the placebo) from its taste. Taking this as an advantage, they did an analysis to compare the occurrence of colds among people who exactly knew, did not know and guessed wrong about which pill they were taking. There was an interesting distribution among those who guessed wrongly. The occurrence of cold was higher among people who were taking the Ascorbic acid thinking they were taking the placebo than people who were taking the placebo thinking they were taking the Ascorbic acid. Here you can see a placebo effect. Placebo is used not only to mask patients from knowing which group they are in, but also to extract the true effect of the intervention.

[7] Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. JAMA. 1975; 231: 1038-1042.

4. Quasi-experimental study

Although the conceptual understanding of randomization and blinding is not difficult, it is difficult to apply in the real world. Can you convince your hospital director or municipal head about random assignment? Can people accept being randomized? Can you handle the logistics of blinding? Although the premise of an intervention study is that you do not know about the efficacy of an intervention, if there is a slight chance that the new intervention is going to be better, people will prefer to be in the intervention group. In general, public health providers in Asian communities prefer for everyone to have an equal opportunity to receive the services being provided, henceforth the concept of randomization is incompatible.

RCTs evaluate efficacy of interventions in an ideal setting. There are cases when implementing a RCT is infeasible, or when you want to assess the feasibility and effectiveness of an intervention in the real world in your target community. In such a case, quasi-experimental designs can be useful. They can be classified in a number of ways,[8] but most commonly into two types: one without a control group and another with control group but without randomization.

Table 9.1. Quasi-experimental studies: Kinds of data to compare

	Without controls	With controls
Post-test only		Post-test of cases vs. Post-test of controls
Pre- and post-tests	Post-test vs. Pre-test	Pre&post test difference of cases vs. Pre&post test differences of controls

[8] Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, Finkelstein J. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc. 2006; 13: 16-23.

They have an advantage of providing evidence of effectiveness in a natural setting, but one must be aware of it major weaknesses. First, it is difficult to control for important confounders. Second, repeat testing of one group will lead to regression to the mean – a decline or increase in an indicator may happen even without intervention statistically. Third, there might be a maturation effect that is related to natural changes with the passage of time rather than the effect of intervention.

5. Study example

Here is an actual example of a quasi-experimental intervention study.[9] Since scientific research on parental interventions in Asia is scarce, we adopted a Canadian multi-language parenting support program into a Japanese public health service setting and evaluated its impact. We compared changes in mental statuses of 32 mothers attending the intervention with data of 156 mothers attending child health checkups. The study was conducted in collaboration with a municipal health center, and it was difficult to randomize the intervention, which was offered as a public service. On the other hand, we had an advantage of having an access to health checkup files kept at the health center. What we compared was a change (post-intervention minus baseline) in self-efficacy levels among mothers attending the intervention and a change in the same indicator during the same period among mothers according to comparative data. The study indicated that the new program was feasible as a public service and has a potential positive impact on mothers' self-efficacy.

---

[9] Goto A, Yabe J, Sasaki H, Yasumura S. Short-term operational evaluation of a group-parenting program for Japanese mothers with poor psychological status: adopting a Canadian program into the Asian public service setting. Health Care for Women International. 2010; 31: 636-651.

Key Concepts in Biostatistics

Nguyen Quang Vinh, Nguyen Thi Tu Van

1. Introduction

To explore an uncertainty, one should have its data. Statistics is a discipline to turn available data into relevant information showing a strongest possible conclusion about the population being studied. With abundant information available through articles, books, media and the Internet in our modern society, a clinician should be able to not only to read, but also interpret data in a way to utilize the scientific evidence in clinical practice.

The two main objectives of doing an epidemiological research are (1) to summarize data of a study sample and (2) to reach inferences from a study sample to the population. Statistics is an important tool as it provides aspects of descriptive statistics, which is used to serve the first objective, and inferential statistics for the second.

In this and the following chapter of biostatistics, we avoided mathematical formulas as much as possible and focused on helping you understand the main concepts. Further calculations will be explained in our lectures, and can be explored with a statistical package.

2. Descriptive statistics

Descriptive statistics help you to summarize your available data by grouping and calculating measures to describe its distribution.

1) Category

A well-designed study provides us with good raw data, but it needs to be further organized. In order to convey relevant information, the raw data needs to be presented in a clear manner. Data must be grouped into categories that are contiguous and non-overlapping. The number of categories should not be too many (not summarizing) or too few (not enough information). The width of categories can be the same or different. One classic example of a category with the same width is structuring reproductive age into 5-year groups: 15-19, 20-24, 25-29, 30-34, 35-39, and 40-44. When a different width is chosen for data presentation, it should be reasonable and depends on the objectives of the study. For example, in a study of reproductive tract infection among women in their reproductive age, they may be divided into 3 categories: under 20, 20-39, and 40 or over; the rationale being a difference in vaginal wall thickness due to changes in hormonal levels and sexual activities. A table to present grouped data is shown as an example below.

Category	Frequency	Relative frequency	Cumulative frequency	Cumulative relative frequency
Under 20	100	20%	100	20%
20-39	350	70%	450	90%
40 or over	50	10%	500	100%

2) Summary measures

Rather than categorizing data, there are three <u>summary measures</u> that describe the central tendency of data.

· <u>Mean</u> (average) = Sum of all data / n
· <u>Median</u> = 50th percentile value
· <u>Mode</u> = Most frequent value(s)

Calculating the mean is simple, but largely influenced by extreme values and appropriate only when data distribution is <u>normal</u> (bell-shaped). The median is also simple, but is not influenced by extreme values.

Most of the information extracted from people have a high degree of <u>variability</u> among individuals. Therefore, measures of dispersions are shown in order to show how data is spread around the summary measures.

· <u>Range</u> from minimum to maximum
· <u>Variance</u> = Mean of squared differences of values from a mean
· <u>Standard deviation</u> (SD) = Square root of the variance. SD indicates the mean of differences of values from the mean. It measures absolute dispersion.
· <u>Coefficient of variance</u> (CV) = Ratio of SD to the mean. CV indicates the extent of variability in relation to the mean. It measures relative dispersion. When the value is over 100%, it means that the data set has an extreme degree of variability. There is no unit to measure CV, therefore they can be compared between any sets of data.

The <u>measures of position</u> evaluate a given value compared with others in the dataset, and describe the relation of a datum to all data of a variable. Two measures of position are percentile (and quartile), and z-score.

· <u>Percentile</u> = Percent of values at or below
· <u>Quartile</u> = 25th, 50th, and 75th percentiles
· <u>Z-score</u> = a standardized value indicating a distance between the value and the mean divided by (in units of) the standard deviation.

# 3. Inferential statistics

1) Estimation

To find out the parameters in population is a desire of researchers, but to explore an infinite population is impossible. Therefore, the understanding of statistics in a given sample helps investigators estimate inferential parameters in a population, not deferring a conclusion until the whole population is observed. Statistics commonly used for the <u>estimation</u> are mean, proportion, and variance. There are two types of estimation: a point estimate and an interval estimate.

The idea of a <u>point estimate</u> is simple. The statistic from a sample is a point estimate, and also called an "estimator" as an inferred parameter in population. A good estimator should meet two criteria; the data collection was without systematic error and its standard error is smaller than that of other estimators (i.e.. consider whether mean or median is the best estimation for the parameter in your sample).

An <u>interval estimate</u> provides an estimate with a range from a formula, estimator ± (reliability coefficient) x (standard error). When a sample is from a normal distribution, the reliability coefficient is the z-score when a variance is known, but it also can be calculated when the variance is unknown. The interval given by the above formula can be interpreted as follows, "when sampling is repeated, $100(1-\alpha)\%$ of all calculated intervals will include the population mean." The quantity $(1-\alpha)$ is called the confidence coefficient and the calculated interval is called the confidence interval. The values .90, .95, .99 are frequently used for the confidence coefficient, and the corresponding z-scores (reliability factors) are 1.645, 1.96, 2.58, respectively. In previous chapters, the concept of 95% confidence interval (CI) is mentioned. In practice, it is interpreted as "we are 95% confident that the interval contained the true (population) mean." Figure 10.1. shows that if sampling is

repeated 100 times, there are 5 times that a confidence interval does not include the true mean.

Figure 10.1. 95% confidence interval
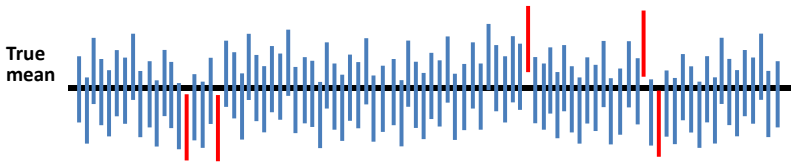
2) Hypothesis testing and p value

To reach a decision regarding a difference, you need to generate a hypothesis. Hypothesis testing is a procedure to aid you in making a decision whether the difference observed in your sample was systematic, or it occurred by chance. A more precise expression will be "hypothesis testing is the procedure to compute the probability of the difference happening only by chance." There are two types of hypotheses: research and statistical hypotheses. A research hypothesis usually results from a mounting observation. This research hypothesis leads directly to a statistical hypothesis, which is written in a mathematical language to be evaluated by an appropriate statistical technique.

There are two types of statistical hypotheses: null and alternative hypotheses. The alternative hypothesis is what you want to conclude about the population (e.g. effect of a new treatment A is different from a traditional treatment B) and the null hypothesis is its opposite (e.g. effect of treatment A is as same as treatment B). The decision to reject the null hypothesis or not depends on the magnitude of a test statistic calculated from this general formula, (relevant statistic of your sample - hypothesize parameter in a population) / standard error. Based on the calculated test statistic, you can find a corresponding p value in a table, which is usually attached in the end of biostatistics textbooks. There are usually several tables attached, and you need

to select the one that matches the distribution of your test. When you are using a statistical package, this step is embodied in the package. A basic rule is that when the p-value is small, you reject the null hypothesis (i.e. treatments A and B are not the same) and support the alternative hypothesis (i.e. treatments A and B are different) making a conclusion that there is a difference. There are two ways to hypothesis testing: one-sided and two-sided hypotheses tests. A thoughtful investigator should use a two-sided approach unless s/he knows very well that the difference occurs only in one direction (e.g. treatment A must be better than B and B will never be better than A). Keep in mind that there is no test that can "prove" a hypothesis. The hypothesis testing procedure only indicates whether or not the hypothesis is "supported" by the available data from a sample data. The convention is to determine whether a hypothesis should be rejected, as opposed to whether a hypothesis should be accepted.

The p value is not a binominal indicator showing whether to reject a hypothesis or not, but has more meaning attached to it. It shows the level of investigator's belief in the null hypothesis. It is a probability (p) of one's belief that the null hypothesis is true. The concepts behind the p value are shown in the table below. The key point is that when deciding whether the null hypothesis should be rejected or not, the "truth" in the population is unknown. The α refers to the statistical significance, which is a probability allowed for the error of rejecting a null hypothesis when it is true. The β is a probability of the error of not rejecting a hypothesis when it is false. When the error is small, a probability of the test accurately rejecting the false hypothesis increases. One minus β is called the statistical power.

		Truth in the population	
		Null hypothesis is FALSE	Null hypothesis is TRUE
Results from your sample	Reject	Correct decision	Type I error = α (significant level)
	Not reject	Type II error = β (1-power)	Correct decision

Let's review the five-step procedure of hypothesis testing. A statistical package helps you with the step 4, but not for the other steps.

Step 1   Set up null and alternative hypothesis.

Step 2   Select a test (distribution).

Step 3   Decide a significant level.

Step 4   Compute a test statistic and then a p value.

Step 5   Give a clear statement without jargon how the result is interpreted.

Generally, calculating sample sizes is an early consideration in the designing phase of a study. This is to estimate an appropriate number of subjects for a given study design. Numbers that are too small will influence the precision of the point estimation; while a sample size that is too large will incur extra resources. When estimating a sample size for descriptive purposes, investigators just need to consider the Type I error. When estimating a sample size to test a hypothesis, the Type II error will need to be taken into account. Most of statistical packages can calculate a sample size to give a best estimate of results and an appropriate selection of test (one-sample comparison of mean/proportion to a hypothesized value or two-sample comparison of mean/proportion). When the sample size is fixed because of some logistical reasons including financial and/or time constrains, the investigators should think about the precision of the available data, and whether this is meaningful for the study objective.

Basic Tests

Nguyen Quang Vinh, Nguyen Thi Tu Van

1. Introduction

In this chapter, we introduce four basic statistical tests. Chi-square (including NcNemar's test) and Fisher's exact tests are to analyze proportions, and Student's and Mann-Whitney tests to analyze two means. In addition to the four tests, diagnostic tests are summarized in the end.

There are many statistical software, but here are two reliable packages that available online for free.

· OpenEpi: Very easy to handle. It includes a sample size calculator. There is a professional version, which is Epi Info.

http://www.openepi.com/v37/Menu/OE_Menu.htm

· R: Another professional package.

http://www.r-project.org/

2. Chi-square test

Chi-square test is one of the most frequently used tests, and its calculation is based on the chi-square distribution. The principle of this calculation

is to compare observed frequencies to expected frequencies in cells of a contingency table considering a degree of freedom. The expected frequency in a cell, for example a, is calculated as (a+c)x(a+b)/(a+b+c+d).

	Outcome +	Outcome -	
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	a+b+c+d

The <u>degree of freedom</u> in a chi-square distribution is calculated from a number of cells. It determines a shape of distribution, based on the p value that is being used. The chi-square statistic is applied to three tests: test of goodness-of-fit, test of independence, and test of homogeneity. Testing for difference between two (or more) proportions is the test of independence.

- Null hypothesis: 2 categories are independent.
- Alternative hypothesis: 2 categories are not independent.
- Degree of freedom = (number of cells in row - 1) x (number of cells in column - 1)
- In general, the chi-square test should not be used if $n<20$ or any of expected frequencies<5.
- Chi-square statistic = Sum of [(Observed – Expected)2 / Expected] for each cell
- Below is a part of a table you use to look for a p value given the calculated statistic. When the degree of freedom (df) is 1 and calculated chi-square statistic is 4, p value is less than 0.05 and the result is significant.

df	p value										
	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.6	11.98	13.82

Of note, in the case of matched data as mentioned in Chapter 8, you should select the <u>NcNemar's (Chi-square) test</u>. The null hypothesis is that proportions of b and c are equal.

2x2 contingency table in a matched case-control study		
	Controls	
Cases	Exposed	Unexposed
Exposed	*a*	**b**
Unexposed	**c**	*d*

3. Fisher's exact test

Fisher's exact test is used when the expected value of any cell in a contingency table is small (say, smaller than 5). The test calculates the probability p of getting the observed set, or one that is more extreme. The test lists all combinations of results which give the same marginal totals, calculate probabilities attached to them, and calculate a exact probability of getting the observed set. In short, the Chi-square test gives an approximation and the Fisher's exact test an exact probability.

4. Student's t test

The principle of Student's t test is to compare the sample mean with the hypothesized mean, considering the standard error, with a corresponding degree of freedom. Just as in chi-square test, the degree of freedom determines a shape of t distribution, based on which the p value is being determined. For a given test statistic, the p value gets smaller as the degree of freedom gets larger.

- Null hypothesis: the sample mean and the hypothesized mean are the same
- Alternative hypothesis: the sample mean and the hypothesized mean are different
- Degree of freedom = n (sample size) - 1
- t statistic = (sample mean – hypothesized mean) / standard error

Of note, just like the Chi-square test, this t-test also has its matched version, paired t-test. It calculates differences of paired values, and the null hypothesis is that the difference in the mean values equals to zero.

5. Mann-Whitney test

Mann-Whitney test is a substitution of 2-sample t test and used when sample size is small and distribution of tested variable in population is not normal. The test considers the rank of observations rather than the value of observations, and compares the sum of ranks of the observations between two groups.

- Null hypothesis: 2 populations have identical of the probability distribution
- Alternative hypothesis: 2 populations differ in distribution

6. Diagnostic tests

Various measurements including medical tests are used for various purposes including diagnosing a disease, assessing health status, screening for a risk factor, and estimating prognosis. The test results can be dichotomous, categorical, or continuous. A diagnostic test can be assessed by comparing it with current gold standards. Major indicators are shown in the second table below. A diagnostic test is valid if it has a high sensitivity, a high specificity, and a high positive predictive value. The best measure of the usefulness of a test is the likelihood ratio.

	Disease positive	Disease negative
Test positive	a	b
Test negative	c	d

Indicators	Calculation	Address the question of...
Sensitivity (Sen)	a / (a+c)	How good is the test in detecting those with the disease?
Specificity (Sp)	d / (b+d)	How good is the test in detecting those without the disease?
Positive predictive value	a / (a+b)	What is the probability that a test positive person has the disease?
Negative predictive value	d / (c+d)	What is the probability that a test negative person do not have the disease?
Accuracy	(a+d) / (a+b+c+d)	What is the proportion of correct results?
Positive likelihood ratio	Sen / (1-Sp)	How much likely is a positive test found in a person with the disease than the without?

Deciding where to draw a line between positive and negative, a value called the cutoff point, requires a consideration of a trade-off between sensitivity and specificity. You must weigh relative importance of the sensitivity and specificity. If a false positive leads patients to a risky treatment, the cutoff point should be set to maximize the specificity. If a false negative leads to a misdiagnosis of a treatable serious condition, the cutoff point should be set to maximize the sensitivity. Another systematic way to set the cutoff point is to use a receiver operator characteristic (ROC) curve. As a guide only, the cutoff point is the corner where the ROC curves turns from steep section to flat region.
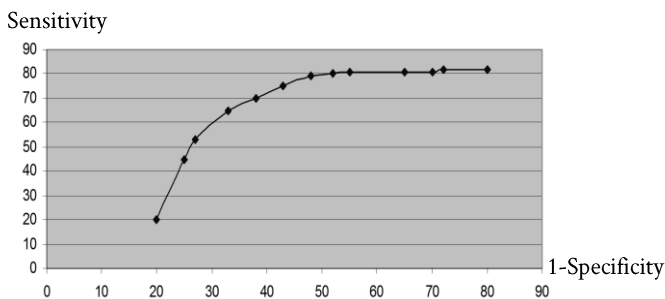
A predictive value is determined after knowing the test results, and is also called posterior or <u>post-test probability</u>. It is influenced by test characteristics (sensitivity and specificity) and a community characteristic (prevalence of the disease). The positive predictive value depends on the prevalence of the disease in the population being tested. For clinical use, the more specific test (less false positives) is needed for a rare disease (low prevalence). Otherwise, a test for positive will likely to be false positive. In contrast, when the disease is more common (high prevalence), a highly sensitive test (less false negatives) must be chosen. Otherwise, a test for negative will likely to be a false negative. The test is more useful when prevalence is not very extreme, say between 0.3 and 0.7.

The <u>likelihood ratio</u> is a highly useful measure to show characteristics of the test at different optional cutoff points. A likelihood ratio that is greater than 1 indicates that a positive test is more likely to occur in people with the disease than in people without the disease. In other words, a likelihood ratio that is greater than 1 indicates that the test result is associated with the presence of the disease. On the other hand, a likelihood ratio less than 1 indicates that the test result is associated with the absence of disease. A test with a likelihood ratio above 10 or below 0.1 is considered to provide strong evidence to rule in or rule out diagnoses, respectively. When more than one test is independently used for making a diagnosis for a disease, all the likelihood ratios of different tests can be multiplied together to give an overall likelihood ratio.

Academic Dishonesty

Chihaya Koriyama, Nguyen Thy Khue

1. Basic rules

The inquiry into the truth is our responsibility and accountability as scientists. We are expected to make sincere efforts to address scientific topics during the research process. Here we attempt to furnish three general definitions of academic misconduct for your initial reference.

1. Fabrication: to create data that is not based on a survey, measurement, experiment or any other research methodology,

2. Falsification: to alter original data intentionally at the convenience of author(s),

3. Plagiarism: to present the work of another as one's own without proper acknowledgements of the source of data/information being used intentionally. Recently, self-plagiarism (presentation of work or idea from your studies published previously) is also a concern in relation to publisher copyright issues.

Since it is difficult for reviewers and editors to distinguish unintentional/careless mistakes (honest errors) and falsification or plagiarism, author(s) are advised to be cautious in all research processes. To put it simply, any

form of deliberate deviation beyond what your data reflects, and the duplication of other's work is not acceptable.

2. More rules

In an expanded sense, the followings are also considered as scientific misconducts:

1. Citation of reference defectively or improperly

2. Hyperbole - exaggerated description of the finding(s)

3. False description of the novelty of finding(s)

4. Double publication: to submit identical data/work to more than one peer-reviewed scientific journal

5. Improper authorship: The International Committee of Medical Journal Editors has summarized the definition of authorship and contributorship. Details are v.i.: http://www.bmj.com/about-bmj/resources-authors/article-submission/authorship-contributorship

When conducting research, it is your responsibility to avoid misconducts, even if they are unintentional. Science can only advance if there is integrity reflected by researchers and if knowledge is based on the proper methodologies to extract and interpret data.

Index