

INTRODUCTION TO CLINICAL RESEARCH FOR RESIDENTS

The Saudi Commission for Health Specialties
The Department of Medical Education & Postgraduate Studies

Read
Evaluate
Study
Experiment
Apply
Revise
Circumstantiate
Harness



2014

Prepared by

Hani Tamim, PhD

*Associate Professor, Department of Internal Medicine
Director, Biostatistics Unit, Clinical Research Institute
Director, Fellowship and Residency Research Program
American University of Beirut Medical Center
Beirut, Lebanon*

Reviewed by:

Haroutune Armenian, MD, DrPH

Professor in Residence at the UCLA

*Professor Emeritus of Epidemiology at Johns Hopkins Bloomberg School of Public Health
Los Angeles, USA*

Mohammed Al Onazi, MD

*Consultant Pediatric Surgeon, Department of Pediatric Surgery
Prince Sultan Military Medical City
Medical Education Consultant, Saudi Commission for Health Specialties
Riyadh, Saudi Arabia*

Saud Al Shanafey, MD, MSc, MBA, FRCSC

*Deputy Executive Director Academic & Training Affairs
Associate Professor & Consultant, Pediatric Surgeon
King Faisal Specialist Hospital & Research Center
Medical Education Consultant, Saudi Commission for Health Specialties
Riyadh, Saudi Arabia*

Edited & formatted by:

Vittoriana Crisera

*Journal Manager, Journal of Health Specialties
Saudi Commission for Health Specialties
Riyadh, Saudi Arabia*

© Saudi Commission for Health Specialties, 2014
King Fahd National Library Cataloging-in-Publication Data

Saudi Commission for Health Specialties - Riyadh 2014
Introduction to Clinical Research for Residents

92p, 14.8cm×21cm

ISBN: 978-9960-9832-9-5

1. Clinical medicine - Diagnosis I-Title
616.075 dc 1435/9120

L.D.no. 1435/9120

ISBN: 978-9960-9832-9-5



Introduction to Clinical Research for Residents

Preface.....	1
Section 1: Overview of the research process	2
1. What is research	2
2. Steps in conducting research	2
Section 2: Formulating a research question.....	5
1. What is a research question	5
2. Criteria of a good research question	5
3. Sources of research questions	5
4. Elements of a research question	5
5. Criteria for a bad research question	6
6. Example of a research question	6
Section 3: Conducting an effective literature review	7
1. Definition of a literature review	7
2. The aim of a literature review	7
3. Characteristics of a good literature review	7
4. Steps of a literature review	8
5. Conclusion	14
Section 4: EndNote	15
1. What is EndNote	15
2. Introduction to an EndNote library	16
3. To create a library	16
4. Entering references into a library	16
5. Inserting references and generating bibliographies in Microsoft Word	17
Section 5: Writing a proposal.....	19
1. Definition of a research proposal	19
2. Aims of a research proposal	19
3. Sections of a research proposal	19
Section 6: Ethical issues and IRB	24
1. Ethical issues	24
2. IRB	25
Section 7: Data collection, entry, cleaning and management	27
1. General terms	27
2. Data collection	28

3. Database structure	28
4. Data entry	30
5. Data cleaning	31
6. Data management	32
Section 8: Descriptive statistics	33
1. Descriptive statistics for categorical variables	33
2. Descriptive statistics for continuous variables	35
Section 9: Inferential statistics	38
1. Definition	38
2. Probability distributions	38
3. Types of inferential statistics	39
4. General rule for normal distributions	41
5. Different statistical tests	41
6. Sample size calculation	42
Section 10: Measures of effect	43
1. Absolute risk	43
2. Relative risk	43
3. Risk difference	44
Section 11: Overview of study designs	45
1. Observational studies	45
2. Experimental studies	46
3. General terms	46
4. Biases in clinical research	48
Section 12: Cross-sectional design	50
1. Definition	50
2. Choosing a representative sample	51
3. Strengths and limitations	51
Section 13: Case control study designs	53
1. Definition	53
2. Identifying cases	53
3. Identifying controls	54
4. Exposure assessment	54
5. Matching	54
6. Analysis	55
7. Strengths and limitations	56

Section 14: Cohort design	57
1. Definition	57
2. Types of cohort studies	57
3. Strengths and limitations	58
Section 15: Experimental studies	59
1. Definition	59
2. Phases of clinical trials	59
3. Types of clinical trials	60
4. Ethical consideration	60
5. Randomization	61
6. Blinding	61
7. Analyses	62
8. Strengths and limitations	63
Section 16: Manuscript writing	65
1. Overview	65
2. Manuscript preparation	65
3. Manuscript submission	68
4. Submission response	69
Section 17: Evidence based medicine (EBM)	71
1. What is EBM	71
2. Why we need EBM	71
3. EBM principles	72
4. EBM steps	72
5. Limitations of EBM	74
6. Linking research to EBM	74
References	75
Appendices	79
I – Research proposal	79
II – Manuscript template	91

Preface

I would like to extend my gratitude to the Saudi Commission for Health Specialties (SCFHS), and specifically Prof. Dr. Sulaiman Bin Omran Al-Emran, the Deputy Secretary General of the SCFHS, for his vision, support, and guidance. Moreover, I would like to acknowledge Dr. Mohammad Al-Onazi, Medical Education and Postgraduate Training Department, for his follow-up, review, and feedback.

Moreover, I would like to give special thanks to the team of Pragma-RTS, a Research Training Services Company (www.pragma-rtts.com) for their support in preparing this booklet, specifically Dr. Samaa Al Tabbah. Their contribution was at all the stages of the booklet preparation, from the topic selection, writing of the different sections, and final editing. Their valuable guidance and advise was much appreciated.

Finally, I would like to thank all those who provided their support in the preparation of this booklet, specifically for editing, proofreading and design.

Hani Tamim, PhD

Associate Professor, Department of Internal Medicine
Director, Biostatistics Unit, Clinical Research Institute
Director, Fellowship and Residency Research Program
American University of Beirut Medical Center
PO Box 11-0236 (F15)
Riad El Solh, Beirut 1107 2020
Beirut, Lebanon
E-mail: hani_t@hotmail.com, htamim@aub.edu.lb

Section 1: Overview of the research process

1. What is research

Research is often viewed as the corner stone of scientific progress. It is a systematic process based on scientific method that consists of testing hypotheses, careful observation and measurement, systematic evaluation of data, and drawing valid conclusions.^[1]

There are different research methods which are used in research, all with specific advantages and disadvantages. Which one the scientist uses, depends on the aims of the study in addition to the nature of the phenomenon. The main two types of research designs used in medicine are the quantitative research design, which generates numerical data or information that can be converted into numbers, and the qualitative research design generates non-numerical data that provides a complete description of the research topic. In this section we will focus on the quantitative design.^[2]

2. Steps in conducting research

The key steps of a research process start with a broad idea of the topic of interest which is then focused on formulating a specific hypothesis or a question that guides the following steps of data collection, gathering, interpretation, and analysis to reach a conclusion which can be generalized to clinical practice.

The key steps in conducting research can be summarized as follows:

- a. **Identification of research problem:** It is the first statement made in any research. An example of a research problem, which may be of a local concern for the community is, violence among children. This is the problem to be investigated. At this stage, the identified problem is still too broad in scope. Therefore, the scope of the study needs to be narrowed down, which can only be done after a thorough literature review.
- b. **Carrying out a literature review:** A literature review is the process of searching for information related to the research topic, across multiple databases and information resources; reading, evaluating and analyzing them helps refine the topic and objectives of the overview to be written. A good literature review will help in finding gap(s), asking a good question, and accurately defining a problem, as well as identifying a proper methodology. After doing a thorough literature review, the focus can be

narrowed down into violence among children who watch TV Programs which needs parental guidance.

- c. **Formulating the research question / research objective:** The next step in the research process after refining the objective through a thorough literature review involves translating that research idea into an answerable question. A research question, is the question that a researcher wishes to study or the hypothesis he/she wishes to test. The most important part of the question is to be researchable and answerable using established scientific methods and procedures. Otherwise, the whole idea will have no point and will reach nowhere. Moreover, the research question should attempt to fill a gap in knowledge. For example, an answerable question could be: Does watching violent TV programs increase violence among children?
- d. **Proposal writing:** The main purpose of writing a research proposal is to obtain ethical approval, as well as funding. A well written proposal provides a structured outline guiding the researcher throughout each step of the research process. It is written to justify the postulated research question and to present a detailed methodology in which the research should be best conducted. There is no universal format of the proposal. Each institution has their specific format.
- e. **Institutional Review Board (IRB) approval and ethical consideration:** IRB is specific human-subjects committees that review and determine the ethicality of research. IRBs exist in research institutions and consist of academicians, researchers, clinicians, and a representative from the community. The main task of the IRB is to protect the rights participants in research being conducted. IRBs have the authority to require modifications on the proposal pertaining to all aspects of the research, as well as to disapprove a proposal. Finally, once the IRB grants the researcher its approval, he/she can start conducting the research.^[1]
- f. **Data collection:** Data collection is the process of collecting the information that will be used to answer the research question. Its most important aspect is to select the relevant information needed to answer the research question. The development of the data collection form is a time consuming process, and it should be given enough attention to ensure it is clear, valid and reliable. It has to be piloted on a small sample of subjects to guarantee applicability, as well as identify any potential problems in the tool. Finally, once the necessary fine-tuning has been carried-out, data collection can be done through surveys, questionnaires, interviews, observations, data abstraction, etc.
- g. **Data entry, cleaning, and management:** It is the process of entering the information (data) into the computer. There are many computer programs

used for this purpose. The most widely used is the Statistical Package for Social Sciences (SPSS). Most of the time data entry is done manually. After finalizing data entry, a data cleaning process is carried out to check for any data errors or outliers. Once data is clean, data management starts which includes creating new variables from the existing ones based on pre-defined criteria.

- h. **Data analyses:** Data analysis is the process of analyzing data using statistical techniques in order to draw conclusions that support or reject the hypothesis, or answer the research question. If the data were properly entered, cleaned and managed, then the data analyses process should run smoothly.
- i. **Research dissemination:** Data analyses results should be organized into tables and figures which serves multiple purposes such as: understanding the results, identifying missing analyses, comparing easily with other published articles, and preparing results for presentation. Publishing the research project is the final step of the research process, which entails summarizing the whole research findings in different forms, such as an abstract, presentation, report, or a manuscript published in a journal.^[1,3,4]

Section 2: Formulating a research question

1. What is a research question

A research question is a concise question which expresses what the research project aims to address; it is the question the researcher wishes to answer. The first step in any research project is to clearly define the research question, since it will be the basis for developing and conducting the research project.^[5]

2. Criteria of a good research question

There are different criteria that define a good research question. The best research question is one that attempts to fill a knowledge gap in the literature, enough to be answerable and researchable, should be relevant to the study's objective, should avoid duplication of previous work, should be feasible and cost-effective, and should be doable within a specified time frame allocated for the research project. Moreover, it needs to meet the minimum ethical standards.^[6]

3. Sources of research questions

The researcher formulates a research question through different means. Literature reading might help a researcher identify the areas that need further study, and thus a hint towards a research question. Another method is patients' observation through clinical experience which might pose different researchable questions. Other methods include previous research, journal clubs, and conferences.^[7]

4. Elements of a research question

A "well-built" research question should include four elements; referred to as PICO that identifies the key elements of a research question that need to be addressed. The PICO concept is important in narrowing down the research question, providing search terms, and saving time in literature search.^[5,8]

PICO letters stand for:

- a. **P: Patient or population.** The first step in the PICO process is to identify the patients or population to be studied. More specifically, it describes patients' characteristics, such as age, gender, disease status, or any other patient-related characteristic.

- b. **I: Intervention to be tested.** Identifying the intervention is the second step in the PICO process. It is important to identify the exposure intended to be studied in the research project. This may include the use of a specific diagnostic test, treatment, adjunctive therapy, medication, etc.
- c. **C: Comparison used in the research project.** It is the alternative exposure to which the intervention will be compared, which might be the standard of care or a placebo. The comparison component is the only optional one in the PICO question, since the researcher might study the intervention alone because either due to no interest in comparison or the lack of a comparable group.
- d. **O: Outcome to be measured as a result of the intervention.** It is the evaluation of the intervention's effect. This may include cure or level of control of a disease, efficacy of a medication or a diagnostic test, etc.

5. Criteria for a bad research question

There are different criteria that define a bad research question summarized below:[7,9,10]

- a. Vagueness of the research question.
- b. The research question is too broad to reach conclusive results.
- c. The research question might fail to reveal the relevance of the topic under investigation.
- d. Other criteria which are related to the logistic aspects of the study, such as the time needed to carry out the study, high cost with limited funding, limited resources such as availability of expertise, special equipment, and/or information.
- e. Research question might be limited by ethical considerations or requirement of authorities' approval.

6. Example of a research question

Topic of interest: Women's health
Narrowed topic: Women and cancer
Focused topic: Women smokers and breast cancer

PICO:
P = Women (age more than 35)
I = Cigarette smoking
C = No smoking
O = Breast cancer

Research question: Does smoking among women older than 35 years affect breast cancer risk compared to non-smokers?

Section 3: Conducting an effective literature review

1. Definition of a literature review

Research literature review is a systematic, precise, critical method for reading, analyzing, evaluating and summarizing, the existing body of completed and recorded work produced by researchers, scholars, and practitioners . It is one of the most important early steps prior to initiating any research study.^[11,12]

2. The aim of a literature review

Literature review is a very important step in planning a research project. The reasons for carrying out a literature review are listed and summarized below:^[12]

- a. It helps the researcher understand the existing body of knowledge in a specific medical field. Moreover, the researcher will be updated on the most recent findings in that field.
- b. It identifies areas of consensus and debate among different studies, and highlights the gaps in knowledge that exist in the literature, which in turn justifies carrying out the research project.
- c. It provides details of different research methodology that were adopted by different researchers, which in turn helps in adopting the most appropriate study methodology in the proposed study.
- d. It identifies other researchers who share the same research interests, who might act as support for future queries.

3. Characteristics of a good literature review

A good literature review is characterized by the author's efforts to search, evaluate, and critically analyze the relevant work in the field. A good researcher has to develop good searching expertise, which will allow him/her to efficiently search available resources through the different electronic engines. Once the relevant articles are identified, the researcher has to have enough expertise to assess the content of the article, in terms of relevance, validity, etc. Finally, the researcher has to be able to integrate the important and relevant work into his/her own research project, specifically from the methodology and results point of view.

A good review should have the following characteristics which make it of value:^[13]

- a. **Comprehensive:** Evidence should be gathered from all relevant sources.
-

- b. **Referenced:** Providing full references for reviewed papers.
- c. **Selective:** Using appropriate search strategies to find the most important evidence.
- d. **Relevant:** Focusing on related studies.
- e. **Balanced:** Providing objective evidence from papers with different findings.
- f. **Critical:** Following valid scientific critical appraisal of the literature.
- g. **Analytical:** Developing new ideas and understandings from the evidence.

4. Steps of a literature review

With the vast amount of scientific information available, searching and locating good literature on a research topic is a challenging task. The following steps provide a sense of how researchers should proceed in searching and reviewing the literature:^[12,14-16]

a. Develop a research question

The first step is to define a specific research question, which identifies the research or clinical problem the research is aiming to solve. Evidence based practice proponents advice using four elements in building the research question, specifically, the PICO (Patient, Intervention, Comparison, and Outcome).

b. Types of the sources used in a literature review

The term ‘sources’ refers to material needed to conduct the literature review, which could be summarized in three types:^[14,17]

- i. **Primary source:** Is a direct description of a research study written by a researcher who conducted the study.
- ii. **Secondary source:** Is a review of studies summarizing and providing new interpretations built from and often extending beyond the original study.
- iii. **Tertiary source:** Include perceptions, conclusions, opinions, and interpretations that are informally shared.

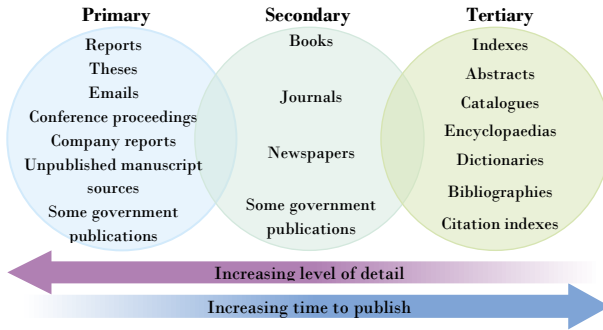


Figure: Literature sources available^[18]

Finally, a literature review might use a combination of primary and secondary sources since its purpose is to document and analyze what has been published on the given topic.

c. Search engines used in a literature review

There are different search engines that might be used to locate relevant material to be used in the literature review, which are summarized in the table below:

Database	URL	Characteristics
PubMed	www.pubmedcentral.nih.gov	• The National Library of Medicine's free search service. • Available through the World Wide Web. • Rapid updates of published material. • Some free full text material.
MEDLINE	www.nlm.nih.gov	• Contains abstracts and references from 1966 to the present. • Combines more than 3,900 medical and nursing journals into a single database. • Includes Canadian Journal of Medical Radiation Technology, Radiography and Radiologic Technology.

CINAHL	www.cinahl.com	- The Cumulative Index to Nursing and Allied Health contains abstracts and references from 1966 to the present. - As well as more than 1,700 journals, CINAHL provides access to health care books, dissertations, selected conference proceedings, standards of professional practice, and educational software.
Evidence-based Databases		
Cochrane Database	www.cochrane.org	- Includes full text of the regularly updated systematic reviews of the effects of healthcare prepared by The Cochrane Collaboration. - Free summaries are available; the full version is subscription only.
BMJ Clinical Evidence	www.clinicalevidence.bmj.com	The British Medical Journal's online decision-support resource describing the best available evidence from systematic reviews, randomized controlled trials, and observational studies.
Health Information Gateways		
Health on the Net	www.hon.ch	Non-profit, non-governmental organization, accredited to the Economic and Social Council of the United Nations for patients and health care professionals.
National Library for Health	www.library.nhs.uk	Funded by the UK's National Health Service for patients and health care professionals.

In addition, websites of different health organizations and associations could be a good source for literature review (such as: World Health Organization, Health Canada, Canadian Medical Association, etc.)

d. Establish the keywords and search strategy

To effectively conduct a search, keywords to be used in the search need to be identified, which are mainly the PICO elements. Moreover, a specific search

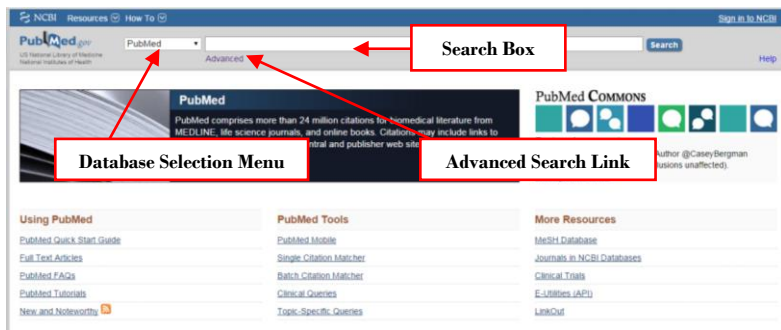
strategy using the *Boolean operators* (AND, OR, and NOT) should be developed, which can be used to combine the keywords and concept in a search.

e. Conducting the search

The search could be carried out by the different search engines detailed above, nevertheless, PubMed is the primary database for researchers in the fields of biochemistry, molecular biology, and related life sciences. It comprises over 20 million references to articles published in more than 5,200 current biomedical journals from the United States and over 80 foreign countries. It was developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM). It is one of several databases under NCBI's Entrez retrieval system (the text-based search and retrieval system used at NCBI for all of the major databases, including PubMed, and many others). PubMed can be directly accessed at: <http://pubmed.gov> or the National Library of Medicine's homepage: <http://www.nlm.nih.gov>.

The main screens of Pubmed

The most important features in PubMed which a researcher needs to be familiarized with are: the main page (including database selection menu, search box, and advanced search link), the search results page, and the My NCBI.



The **main page** includes a database selection menu, a search box, and an advanced search link.

- On the PubMed's home page the **database selection menu** is displayed, where the researcher can choose between PubMed and other NCBI databases (the last four databases you searched will appear at the top).

- A **search box** appears where keywords are entered. Automatic suggestions will display the search terms typed (click Turn off to temporarily disable the auto suggest feature). Any combination of search terms can be typed in the search box. Click the 'Search' button or 'Enter key' to launch the search.

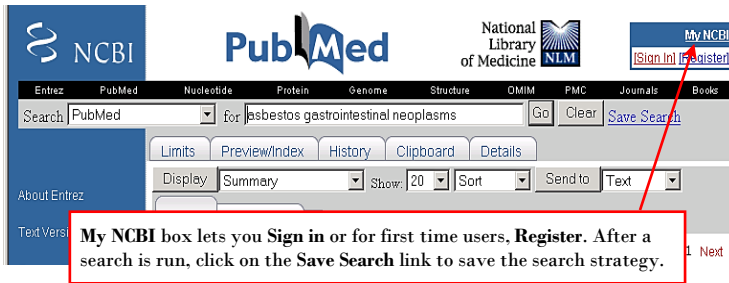
Following are some 'top tips' for focusing a search:

- Combine search terms with '**AND**' or '**OR**'
 - Use '**Limits**' (Age group, Publication type, language, etc.)
 - Search for your term as a word in the **title** or **title or abstract** (using Limits)
 - Use MeSH, with subheadings
 - Try PubMed's **Clinical Queries** or **Topic-Specific Queries**
 - Use the **Related Articles** link, once you find a set of relevant citations
- The **advanced search link** is where the researcher can construct a tailored search.

The results of the search are displayed in the **Search Results Page** (see below). Use the **Display Settings** menu to change the display format, the number of citations per page and the sort order of your results. To view selected citations: Click on the box found to the left of each item number of interest.

The screenshot shows the PubMed search results page. At the top, there's a search bar with the query 'Human genetic resistance malaria' and a 'Search' button. Below the search bar, there are links for 'PubMed', 'RSS', 'Save search', and 'Advanced'. The main content area displays 'Results: 1 to 20 of 801'. The first result is titled 'Monitoring parasite diversity for malaria elimination in sub-Saharan Africa' and lists authors Ghanisah A, Amenga-Etego L, Amambua-Ngwa A, Andagatu B, Apirion T, Bouyou-Akoton M, Cornelius V, Olotassia L, Andriananjana V, Ishengoma D, Johnson K, Kamau E, Maiga-Ascofaré O, Mumba D, Tiedana P, Tafelu-Kibuto A, Randriananelojosa M, Williams Y, Kwaikowski DP, Djimde AA. The second result is titled 'Tracing evolutionary relics of positive selection on eight malaria-related immune genes in humans' and lists authors Huang BH, Liao PC. The third result is titled 'Absence of the human CYP2C8*3 allele in Ugandan children exposed to Plasmodium falciparum malaria' and lists authors Paganotti GM, Cosentino V, Russo G, Tabacchi F, Gramolelli S, Coluzzi M, Romano R. On the right side, there are filters for 'New feature', 'Titles with your search terms', and '324 free full-text articles in PubMed Central'.

My NCBI is a feature of the NCBI databases that allows the user to save records and searches, and customize the results display with filters and other options. Moreover, updates on saved searches might be requested to be sent to a personal e-mail.



f. Choosing the material to be included in the literature review

The choice of the articles found in the literature search to be included in the literature review may seem confusing as the number might be overwhelmingly high. Thus, filtering the found articles is important to make sure that the most relevant ones to support the research will be chosen.

In narrowing the literature selection, more focused screening criteria are taken into consideration, such as:

- **Date of publication:** ex: only studies conducted between 2005 and 2012
- **Participants or subjects:** ex: children 6 to 12 years of age
- **Publication language:** ex: documents written in English
- **Research design:** ex: clinical trials
- **Authors:** ex: well-known author in a specific field
- **Journal:** ex: high impact journal, such as New England Journal of Medicine
- **Relevance:** ex: similar objectives addressed and methodologies adopted

Constructing a summary table of selected material will serve as a tool to organize a researcher's work.

g. Critically analyze and evaluate the information

Critical analyses of the chosen documents refer to the process of looking at each document closely, reading the introduction, methodology, results and discussion sections. As for the critical evaluation of the document, it refers to assessing the validity of the methodology adopted, and relevance of the results reported. Biases affecting each of the documents should be taken into consideration, and evaluated accordingly. Finally, integrate the reported results into the scope of the proposed research.^[16]

h. Cite literature properly

A researcher is supposed to protect the intellectual property of other researchers by acknowledging any work that has influenced the proposed research. This is done by citing other people's work by denoting the names of the authors, the title of the paper, the journal where it is published, as well the year of publication. More specifically, a citation in the body of the proposal or manuscript is to be included, as well as its corresponding details in the references section of the document.^[16]

There are four main reasons why it is important to cite literature properly:

- To acknowledge the author(s) of the work that the researcher used.
- To provide context to the research and demonstrate that the research is well-supported.
- To allow readers to find the original source and learn more about some aspects mentioned in the document.
- Avoid plagiarism, which occurs when a writer deliberately uses someone else's language, ideas, or other original material without acknowledging its source.

There are bibliographic management software programs that allow the researcher to search, collect and organize citations, and insert the citations into a word processing program in formatted bibliographic styles. Such available programs are: EndNote, Reference Manager, and ProCite.

5. Conclusion

An effective review will increase likelihood of funding, generate new ideas and directions for investigation, and improve the quality (and likelihood) of peer-reviewed publication of primary research.^[14] At all stages of the process it is vital that the search process is evaluated, since the inability to find relevant information can be attributed to a poorly constructed search strategy, inappropriate search terms, poor retrieval methods or inappropriate source.

Section 4: EndNote

1. What is EndNote

EndNote is a software program that works with Microsoft Word to automatically format in-text citations and end-of-paper reference lists with your chosen style (APA, MLA, Chicago, etc.). EndNote can also be used as a personal database to gather and store citation records from hundreds of online resources for references and PDFs including PubMed and many more.^[19]

By storing citations in an EndNote library rather than a Microsoft Word or Excel file, you can:

- Automatically insert well-formatted citations into your paper.
- Automatically reformat citation style in one click.
- Search live within your EndNote library.
- Sort your library by author, title or date.
- Build bibliographies in over 5,000 styles.
- Access and manage your research from your desktop, online, and your iPad.
- Find full text for your references in one click.

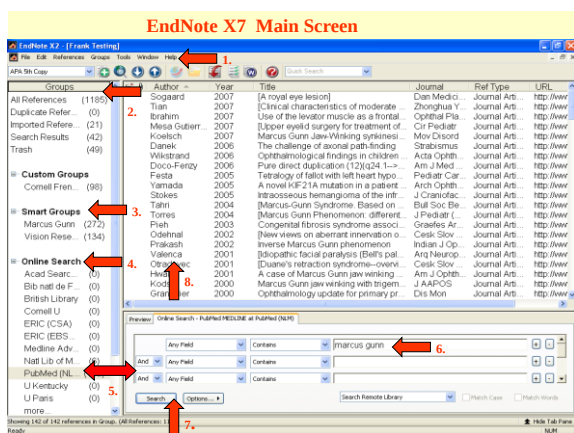


Figure: is a snapshot of the EndNote X7 Main Screen including references of a specific clinical topic.^[20] The main features of the screen are listed below according to the inserted red arrows:

- The main drop down menu
- Groups section, presents summary of available libraries

- Custom and Smart Groups are a list of the subset libraries created by the user
- Online Search is where users can select and search internet databases
- Selected online Database, the PubMed
- Search example is where users can search for a reference using multiple keywords combined by with AND, OR, NOT
- Search button is through which a user can perform the search process and bring the reference to EndNote

2. Introduction to an EndNote library

An EndNote library is a collection of references (or citations) that are usually on a single topic (i.e. a thesis). Each reference in a library is composed of different fields, such as author, date, journals, etc.

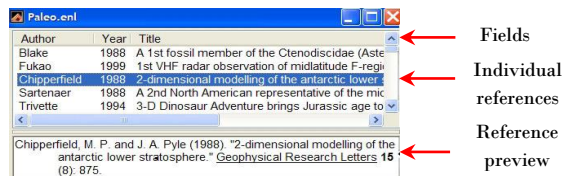


Figure: is an example of an EndNote library window screen shot showing the fields, individual references, as well as a preview of a selected reference.^[21]

3. To create a library

- Open EndNote
- Go to File
- Select New
- Name your Library
- Add Citations to an EndNote Library

4. Entering references into a library

There are 3 ways to enter references into an EndNote library:

- Type data by hand
 - At the top of the screen go to References
 - Choose New Reference
 - Select the correct Reference Type

- Enter the citation title, author, year, ISBN, abstract etc.
- EndNote will automatically save on closing
- b. Download Data from a database and import into your library. Most databases, for example PubMed, allow you to export records directly into EndNote.
 - Conduct a database search
 - Select references to export to EndNote
 - Locate and click on 'Export', 'Send to' or 'Download' button
 - The reference will be automatically exported to EndNote
- c. Connect directly to and search a database from EndNote
 - At the top of the EndNote screen, go to Tools
 - Online Search
 - Choose a Connection
 - Choose the name of the appropriate database (PubMed, Business Source Premier, etc.)
 - Enter your search and click “OK” to retrieve citations
 - Citations will be automatically copied to your EndNote library as soon as you finish the search

5. Inserting references and generating bibliographies in Microsoft Word

- a. From the EndNote library, highlight the reference you want to insert into the document
- b. Go to Word
- c. Put the cursor in the text where it needs to be inserted
- d. From the tools bar choose EndNote
- e. Choose Insert Citation
- f. Insert Selected Citation
- g. Change citation style of paper (MLA to APA, etc.)
- h. In your paper, choose Tools EndNote
- i. Format Bibliography
- j. Choose desired Output Style
- k. All EndNote tools are gathered in the “EndNote X3” tab at the top.^[22]

For example, here are some citations from Gray's Anatomy using several different styles:

- Gray H. Anatomy, descriptive and applied. 18th ed. Philadelphia: Lea & Febiger; 1910. (APA 5th citation style)

- Gray, H. (1910). *Anatomy, descriptive and applied* (18th ed.). Philadelphia: Lea & Febiger. (MLA citation style)
- Gray, Henry. *Anatomy, Descriptive and Applied*. 18th ed. Philadelphia: Lea & Febiger, 1910. (New England J Medicine citation style)

To get the EndNote software: <http://www.endnote.com/demo.asp>

Section 5: Writing a proposal

1. Definition of a research proposal

Research proposal is a very important step in the research process, as it is a summary of the suggested process to be used to answer the research question. It is done through gathering information, reading, integrating, organizing ideas, and planning. The proposal should include all key elements involved in the proposed research process and include sufficient information for readers to evaluate the project. It is intended to convince the reader that the investigator has an important research question to be answered, and that he/she has the competence to carry out the project. In summary, the research proposal should answer the following questions regarding a research project: Why, What, How, Where, and When?^[23,24]

2. Aims of a research proposal

The objectives of writing a research proposal are:^[24,25]

- a. **Research road map:** The research proposal serves as a road map, where each step of a research project is described in detail.
- b. **Ethical approval:** Research projects should obtain ethical approval prior to initiation. Ethical approval is granted through the IRB committee. A research proposal is an important document for IRB members to evaluate a research process to grant approval.
- c. **Funding:** In some cases, research needs financial support to be carried out, and thus, a research proposal will be an application based on which the funding agency will grant a fund to the researcher in order to carry out the study.

3. Sections of a research proposal

The format of a research proposal depends on the institution at which the study will be carried out, but the key point is to adhere to the specific format required. Nevertheless, almost all institutions require a research proposal including the following main sections: abstract, background and literature review, gap in knowledge, objective of the study, methodology, ethical considerations, budget, and references.

The following are brief descriptions of the different sections of a research proposal similar to a template provided in Appendix I:^[23,24,26,27]

a. Title, Investigators, and Affiliations

This subsection gives logistic information about a research project, such as: title, investigators, and their respective affiliations. The title should be concise, descriptive, and informative. It should accurately reflect the content and scope of the proposed study. Moreover, information about the name of the principal investigator (PI), the co-investigators, as well as their affiliations and contact information must be provided correctly. It is important to note that the PI and the department head at which the study will be carried out should sign the form indicating their knowledge and approval to conduct the study.

b. Abstract

An abstract should concisely describe the background of the proposal, critically evaluate the rationale to carry out the study or the gap in knowledge, and it should highlight the objectives of the project. It also should briefly describe the methodology, data analysis, as well as the significance and relevance of the project.

c. Background and literature review

The background and literature review subsection is necessary for investigators to provide a summary of the existing knowledge about the research problem as well as to find out whether or not others have investigated the same or similar problems, as well as to establish the link with other reported studies. Its purpose is to establish a framework for the research, so that readers can understand how it relates to other research. This is accomplished by a thorough and critical review of the literature. It should be written in a way to create interest about the topic, and lay the broad foundation for the problem that led to the proposed study. It should also clearly identify the gap in knowledge that needs to be filled through this study. This should in fact answer the question: Why is this study needed and what is its relevance? In addition, the output of the literature review process should demonstrate that the proposed research will contribute/add to the existing body of knowledge. Finally, it highlights the researcher's expertise and knowledge about the area in which the study will be carried out.^[24]

d. Objective of the study

The purpose of the study should be very clearly specified in a sentence or two that summarizes the research question to be answered by the proposed research.

A research project should have one primary objective, and one or more secondary objectives which should all be detailed in this section.

e. Methodology

The methods section is the heart of the proposal. It should be written in detail, because it is the main section that will be assessed by the reviewers. It is expected to follow a logical continuity in the research process. The following are some important subsections that a reviewer expects to find in the methods section, keeping in mind that all the subsections should be directly related to the objectives of the study:[25,27]

- **Study design:** In this subsection, the type of study design to be used to answer the research question must be provided, such as cross-sectional, case-control, cohort, or interventional study. A valid justification should be provided for selecting the study design, as well as the reason behind ruling out other study designs.
- **Setting:** The area or location where the study will be carried out is described in this section. More specifically, detailed description about the institution where the study will be carried out should be provided, such as, type of institution, load of patients, staff available, community served, as well services provided.
- **Time period:** It is important to set the time frame for the proposed work. The accomplishment of tasks or specific aims and the division of labor during that time-period must be clearly set and defined within time frames throughout the duration of the study.
- **Inclusion/exclusion criteria and sampling technique:** In this subsection, a description of the individuals who will participate in the research study and how they will be recruited should be provided. Inclusion criteria are the characteristics that are set by a researcher that will define eligibility, whereas exclusion criteria are those which make the subject ineligible. More specifically, the following are some examples for such criteria: demographic characteristics (such as age and gender), geographic location, disease status, pharmacological characteristics, lab results, etc. Finally, the procedure through which the eligible subjects are selected into the study is called the sampling technique. The most frequently used sampling technique is the random selection. Other sampling methods include convenient sampling, consecutive sampling, etc.
- **Sample size:** Sample size is a statistical calculation of the number of subjects that are required in a study to meet the objective. Fewer than needed number of subjects would lead to a study that is underpowered, whereas

more than needed number would lead to wasting resources on the additional subjects. Sample size should not be selected haphazardly by the investigator; rather it should be justified by proper statistical calculation.

- **Data collection methods:** This subsection sheds light on the data to be collected and the methods of collecting it. Information to be collected could be done by using questionnaires, data collection forms, online surveys, etc. Details of these forms should be specified in this section, as well as the validity and reliability of these tools. Moreover, the methods used for collecting the data should be specified, such as interviews, chart reviews, etc. Including the data collection form in the research proposal is important.^[24]
- **Apparatus or instruments:** Any apparatus and/or instruments proposed to be used in the research study should be listed and described in this subsection. The following information should be included: a general description of the apparatus or instruments to be used, why they are used, variables measured by these instruments, and finally their reliability and validity.
- **Exposure and outcome definitions:** In this subsection, detailed information about exposure and outcomes considered in the study should be provided. It is worth noting that some studies might include multiple exposures (with one being primary) and/or multiple outcomes (with one being primary). Mentioning which is the primary and which are the secondary exposures and outcomes is important.
- **Data analyses:** This subsection describes the proposed statistical analyses to be carried out, which should address different levels of analyses (univariate, bivariate, and multivariate), as well as information about the program to be used for data entry and analyses.

f. Ethical considerations

A specific subsection should be included about ethical considerations of the study which are required by the IRB. More specifically, the investigator has to provide information on how would confidentiality and anonymity of the subjects be maintained. It is also required to declare any conflicts of interest, for example, a researcher affiliated to a company closely related to the proposed area of research.

g. Budget

Budget is the total amount of money needed by the researcher to carry out the study. Details of the budget should be itemized and specified to allow judgment

of its appropriateness, based on the description of the research design and methods. The budget should be divided into parts, for example: personnel (salaries and benefits), supplies and equipment, other costs (compensation for subjects, transportation costs, and medical intervention costs), etc. Justification for each of the parts should be included, if necessary. Finally, it is important that the budget remains in line with the proposed timeframe of the study.^[27]

h. References

The references used in the write up of the proposal should be included in the sequence they appear in the proposal. Only references that are cited in the text should be included. They should be relevant to the topic and updated. Referencing programs, such as EndNote, could be used to facilitate the process of referencing.

i. Additional forms to be filled and submitted

Along with the research proposal, the following are some forms that might be needed to be submitted according to the institutional guideline.

- **Submission Checklist:** It includes all the required items in the submission
- **Cover Sheet:** It includes signature of the PI, each co-investigator, and the chairperson of the PI's department
- **Informed Consent Form:** For any research involving human subjects
- **Other forms:** Includes any additional forms needed

j. General tips

Finally, a general tip for the write up of the research proposal is to give the whole process enough time and not to do it in a rush and under pressure. A researcher has to keep an open eye on the deadlines for submission of his/her proposal. He/she has to know what to do before they start putting it into words. A good proposal would go through few rounds of writing and modification before finalization.

Section 6: Ethical issues and IRB

Ethics in clinical research involves a set of relevant rules considered in the conduct of clinical studies. The ethical conduct of a clinical research does not end with the formulation of the study design and a signature on the informed consent form. Protecting the rights, interests, and safety of research subjects must continue throughout the study duration. Subject safety monitoring is the responsibility of several groups, including IRBs, investigators and their research staffs, sponsors, and data monitoring committees (DMCs).

1. Ethical issues

Various organizations have created guidelines for human subject research for the various kinds of research and situations. A set of guidelines was published to guide researchers who work with human subjects. Good Clinical Practice (GCP) enforces guidelines on ethical aspects of clinical research. GCP aims to protect and preserve human rights through making sure that studies are conducted with high scientific standards with proper documentation. It provides principles on how clinical research should be conducted, define the roles and responsibilities of sponsors, clinical researchers, and investigators.^[28,29]

The following are a few points essential for ethical considerations in clinical research:

a. Informed consent

An informed consent is a voluntary agreement to participate in a clinical research study. It is not merely a form that is signed but is a process, in which the subject has an understanding of the research and its risks. It must be obtained for all types of human subjects research including: diagnostic, therapeutic, interventional, social and behavioral studies.

The goal of the informed consent process is to provide sufficient information so that a participant can make an informed decision about whether or not to enroll in a study or to continue participation. It is a document that should be written in a lay language, which makes it easily understood by the participant. It should include information about the objectives and procedures of the study, and potential risks and benefits of participation. It should indicate information about the right of the patient to withdraw from the study at any time without affecting the healthcare they are receiving. Moreover, the contact information of

the PI should be provided for any further communication and clarification. Subjects must be given sufficient time to consider participation. Informed consent should be obtained before enrolling a participant and its regulations should be maintained throughout the whole research process.^[30]

b. Vulnerable groups

Vulnerable groups should be given extra attention and precautions in terms of ethical considerations when to be included in clinical research. Researchers refer to populations which have low autonomy as "vulnerable populations". These are groups which may not be able to fairly decide for themselves whether to participate in clinical research or those who are relatively (or absolutely) incapable of protecting their own interests. Examples of groups which are considered as vulnerable populations include: prisoners, children, pregnant women, human fetuses, neonates, racial minorities, as well as people who are very sick, physically handicapped, mentally disabled, economically or educationally disadvantaged, or institutionalized.^[28,31]

2. IRB

An IRB, also known as an independent ethics committee or ethical review board, is a committee that has been formally designated to approve, monitor, and review the ethicality of biomedical and behavioral research, in protecting the rights of participants in a research projects.

a. Principles

The three main principles that guide the IRB in making its decision are derived from the Basic Ethical Principles enumerated in the Belmont Report document dated April 18, 1979. They are as follows:^[1,31]

- **Respect for Persons:** The principle of respect requires that subjects participating in the research should be fully aware of the nature of such research and assured that such participation is voluntary, with no pressure or duress. They should also be aware of the physical, psychological, and socio-economic risks that such participation might bring to the subject immediately or in the future. This requirement is imperative, even if the risks were described as minimal or insubstantial.
- **Beneficence:** Beneficence requires researchers to maximize the potential benefits to the subjects and minimize the potential risks.

- **Justice:** The principle of justice requires an equitable and fair selection of subjects and a fair and equitable distribution of risks and benefits of research.

b. IRB structure

IRBs are formed by academic, research, and other institutions, and they include at least five members of different professions, having enough expertise to make an informed decision on whether the research is ethical, informed consent is sufficient, and appropriate safeguards have been put in place.^[29]

c. Tasks of IRB

The IRB is empowered to:^[29]

- Review research proposals submitted to the committee for approval
- Identify the risks entailed due to participation in the research study
- Assess steps proposed by the investigators to minimize the risks on participants
- Assess safeguards to be used for maintaining confidentiality, anonymity, and well-being of participants
- Conduct risk-benefit analysis in an attempt to determine whether or not research should be done
- Give a final decision on the proposal submitted, which might be approving, requiring modifications, or disapproving the proposal

d. Review process

Before conducting any human research, the researcher must have the study approved by the IRB. Once the written proposal is completed, it is submitted to the appropriate IRB along with any additional materials needed (e.g., consent forms, test materials, and questionnaires). This proposal will be sent to reviewers who will provide detailed feedback pertaining to the ethical conduct of the study. Upon receiving the reviews, the IRB committee will meet to reach a final decision. Researchers are responsible for complying with all IRB decisions, conditions, and requirements. Upon granting approval to the investigators, the research project could be initiated.^[1,32,33]

Section 7: Data collection, entry, cleaning and management

The main step in a research study is to collect data that will be analyzed and verified statistically to reach a conclusion about the research question. It is important in this section that the reader should be acquainted with some of the most commonly used terms during the phase of data collection, which are explained below:

1. General terms

The following is a brief explanation of some of the most commonly used terms that relate to this phase of research when data is collected.

a. Variable

A variable is anything that is not fixed or has the potential to change. A variable in clinical research is any piece of information collected on a patient through the data collection process. The variability aspect of a variable is the potential of changing between different patients rather than within the same patient. For instance, gender is fixed for each patient, but it will be a variable when some patients are males and some patients are females.

b. Types of variables

There are 2 main types of variables:^[34]

- **Categorical variables:** They are those that have the responses falling into fixed categories. Different types of categorical variables exist, which depends on the potential responses.
 - **Dichotomous variable:** It is where the answer could be one of two possibilities, such as ‘hypertensive’ or ‘non-hypertensive’.
 - **Nominal variable:** It is a categorical variable with more than two responses, where there is no specified order in the responses, such as blood type A, B, AB, and O.
 - **Ordinal variable:** It is where a categorical variable could have more than two responses with a pre-defined order, such as severity of disease (mild, moderate, and severe).
- **Continuous variables:** They are those that consist of any value within the normal defined limits. An example of a continuous variable is age, blood pressure, weight, height, etc.

c. Coding

Coding is a step where non-numeric variables are given numbers. For instance, instead of entering 'F' for females, and 'M' for males, a code of '1' is given for males and a code of '2' for females.

2. Data collection

Data is defined as the raw facts that are not yet interpreted, organized or evaluated. Data collection is the systematic gathering of data to serve a particular purpose. There are different sources used for data collection, and these include questionnaires, interviews, observation, already existing records, or electronic means. Most of the times, data are collected on papers.

MI admission to an Emergency Department, Saudi Arabia		
SERIAL NUMBER:	_____	
A. DEMOGRAPHIC:		
MR NUMBER:	_____	
Age at presentation:	_____	
Sex:	☐ Male	☐ Female
Marital status:	☐ Married	☐ Single ☐ Widowed
Arabic speaking:	☐ Yes	☐ No
Nationality:	☐ Arab	☐ Non-Arab
Weight:	_____	
Height:	_____	
B. CLINICAL AND OUTCOME INFORMATION:		
1. Hypertension:	☐ YES	☐ NO
2. Diabetes:	☐ YES	☐ NO
3. Cigarette smoking:	☐ YES	☐ NO
4. Family history of CAD	☐ YES	☐ NO
5. Prior MI:	☐ YES	☐ NO
6. Prior CABG:	☐ YES	☐ NO
7. Prior Angioplasty	☐ YES	☐ NO
8. Date of presentation	_____	

Figure: is a snapshot of a data collection form for a study to assess characteristics of patients admitted with acute myocardial infarction to an emergency department in Saudi Arabia.

3. Database structure

A database is a collection of data that is organized so that it will be easily accessed, managed, and updated. Databases are used to manage and archive

large amounts of data. Databases could be structured in different computer programs, such as Microsoft Excel, Microsoft Access, etc. The most frequently used computer program used for data entry, as well as other data manipulations, is the Statistical Package for Social Sciences, or better known as SPSS.^[35]

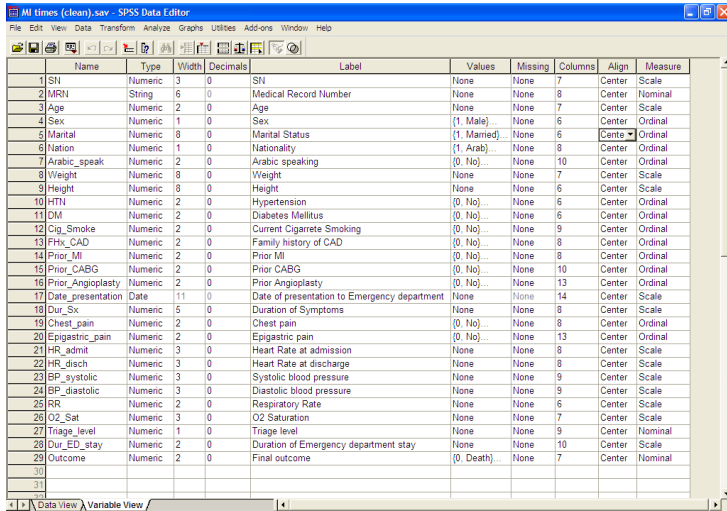
Although Microsoft Excel is usually used for data entry, the SPSS is considered superior for different reasons.

- a. Microsoft Excel does not allocate the first line for a patient's record; instead it is used to include the variable names. On the other hand, SPSS permits the user to define the variables included in the study, while keeping the first line for a patient's record.
- b. Microsoft Excel does not allow an easy coding of variables. This is important as all data to be analyzed have to be entered as numbers. In contrast, SPSS will provide an easy way to define the codes given to each variable.
- c. Unlike SPSS, Microsoft Excel could be used to analyze data, it needs an expertise to do it.

A database structure is a process where the variables included in the study are defined. Mainly the following are provided:

- a. **Name:** It is the short name of the variable that should not include any spaces.
- b. **Type:** It specifies the format of the variable, such as numeric (anything that includes only numbers), dates (anything where date is involved), or string (where a combination of letters, numbers or symbols are included).
- c. **Width:** It is where the space allocated for the specific variable is determined.
- d. **Decimals:** It is where the number of decimals needed for the variable is specified.
- e. **Label:** It is a place where the full definition of the variable could be provided.
- f. **Values:** It is where the coding of the variable is specified.

Other options for each variable are available that allow the user to control the variables.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	SN	Numeric	3	0	SN	None	None	7	Center	Scale
2	MRN	String	6	0	Medical Record Number	None	None	8	Center	Nominal
3	Age	Numeric	2	0	Age	None	None	7	Center	Scale
4	Sex	Numeric	1	0	Sex	(1, Male)	None	6	Center	Ordinal
5	Marital	Numeric	8	0	Marital Status	(1, Married)	None	6	Center	Ordinal
6	Nation	Numeric	1	0	Nationality	(1, Arab)	None	8	Center	Ordinal
7	Arabic_speak	Numeric	2	0	Arabic speaking	(0, No)	None	10	Center	Ordinal
8	Weight	Numeric	8	0	Weight	None	None	7	Center	Scale
9	Height	Numeric	8	0	Height	None	None	6	Center	Scale
10	HTN	Numeric	2	0	Hypertension	(0, No)	None	6	Center	Ordinal
11	DM	Numeric	2	0	Diabetes Mellitus	(0, No)	None	6	Center	Ordinal
12	Cig_Smoke	Numeric	2	0	Current Cigarette Smoking	(0, No)	None	9	Center	Ordinal
13	FHx_CAD	Numeric	2	0	Family history of CAD	(0, No)	None	8	Center	Ordinal
14	Prior_MI	Numeric	2	0	Prior MI	(0, No)	None	8	Center	Ordinal
15	Prior_CABG	Numeric	2	0	Prior CABG	(0, No)	None	10	Center	Ordinal
16	Prior_Angioplasty	Numeric	2	0	Prior Angioplasty	(0, No)	None	13	Center	Ordinal
17	Date_presentation	Date	11	0	Date of presentation to Emergency department	None	None	14	Center	Scale
18	Dur_Sx	Numeric	5	0	Duration of Symptoms	None	None	8	Center	Scale
19	Chest_pain	Numeric	2	0	Chest pain	(0, No)	None	8	Center	Ordinal
20	Epigastric_pain	Numeric	2	0	Epigastric pain	(0, No)	None	13	Center	Ordinal
21	HR_admit	Numeric	3	0	Heart Rate at admission	None	None	8	Center	Scale
22	HR_disch	Numeric	3	0	Heart Rate at discharge	None	None	8	Center	Scale
23	BP_systolic	Numeric	3	0	Systolic blood pressure	None	None	9	Center	Scale
24	BP_diastolic	Numeric	3	0	Diastolic blood pressure	None	None	9	Center	Scale
25	RR	Numeric	2	0	Respiratory Rate	None	None	6	Center	Scale
26	O2_Sat	Numeric	3	0	O2 Saturation	None	None	7	Center	Scale
27	Triage_level	Numeric	1	0	Triage level	None	None	9	Center	Nominal
28	Dur_ED_stay	Numeric	2	0	Duration of Emergency department stay	None	None	10	Center	Scale
29	Outcome	Numeric	2	0	Final outcome	(0, Death)	None	7	Center	Nominal

Figure: is a snapshot of a database structure for the study mentioned above done through SPSS.

4. Data entry

Data entry is the process of entering text or numbers into a computer program either by typing the fields through the keyboard, or by scanning the documents. It could be done through single data entry or through double data entry.

A single data entry means that the data is entered once into the computer while the double data entry is carried out when the data is entered twice into two different computers. This can be done for more control of data entry errors. The two databases will then be compared and any discrepancy will be resolved by the checking the original data collection forms.

The screenshot shows the SPSS Data Editor window for a file named 'All times (clean).sav'. The main window displays a data grid with 30 rows and 15 columns. The columns are labeled: SN, MRN, Age, Sex, Marital, Nation, Arabic_speak, Weight, Height, HTI, DM, Cig_Smoke, Fhs_CAD, Prior_MI, and Prior_Stroke. The data is entered as numbers, but the variable view at the bottom shows that some of these numbers represent categorical variables. For example, 'Sex' has values 1 for Male and 2 for Female; 'Marital' has values 1 for Married and 2 for Non-Married; 'Nation' has values 1 for Arab and 2 for Non-Arab; 'Arabic_speak' has values 1 for Yes and 2 for No; 'Weight' has values 1 for 65, 2 for 72, 3 for 85, 4 for 55, 5 for 58, 6 for 115, 7 for 43, 8 for 75, 9 for 75, 10 for 70, 11 for 85, 12 for 64, 13 for 86, 14 for 68, 15 for 94, 16 for 92, 17 for 42, 18 for 72, 19 for 65, 20 for 58, 21 for 60, 22 for 80, 23 for 71, 24 for 43, 25 for 87, 26 for 58, 27 for 69, 28 for 68, 29 for 70, and 30 for 68. The variable view at the bottom shows the variable names and their corresponding values.

Figure: is a snapshot of a database for the study mentioned above done through SPSS. Note that although the entries have been entered as numbers, they appear as words, since the coding for these numbers were specified in the “variable view” earlier.

5. Data cleaning

Data cleaning is an important step that is usually overseen by a lot of researchers. No matter how cautious the person is in data entry, errors still occur. Thus, data cleaning is a process where the data is checked for entry errors or extreme values in the dataset.

There are two steps in the data cleaning step:

a. Identifying the errors

This is a step where an error is identified through different methods. The most obvious type of data cleaning is to check the two extreme sides of the distribution of the variable. For instance, in a specific study, age might be expected to be between 18 and 80, thus by sorting the data in an ascending or descending order, one can identify any outliers or data entry mistakes.

b. Correcting the errors

Once the error is identified, the next step is to correct it. This could be done through going back to the data collection form used in the data collection, going

to the patient's record, calling the patient, or any other method to find the correct information. The method of correcting the errors depends on the method of data collection. Leaving an empty space on the data collection sheet (missing data) means failure to correct the error.

6. Data management

Data management is a crucial step in the preparation of the data for analyses. It includes steps to create new variables based on the available ones. There are three types of data management that researchers encounter in clinical research:

a. Recoding

As mentioned earlier, coding is assigning a number to a non-numeric variable; thus, recoding means to modify the coding of an already coded variable. For instance, smoking might be defined as: non-smoker (code = 0), ex-smoker (code = 1), and current smoker (code = 2). However, the researcher might be interested in having a variable that provides information on whether the patient is a smoker or a non-smoker. Thus, recoding could be done where the non-smokers and ex-smokers are grouped together in a new variable (code = 0), and current smokers (code = 1).

b. Categorization

It is a process where a continuous variable is divided into groups, or categorized. More specifically, a continuous variable is transformed into a categorical variable through this step. For instance, Systolic Blood Pressure (SBP) might be divided into three groups: hypotensive (SBP < 120), normal (SBP between 120 and 140), and hypertensive (SBP > 140).

c. Computation

It is a step where a new variable is created using a specific mathematical equation. A simple example of a computation process is calculating the Body Mass Index (BMI) using the following equation:

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height}^2 (\text{m}^2)}$$

Other examples include creating scores for different scales, such as Quality of Life, Depression scores, etc.

Section 8: Descriptive statistics

Once data is entered in a database, cleaned, and properly managed, the next step is to carry out the statistical analyses. Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It is divided into two parts: Descriptive statistics, and inferential statistics. In this section, descriptive statistics will be covered.

At this stage, a description of the sample collected, in terms of general characteristics, is done through some analyses. Descriptive statistics are the techniques used to summarize and describe the main features of a sample. It does not involve generalizing beyond what is available in the dataset.

The type of descriptive statistics that needs to be carried out depends on the type of variable under consideration.^[34,36] Thus, two types of descriptive statistics could be done, based on whether the variable is categorical or continuous variable.

1. Descriptive statistics for categorical variables

There are two types of descriptive statistics for categorical variables, either by providing frequency distribution or constructing a graph.

a. Frequency distribution

A frequency distribution lists, for each value (or small range of values) of a variable, the number and percent of times that observation occurs in the study sample. An example of “each value” is hypertension (either yes or no), whereas “small range of values” is when a continuous variable is categorized, such as SBP (< 120, 120 - 140, > 140).

The following is an SPSS output of the frequency distribution of hypertension (either yes or no) in a sample of 291 patients.

		Hypertension			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	217	74.6	75.1	75.1
	Yes	72	24.7	24.9	100.0
	Total	289	99.3	100.0	
Missing	System	2	.7		
Total		291	100.0		

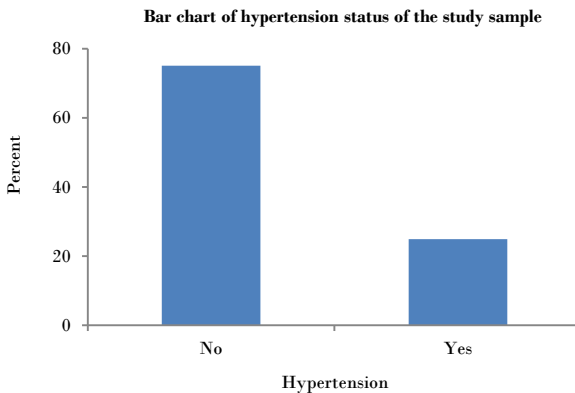
The first column “Frequency” presents the number of patients in each group, as well as those who have missing information. The second column “Percent” presents the percentage of each category out of the total, including the missing values. The third column “Valid Percent” is the percentage in each category among those who have known information, i.e. excluding the missing ones. Finally, the fourth column “Cumulative Percent” presents the addition of the valid percent at each level.

If this variable is to be summarized, providing the number alone will not be adequate since it will not be clear enough, as it is dependent on the total sample. Similarly, providing the percentage alone will not be clear as well, as it will have no reflection on the sample size. Thus, the right way to summarize a categorical variable is by providing both the number and the valid percent in the following format: 72 (24.9%).

b. Graphical representation

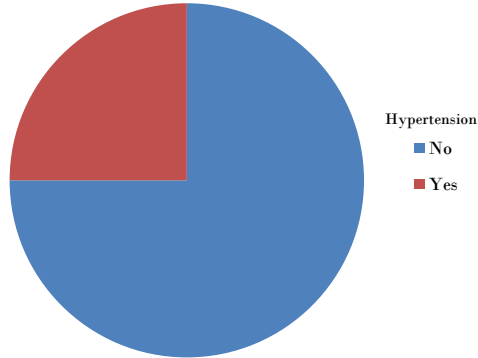
There are two types of graphs that are most frequently used to describe a categorical variable, and these are the bar and pie charts.

A bar chart consists of parallel, usually vertical bars with their lengths corresponding to the frequency or percentage of each value. The bars are separated from each other by a space as to reflect on the categorical aspect of the variable. Following is an example of a bar chart of the above mentioned hypertension variable.



A pie chart is a circular chart which is divided into sectors; each sector represents the corresponding frequency or percentage of each value. The following is an example of a pie chart of the hypertension variable.

Pie chart of the hypertension status of the study sample



2. Descriptive statistics for continuous variables

There are two types of descriptive statistics for continuous variables, either by providing summary values or constructing a graph.

a. Summary values

A continuous variable cannot be summarized by providing the frequency and percentage, on the contrary, specific numbers are provided to represent a variable. There are again two measures that are presented for a continuous variable:

- **Measures of central tendency:** These represent a value around most of the data cluster around. In other words, it is the value that best represents the continuous variable. Three measures of central tendency are available:
 - **Mean or average** is a value that is calculated by summing up each value of the continuous variable and dividing it by the sample size.

The formula of a mean is:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Where
$$\sum_{i=1}^n X_i = X_1 + X_2 + X_3 + \dots + X_n$$

One of the disadvantages of the mean is its sensitivity to extreme values. For example, the SBP of 5 patients are provided as follows: $X_1 = 120$, $X_2 = 80$, $X_3 = 90$, $X_4 = 110$, and $X_5 = 95$

The mean will be:

$$\bar{X} = \frac{120 + 80 + 90 + 110 + 95}{5} = 99 \text{ mmHg}$$

- **Median** is the middle number, or the number that divides the data in two halves. To calculate the median, the values of the continuous variables have to be sorted in an ascending or descending order, after which the middle value is chosen. For the above example, the median will be:

80 90 **95** 110 120

If the number of patients is even, then the median will be the average of the middle two values.

One of the advantages of a median is that it is not sensitive to extreme values, but the disadvantage would be that all the patients will be ignored except of the middle one.

- **Mode** is the most frequently occurring number. In the above example, there is no mode, since each value appears once. On the other hand, if the following data is provided for 6 patients; accordingly, the mode will be 110, as it appears twice compared to all the other values that appear only once.

80 90 95 110 120

- **Measures of dispersion:** These provide an index of how much there is dispersion or variability in the values of a continuous variable. The three measures of dispersion are:

- **Range** is the difference between the highest value of the continuous variable and the lowest value. For the following data, the range will be $120 - 80 = 40$.

80 90 95 110 120

One of the disadvantages of a range is ignoring all the values and taking only the lowest and the highest, and thus being sensitive to

extreme values (the range will be greatly affected by one extreme value).

- **Variance** is a measure of how much the values of a continuous variable are away from the mean. The formula of the variance is:

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Where X_i is the value of the first patient and $\bar{X}$ is the average, and n is the sample size.

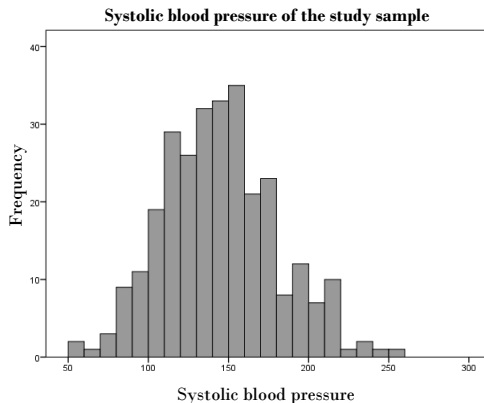
Accordingly, a small variance indicates that values are close to each other, whereas a large variance indicates that values are far away from each other.

- **Standard deviation (SD)** is the square root of the variance and the most frequently used measure of dispersion. The formula of the variance is:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

b. Graphical representation

The most frequently used graph constructed for a continuous variable is the histogram, which is similar to the bar chart except that the bars are connected to reflect on the continuity aspect of the variable. The following is an example of a histogram of the BP of patients in a specific study:



Section 9: Inferential statistics

1. Definition

Inferential statistics is a type of statistics that is used to draw conclusions on a population using data that was collected on a sample. It is the most important step taken by an investigator to generalize the results found in the study to the population under consideration. All clinical research is carried out to answer research questions postulated to a certain population, whereas studies usually entail taking a sample and answering questions on that sample. Thus, it is crucial to make the link between the results of the sample obtained and the population which is the target of the research question. This step is the inferential statistics. The link recognized is the inferential statistics.^[37]

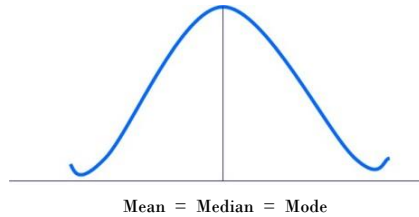
Example: To illustrate the concept of inferential statistics, a simple example is taken into consideration. Assume that a group of researchers are interested in assessing the SBP among patients admitted to the Emergency Department (ED) after an MI. Although this question applies to a total population which is any patient that meets the criteria (i.e. having had an MI and admitted to the ER); nevertheless, a sample is taken and the outcome (SBP) is assessed. In a hypothetical example, a sample of 291 patients were included in the study, and their SBP was found to be 144 mg/Hg (SD = 35). This calculation is what the descriptive statistics involves. The next step is to make inferences on the population, which is the inferential statistics.

Note that more complex examples are applicable, such as the RR, OR, etc.

2. Probability distributions

Inferential statistics is highly related to probability distributions. A probability distribution is a statistical function describing the probability of all possible values of a continuous variable. The most frequently used distribution is the normal distribution, which is unique in its characteristics, mainly being symmetrical and having its mean, median and mode equal to each other. It is completely described by the mean and its SD.

Another major characteristic of any normal distribution is the fact that the area (which is referred to the probability) between -1 SD and +1 SD is 68%, and it is 95% between -2 SD and +2 SD, and is 99% between -3 SD and +3 SD.^[37]



3. Types of inferential statistics

There are two main types of inferential statistics, mainly, confidence interval, and hypothesis testing (p-value). The following is a brief explanation of each.

a. Confidence interval

The confidence interval (CI) consists of 2 numerical values defining a range of values that with a specified degree of confidence includes the parameter being estimated.^[38] Practically, since a researcher is interested in estimating the average SBP in the specified population, it will be impossible to estimate this value with a single number, thus, the best way is to estimate it by providing a range of values, which are attached with a specific confidence. Of course, the narrower the CI indicates a better estimation, whereas the wider the CI indicates a worse estimation. Applying the normal distribution's characteristics, there will be a 95% probability that the average obtained will be between 2 SD's from the mean. The equation used for this calculation will be:

$$144 \pm 1.95 \times \left(\frac{35}{\sqrt{291}} \right)$$

The term $\left(\frac{SD}{\sqrt{n}} \right)$ is called Standard Error (SE), which is an error indicated by taking a sample. The bigger the sample size, the smaller the SE, and the smaller the sample size, the bigger the error. The reason the SD is divided by the square root of the sample size is to account for the sample size. In other words, the bigger the sample size, the estimate will be better by obtaining a CI which is narrower (by having a small SE), whereas the smaller the sample size, the estimate will be worse, by obtaining a CI which is wide (the SE will be big).

From the above calculation, the 95% CI was found to be 140 - 148. This indicates the population average SBP for the population under study is between 140 and 148, with a 95% confidence. Accordingly, there is an error of 5%.

b. Hypothesis test (p-value)

The other type of inferential statistics is related to a hypothesis testing, which is related to the p-value. The researcher will evaluate a hypothesis about a population rather than simply estimating it.^[38] A hypothesis is a statement that is not proven, which the investigator postulates, and according to the data obtained will either accept it or reject it. There are two types of hypotheses:

- **A null hypothesis (H_0):** This indicates there is no association or no difference
- **An alternative hypothesis (H_a):** This indicates there is an association or a difference

In the previous SBP example, the researchers were interested to compare the specific population's SBP to the normal value (considered 120).

Thus, the hypotheses will be:

- $H_0: \mu = 120$
- $H_a: \mu \neq 120$

Accordingly, a sample was taken and the average was 140. The next step is to decide whether this sample is supportive of the null or alternative hypothesis. There will be two scenarios:

- If the sample supports the null hypothesis, the conclusion will be that the population average is not significantly different from the normal value (120)
- If the sample supports the alternative hypothesis, the conclusion will be that the population average is significantly different from the normal value

The basis on which it is decided whether the sample supports the null or alternative hypothesis is mainly based on statistical grounds.

The difference between the observed value and the normal value is 20, and thus the question would be, "could this observed difference be due to chance?" If it is shown that this difference is due to chance then the conclusion would be that there is no significant difference, whereas if it was shown that this difference is too big to be due to chance then the conclusion will be that there is a significant difference.

The p-value is a probability that the observed difference is due to chance or due to sampling error. If this p-value is small (conventionally taken as < 0.05), then

it indicates that the difference is unlikely to be due to chance, which leads to accepting the alternative hypothesis and the conclusion will be that there is a significant difference. On the other hand, if this p-value is large (conventionally taken as ≥ 0.05), then it indicates that the difference is likely to be due to chance, which leads to accepting the null hypothesis and the conclusion will be that there is no significant difference.^[39]

The α level (also referred to as type I error) is the cut-off point at which significance is indicated, and is usually taken at 5% (0.05). More specifically, it is the error taking place when making the inference, and is the probability of rejecting a true null hypothesis. On the other hand, the β (also referred to as type II error) is the error when making the inference, and is the probability of accepting a false null hypothesis.

4. General rule for normal distributions

It is only logical that not all distributions of continuous variables are normally distributed, which implies inapplicability of the rules discussed earlier. It has been shown by mathematicians that the distribution used for the calculation for the confidence interval and p-value (called sampling distribution of the mean) is always normal, given that the sample size is larger than 30. This will be the reason why the above calculations are relevant even in non-normal.^[38]

5. Different statistical tests

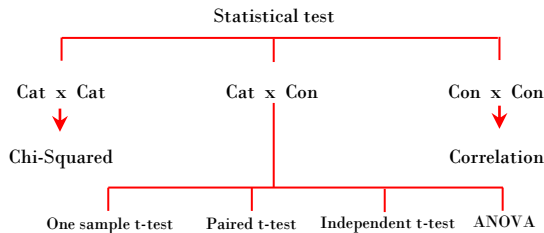
The p-values for different types of variables are calculated through different statistical tests. The association between two variables depends on whether they are categorical or continuous ones. The following is a list of the most commonly used tests for the calculation of p-values.^[38]

- a. **Chi-square test:** It is a test used for assessing the association between two categorical variables, such as association between physical activity (yes versus no) and hypertension (yes versus no).
- b. **One sample t-test:** It is a test used for assessing the difference between a continuous variable and a fixed reference value, such as the difference between the observed SBP and the normal value (of 120 as an example).
- c. **Paired t-test:** It is a test used to assess the association between two continuous measurements, which are related to each other, such as heart rate before and after taking a certain medication.
- d. **Independent t-test:** It is a test used to assess the association between two a categorical variable and a continuous one, where the measurement between

the two groups are independent, such as the SBP between males and females.

- e. **Analysis of variance (ANOVA):** It is a test to assess the difference between a categorical variable with more than two levels and a continuous one, such as the difference in age between patients of different severity of illness (mild, moderate, and severe).
- f. **Correlation:** It is a test used to assess the association between two continuous variables, such as SBP and age.

All of the above yield a p-value which is interpreted in the same way.



6. Sample size calculation

Sample size calculation is an important step carried out to calculate the number of subjects needed to be included in the research project.^[40] It is a statistical method used to do the calculation. Different research questions require the use of different equations for the calculation of the sample size. Typically, the following are the basic information needed for sample size calculation:

- a. Level of statistical significance (α), and is usually considered at 0.05 (5%).
- b. The value of the power desired ($1-\beta$), and it is usually considered at 0.8 (80%).
- c. An estimate of the expected prevalence or incidence rate in the control group (or unexposed), as well as the expected difference in response rates to be detected between the two groups. This could be achieved by considering the difference that would be clinically important in management of the specific patients, or from previous work carried out on the same topic.

Section 10: Measures of effect

Measures of association are used to compare the outcome between two groups, who are usually different according to their exposure status. The most relevant measure of effect used in clinical research is the risk, which is the probability of developing the outcome in a certain group. There are 2 types of risk.^[41]

1. Absolute risk

Absolute risk indicates the probability of an event to occur. The following is an example of a contingency table for the association between physical activity (exposure) and hypertension (outcome). In this group, 110 subjects were diagnosed with hypertension which gives an absolute risk of hypertension of 18.3%.

		Hypertension		Total
		Yes	No	
Physical Activity	Yes	30	170	200
	No	80	320	400
	Total	110	490	600

2. Relative risk

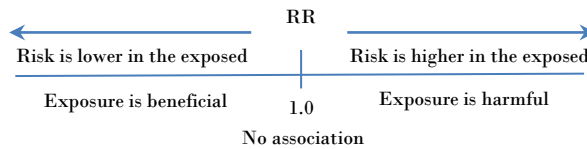
Relative risk is a measure to assess the association between two different groups. In the previous example, the risk of hypertension in the physically active group was 15%, whereas that for the non-physically active group was 20%. Thus, a relative risk is the ratio of these two risks, which yields a relative risk of 0.75.

The following is an interpretation of three different values of relative risk:

- A relative risk of 1:** It indicates that the risk of the outcome in the exposed group is the same as that in the unexposed group, indicating no association between the exposure and outcome. For instance, in the above example, if the risk of hypertension in the physically active group was 20% and that in the non-physically active group was 20%; then, the relative risk will be 1. The conclusion would be that whether the person is physically active or inactive, the same risk of hypertension applies.
- A relative risk of more than 1:** It indicates that the risk of outcome in the exposed group is greater than that in the unexposed group. For instance, in the above example, if the risk of hypertension in the physically active group was 20% and that in the non-physically active group was 10%, the relative risk will be 2. This indicates that physically active people have more risk of

being hypertensive as compared to those who are non-physically active, and more specifically, their risk is twice, or there is 100% increase in the risk of hypertension.

- c. **A relative risk of less than 1:** It indicates that the risk of outcome in the exposed group is less than that in the unexposed group. For instance, in the above example, if the relative risk was found to be 0.75 (15% in the physically active group and 20% in the physically inactive group), which indicates that physically active people have less risk of being hypertensive as compared to those who are non-physically active, and more specifically, there is 25% decrease in the risk of hypertension.



3. Risk difference

Risk difference is also called attributable risk and it measures the difference in the risks between the exposure groups. The following is an interpretation of three different values of risk difference:

- a. **A risk difference of 0:** It indicates that the risk of the outcome in the exposed group is the same as that in the unexposed group, indicating no association between the exposure and the outcome. For instance, in the above example, if the risk of hypertension in the physically active group was 20% and that in the non-physically active group was 20%, the risk difference will be 0. The conclusion would be that there is no effect physical activity on hypertension.
- b. **A positive risk difference:** It indicates that the risk of the outcome in the exposed group is bigger than that in the unexposed group. For instance, in the above example, if the risk of hypertension in the physically active group was 20% and that in the non-physically active group was 10%, the risk difference will be 10%. This indicates that there is 10% increase in the risk of hypertension which is due or attributable to physically activity.
- c. **A negative risk difference:** It indicates that the risk of outcome in the exposed group is smaller than that in the unexposed group. For instance, in the above example, the risk difference was found to be -5 (15% in the physically active group and 20% in the physically inactive group), which indicates that there is 5% decrease in the risk of hypertension which is due or attributable to physically activity.

Section 11: Overview of study designs

A study design is a scientific method that a researcher follows to assess the association between an exposure and an outcome. Mainly, it is the method based on which subjects included in the study are selected, observed, followed, and studied.

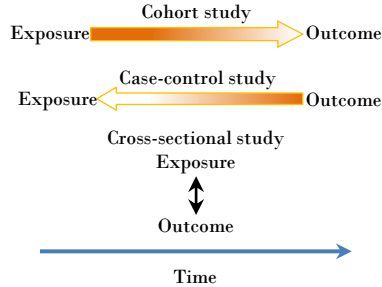
In clinical research, there are two broad categories of study designs, mainly observational and experimental:

1. Observational studies

Observational studies are studies where the investigator assesses the association between an exposure and an outcome by just observing what is happening among a group of subjects. Thus, the allocation of subjects in different groups of the exposure is beyond the control of the investigator. Under observational studies, there are two broad categories of studies, mainly:

- a. **Descriptive studies:** It is where the main objective of the research is to describe the main features of the population. There are different types of descriptive studies, mainly:
 - **Case report:** It is a report that documents unusual medical occurrences that can represent the first clue in the identification of new disease or adverse effect of exposures (based on one patient). It includes a summary of the disease, such as the presentation signs, symptoms, diagnostic studies, treatment course and outcome.
 - **Case series:** It is a collection of different case reports, thus based on more than one patient
 - **Cross-sectional study:** It is a study in which data is collected at one point in time and which reflects information on prevalence of outcomes.
- b. **Analytical studies:** These are studies designed for research with the main objective being beyond description, and where a specific association is assessed. The main two types of analytical studies are:
 - **Case-control study:** It is a study where participants are recruited based on their outcome status; a group having the disease and another who are disease-free. Information on past exposure is collected and compared between the two groups.
 - **Cohort study:** It is a study where participants are recruited based on their exposure status; a group having the exposure under study, and

another group who are exposure-free. The subjects are followed in time to observe the occurrence of the outcome in both groups.



2. Experimental studies

Experimental studies are those studies where an investigator allocates different exposures to subjects based on different criteria, mainly in a random fashion, and consequently, the investigator is no longer observing the subjects, but is actually carrying out an experiment on them. The main type of experimental studies is the randomized clinical trials.

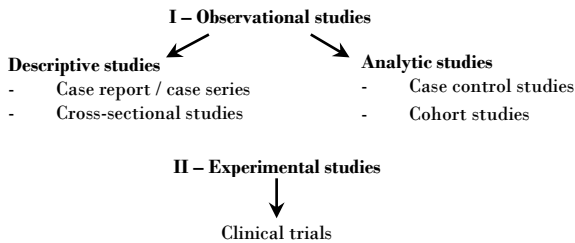


Figure: illustration of the different study designs available in clinical research

3. General terms

The following are some terms that are useful when addressing study designs:

a. Prevalence

Prevalence is an indication of how frequent a specific outcome or disease is present in a specified population. It is calculated by dividing the number of subjects with a particular outcome or disease in a given population by the total number of people in that population at a specific point in time. It is represented

as a fraction, proportion, or a percentage. For example, the prevalence of diabetes in the Kingdom of Saudi Arabia was estimated to be 30%, calculated by dividing the number of diabetic patients (1,792 patients) over the total number of subjects included in the study (6,024 subjects).^[42]

b. Incidence

Incidence is an indication of how fast the outcome or disease is growing in a specific community. It is calculated by dividing the number of newly diagnosed subjects with a specific outcome or disease by the total number of people in that population over a specific period of time. For example, the incidence of intrauterine fetal deaths after 26 weeks of gestation was found to be 6.1 per 1,000 total births, which was calculated by dividing the number of deaths (103 cases) by total number of pregnancies (16,882).^[43]

c. Exposure and outcome

Exposure and outcome are two terms that researchers use when they carry out clinical research. An exposure might also be called a risk factor or an independent variable, and they all reflect the variable one is interested in seeing; the effect of on a certain outcome, also called disease or dependent variable. As an example, a researcher might be interested in studying the effect of physical activity (exposure) on obesity (outcome), but another researcher might be interested in studying the effect of obesity (exposure) on diabetes (outcome).

d. Contingency table

Contingency table also called a two by two table or cross tabulation, is a method used to display the frequency distribution of two variables across each other, for the purpose of assessing the association between these two variables. It is a method used in the different study designs, but what differs is the measure of association calculated, which is dependent on the type of study carried out. The following is an example of a contingency table for the association between physical activity (exposure) and hypertension (outcome):

		Hypertension		Total
		Yes	No	
Physical Activity	Yes	30	170	200
	No	80	320	400
	Total	110	490	600

This two by two table indicates the number of subjects in each of the different categories of both the exposure and the outcome. There are 200 physically active

subjects, and 400 non-physically active subjects, whereas there are 110 and 490 hypertensive and non-hypertensive subjects, respectively. There are 30 subjects who are physically active and hypertensive.

4. Biases in clinical research

Bias is the systematic error that occurs in the design, conduct, or analyses of a study, which results in a wrong measure of association. There are three types of biases:^[41]

a. Selection bias

Selection bias is a systematic error that occurs at the stage where improper procedures are followed when recruiting subjects into the study. Selection bias occurs when the results in the measure of association for those who participated in the study are different from those in the target population. An example of selection bias is when subjects volunteer to participate in a study, since those volunteers might have different characteristics than those who do not. Another example is the “loss of follow-up”, as those subjects lost over the course of the study might be different than those who remain in the study. The way to avoid selection bias is by employing strict inclusion/exclusion criteria, as well as applying a random selection from the target population.

b. Information bias

Information bias is another systematic error that occurs at the stage of data collection, where the information collected are inaccurate, and this could be at the level of the exposure, the outcome, or other factors. This could take place when imprecise measurement instruments or invalidated questionnaires are used. Another source of information bias is introduced by subjects, such as recall bias, or reluctance to tell the truth. A third source of information bias is introduced by the interviewer when he/she probes for specific answers. In general, information bias is avoided by using validated questionnaires as well as following a structured method to collect information.

c. Confounding

Confounding is the third type of bias where a specific characteristic (that is associated with the exposure and is a risk factor for the outcome) leads to a wrong measure of association. A classic example about confounding is when the association between coffee drinking and lung cancer is assessed. Although it has

never been reported that coffee drinking increases the risk of lung cancer; yet, confounding might lead to a conclusion of a positive association. This is due to the confounding effect of smoking, which is associated with the exposure (coffee drinking; those who drink coffee are more likely to smoke), and to the outcome (lung cancer; those who smoke are more likely to develop lung cancer). Thus, confounding is about mixing the effect of an extraneous variable (confounder) with the effects of the exposure on the outcome under study.

There are 5 ways to control for confounding:

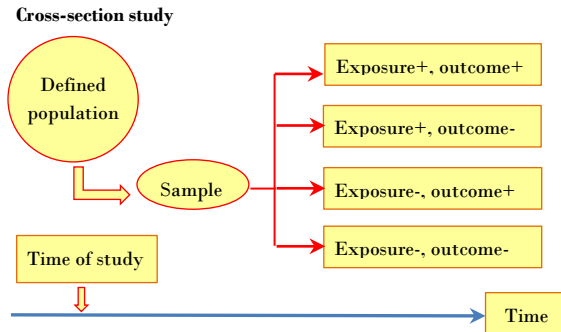
- Randomization (in clinical trial)
- Restriction (including subjects who are homogeneous in terms of the confounder)
- Matching (in case control studies)
- Stratification (assessing the association separately for the levels of the confounder)
- Multivariate analyses (statistical technique that removes the effect of the confounder)

The most effective way to control confounding is the multivariate analyses which necessitates collecting relevant information on the potential confounder so that it can be adjusted in the analyses.

Section 12: Cross-sectional design

1. Definition

A cross-sectional study is an observational descriptive study, where the main objective is describing a particular status among a specific community. Assume for instance, a researcher is interested in estimating the prevalence of hypertension in the Kingdom of Saudi Arabia. If a cross-sectional study is to be carried out, the researcher will recruit a group of subjects, and assess the hypertension status among them and accordingly will be able to estimate the prevalence of hypertension. Such a study is observational since the researcher is not intervening in any way with the subjects, whereas the researcher is only observing the subjects and recording information on them.^[41]



A cross-sectional study represents a snap shot about a particular status in a specific community, at a specific point in time (i.e. one time). This is where the cross-sectional aspect of the study comes from. Although the study might take few months to be conducted, the cross-sectional aspect is specific to each subject, where he/she will be assessed only once, and no follow-up is done. Cross-sectional studies are usually carried out to assess the burden of disease or health needs in a specific population, and to plan and allocate health resources.

As an example, in 2011, a study was carried out by Saeed et al. to estimate the prevalence, awareness, treatment and control of hypertension among Saudi adult population. Out of the 4,758 subjects included in this cross-sectional study, 1,213 were found to be hypertensive, which gave a prevalence of 25.5%.^[44]

Although the primary objective of the cross-sectional study is descriptive in nature, such as estimating the prevalence; most cross-sectional studies will report

some analytical associations between an exposure and an outcome, which will be considered as a secondary objective. For instance, in the above mentioned study, the authors assessed the association between gender (exposure) and hypertension (outcome) and reported the results.

		Hypertension		Total
		Yes	No	
Gender	Males	634	1,706	2,340
	Females	579	1,839	2,418
	Total	1,213	3,545	4,758

Accordingly, the authors reported 27.1% prevalence rate of hypertension among males (634/2,340) as compared to 23.9% prevalence rate of hypertension among females (579/2,418). From these numbers, it was found that a higher prevalence of hypertension was among males as compared to females. To further quantify the association between gender and hypertension, these two prevalence rates might be divided by each other to obtain a prevalence rate ratio of 1.13.

The association between the risk factor and the outcome under consideration, as calculated through a cross-sectional design, might not be considered valid since both the exposure and outcome were assessed simultaneously which is usually hard to confirm which came first; the exposure or the outcome. A bias that might be introduced in such a scenario is called “reverse causality bias”. As an example, when assessing the association between depression and cancer, a cross-sectional study will be biased as it will be hard to ascertain whether depression preceded or came after cancer diagnosis.

2. Choosing a representative sample

One of the most important challenges of a cross-sectional study is the selection of a representative sample. For instance, in a study assessing the prevalence of hypertension among Saudi adult population, the sample should include a random sample of all Saudi adult population; otherwise the results will not be valid because of selection bias.

3. Strengths and limitations

The few strengths of cross-sectional studies are:

- a. Relatively easy to conduct, as the study methodology involves identifying a group of subjects, collecting information on them, and analyzing the data.

- b. It is a study that could be done over short periods of time, since there is no follow-up involved.
- c. Data on all variables (exposure, outcome, and confounders) are only collected once.
- d. Possibility of assessing prevalence for all factors being studied.
- e. Multiple exposures and outcomes can be studied.
- f. The prevalence of disease or other health related characteristics are important in public health for assessing the burden of disease in a specified population and in planning and allocating health resources.
- g. Good for providing descriptive results and for generating hypotheses.

The few weaknesses of cross-sectional studies are:

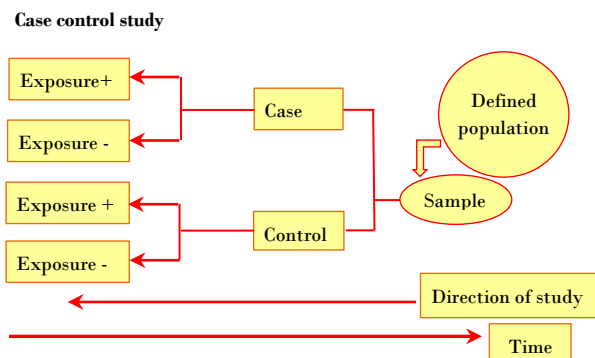
- a. Time sequence of exposure and outcome is not implicated in this design, i.e. it is difficult to determine which one came first, the exposure or the outcome.
- b. Not suitable for studying rare diseases, since very few subjects will be captured in the study, even if it includes a big number of participants.
- c. The incidence of a disease cannot be calculated through this study design.
- d. Susceptible to selection bias because of sampling and low response.
- e. Susceptible to information bias because of recall.
- f. Susceptible to confounding bias due to inability to collect information on all potential confounders.

Section 13: Case control study designs

1. Definition

A case-control study is an epidemiological study where participants are selected based on their disease status. Two groups of subjects are included in this type of study, cases (having the disease under consideration) and controls (not having the disease). Accordingly, both cases' and controls' history of the exposure and other risk factors is checked. Exposure history is compared between the cases and controls to assess whether an association exists between the exposure and the outcome. Thus, sampling of participants into the study is based on the outcome status.^[41]

Case control studies are also called retrospective studies as they assess exposure in the past, and since the direction of the study is backwards. Both exposure and outcome would have already occurred in a case-control study.



For example, a study was carried out to identify the prominent maternal and neonatal risk factors associated with early-onset group B streptococcus (EOGBS) disease in neonates. The authors carried out a case-control study where cases were infants < 7 days of age with invasive group B streptococcus (GBS) disease and controls were healthy infants born in the same hospital during the same period having the same birth weight and gestational age category.^[45]

2. Identifying cases

Case ascertainment can either be retrospective (prevalent cases) or concurrent (incident cases). Incident cases are those derived from ongoing-ascertainment of

cases over time, whereas prevalent cases are derived from a cross-sectional survey.

One very important issue in selection of the cases is to ensure the representativeness of the selected sample. Ideally, cases should represent a random sample of all cases of interest in the source population (e.g. from vital data, registry data). More commonly cases are a selection of available cases from a medical care facility (e.g. from hospitals, clinics).

3. Identifying controls

If cases are a random sample of cases in the population, then controls should be a random sample of all non-cases in the population sampled at the same time. One general rule is that controls should be at risk of the disease. The controls should resemble the cases in all aspects except for the presence of disease (and any as yet undiscovered risk factors for disease). To insure comparability between the cases and controls, different types of controls have been used in different case-control studies, and these include: hospital controls (those admitted to the same hospital for reasons not related to the disease under study), neighbours (door to door, phone, etc.), friends or associates of cases, or population-based controls (a members of a population of a defined area). The selection of one type of controls over the other depends on factors such as feasibility, representativeness, convenience, logistics, cost, etc.

The number of controls selected for each case varies between different studies. It is well established that as the ratio of controls to cases increase, the efficiency of the study will increase, which will reach a plateau at 4 controls per case.^[46] Thus, most case-control studies include between 2 to 4 controls per each case.

4. Exposure assessment

Once the cases and controls are identified, information on past exposure history, as well as information on other risk factors are collected through different means. Most frequently used method of exposure collection is through questionnaires, where the participants are interviewed for detailed recollection of their past exposures. A more standardized way of data collection is using already existing databases, such as drug registry.

5. Matching

To ensure comparability between cases and controls in terms of strong confounding variables, control selection is usually done through matching. This

indicates the process of selecting controls so that they are similar to the cases in regard to certain characteristics. The three most commonly used matching criteria are usually age, gender, and time.

In the paper summarized earlier,^[45] cases and controls were matched on time, weight and gestational age, where controls had to fall in the same category as the cases.

One of the disadvantages of matching is that matching on many variables may make it difficult or impossible to find an appropriate control. Moreover, once a variable has been used to match on, it cannot be used to explore possible association of disease with it.

6. Analysis

In the study summarized above, the authors assessed the association between fetal tachycardia and B streptococcal infection, and the results are summarized in the following table.

		B streptococcal infection		Total
		Cases	Controls	
Fetal tachycardia	Yes	22	7	29
	No	77	193	270
	Total	99	200	299

The calculation of the prevalence rate ratio or the risk ratio will yield wrong estimates, since any of those measures will be dependent on the number of controls per case recruited into the study, which is decided upon by the investigator. Thus, an alternative measure is calculated which is specific to case-control studies, and is called the Odds Ratio (OR).

The OR is calculated by dividing the odds of exposure among the cases over the odds of exposure among the controls. To illustrate the concept of the OR, the following is a hypothetical table:

		Cases	Controls	Total
		a	b	a+b
Exposure	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	a+b+c+d

- The definition of odds is a ratio between two numbers.

- The odds of exposure among cases = a / c
- The odds of exposure among controls = b / d
- Thus, the OR will be calculated by dividing the two odds which will be equal to $(a * d) / (b * c)$

An OR can approximate the relative risk in instances where the disease prevalence is low ($< 10\%$). The OR is interpreted in the same manner as the relative risk with an OR = 1.0 indicating no association, an OR > 1.0 indicating a positive association, and an OR < 1.0 indicating a negative, or protective association. In the above example, the OR will be 7.88, which indicates a positive association between fetal tachycardia and B streptococcal infection.

7. Strengths and limitations

The few strengths of case-control studies are:

- a. Being fast to be carried out, as no follow-up is needed, and at the time of conducting the study, both exposure and outcome have already took place.
- b. It involves low financial cost, again due to the fact that all the data is collected in a relatively short period of time, and the absence of follow-up.
- c. Appropriate for studying rare diseases, since patients are selected based on the outcome status, which gives the opportunity of recruiting an adequate sample of the cases from a specialized health care facility.
- d. Appropriate for assessing the association between different risk factors and one outcome.

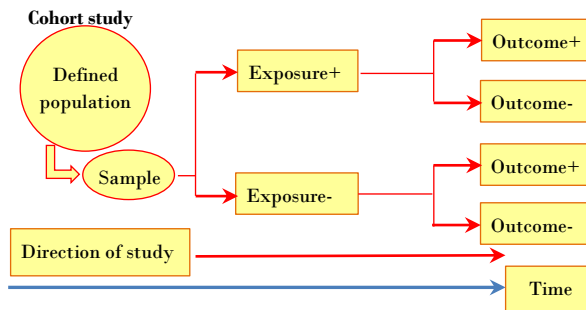
As for the limitations of case-control studies, they include the following:

- a. Prone to selection bias, if the cases or controls are not representative of the underlying population.
- b. Prone to information bias, since information collected has happened long in the past, and it might not be accurate due to recall (people with a condition will be more motivated to recall details about past exposures).
- c. Affected by confounding, as it might not be feasible to collect information on a wide range of confounding variables.
- d. Not appropriate for studying rare exposures, since not enough subjects (whether cases or controls) will be found to be exposed.

Section 14: Cohort design

1. Definition

A cohort study is an epidemiological study where participants are selected based on their exposure status. Two groups of subjects are included in this type of study, exposed (having the exposure under consideration) and unexposed (not exposed to the factor under consideration). Both exposed and unexposed subjects are free of the disease under consideration, but they should be at risk of developing it. Subjects are then followed prospectively over time to identify those who develop the disease. Incidence of the disease will be compared between the two groups to assess whether an association exists between the exposure and outcome. Thus, sampling of participants into the study is based on the exposure status.^[41]



2. Types of cohort studies

Cohort studies could be either prospective or retrospective.^[41]

- Prospective cohort study is what has been discussed earlier. On the contrary, a retrospective cohort study is exactly the same as the prospective one, with the exception being that both exposures and outcomes have already existed in the past. Usually retrospective cohort studies are carried out using existing medical records, where subjects are still divided into exposed and unexposed but it is the exposure that happened in the past. The outcome is then retrieved from the charts, which implies that the investigator does not follow subjects as it is in the prospective cohort studies.

For example, a prospective study was carried out in the Kingdom of Saudi Arabia to assess the effect of BMI in early pregnancy on pregnancy

outcome. Women during their first month of pregnancy were included in the study. Their BMI was recruited and divided into four groups. Follow-up of the patients was done till delivery and outcomes were measured, after which the association between BMI and those different outcomes were assessed.^[47]

- b. Another study was a retrospective cohort study which was carried out to investigate the rate of neonatal intensive care unit (NICU) admissions in pregnant mothers with gestational diabetes in the Kingdom of Saudi Arabia. The medical records of a specific hospital were reviewed and women with gestational age and another group of healthy women were included. Information on NICU admissions were also extracted. The association between gestational diabetes and NICU admissions was reported.^[48]

3. Strengths and limitations

The few strengths of cohort studies are:

- a. Easier and cheaper than experimental studies.
- b. Can directly measure disease incidence, since there is a follow-up inherent in the study design.
- c. At the time of sample selection, the outcome has not occurred yet, and thus temporal relationship can be established between exposure and outcome, thus building evidence for causality.
- d. Appropriate for assessing the association between one exposure and different outcomes.
- e. Appropriate for studying rare exposures, since patients are selected based on the exposure status, which gives the opportunity of recruiting an adequate sample of the subjects with the exposure under consideration.
- f. Being fast to be carried out, if it was a retrospective cohort study.

As for the limitations of cohort studies, they include the following:

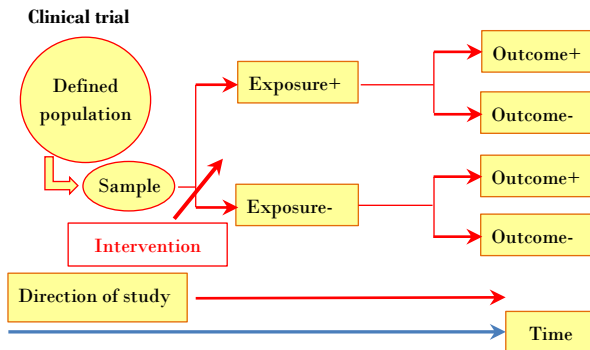
- a. Expensive and takes a long period of time, especially if it was a prospective cohort study.
- b. Imbalances in subject characteristics could exist, which introduces bias.
- c. Affected by confounding, as it might not be feasible to collect information on a wide range of confounding variables.
- d. Not appropriate for studying rare outcomes, since not enough subjects (whether exposed or unexposed) will develop the outcome upon the follow-up.
- e. Loss to follow-up, especially when the duration of follow-up is very long, which might introduce selection bias.

Section 15: Experimental studies

1. Definition

A clinical trial is an epidemiological study where participants are selected based on their exposure status, just like a cohort study. The only difference between a cohort study and a clinical trial is the fact that cohort studies are observational in nature (the researcher does not intervene in assigning the exposure), whereas a clinical trial is experimental in nature (the researcher intervenes and is the one who assigns the exposure to subjects). Thus, a clinical trial is a prospective study where an intervention is allocated to different groups of subjects and are followed-up over time to identify those who develop the outcome under consideration.^[49]

Since a clinical trial is about assigning the exposure to subjects, it imposes the type of exposure to be beneficial in nature, such as vaccines, drugs, treatments, or devices. Mainly, clinical trials are used to assess whether the new intervention is safe, efficacious and effective. Incidence of outcome is compared between the two groups to assess whether there exists an association between the exposure and outcome.



2. Phases of clinical trials

There are different phases of clinical trials that depends on the stage at which the therapeutic agent is at;^[49] which are mainly as follows:

- a. **Phase I (clinical/pharmacological study):** It is the phase of clinical trials that assesses the safety (toxic and pharmacological effects) of the therapeutic

agent, rather than the efficacy of it. Usually it includes a small number of healthy volunteers. The outcomes assessed are usually dose tolerance and frequency, duration of exposure, as well as absorption, distribution, metabolism and excretion.

- b. **Phase II (initial clinical investigation):** It is the first clinical investigation of the safety and effectiveness of the therapeutic agent at a given dose. Such clinical trials include a comparison group. Included in this type of clinical trials are volunteer patients who are monitored in a hospital setting. The outcomes assessed are all types of adverse events that might be due to the agent under consideration, as well as the pharmacokinetics of the drug.
- c. **Phase III (full scale evaluation):** It is the first large-scale controlled trial which assesses the effectiveness of the therapeutic agent under consideration. Included in the study are patients recruited from hospital or clinic settings. The outcomes are all possible adverse events. A drug is approved for marketing once it passes this phase of clinical trials.
- d. **Phase IV (post-marketing surveillance):** It is the phase of a clinical trial where the long-term effects and monitoring of the therapeutic agent which is being used in the market is assessed. Included in the study are patients from hospitals and clinics and are usually big in number.

3. Types of clinical trials

There are different types of clinical trials based on the main objective of the study:^[49]

- a. **Superiority trial:** It is the most frequently used type of clinical trials, where the objective is to show that a new therapeutic agent is superior to the conventional one.
- b. **Equivalence trial:** It is a type of clinical trial which is used to show bio-equivalence, thus, to show that the new therapeutic agent is equivalent to the conventional one.
- c. **Non-inferiority trial:** It is a type of clinical trial which is used to show that the new therapeutic agent is not worse than the conventional one.

4. Ethical consideration

Since a clinical trial is an experiment carried out on humans, a lot of ethical issues come into the picture.^[30] The following are the most important ethical considerations:

- a. **Prior knowledge:** For a clinical trial to meet ethical standards, there should be enough evidence about its efficacy, in other words, enough evidence
-

should exist showing that the therapeutic agent is beneficial. On the other hand, the agent under consideration should not be proven to be superior. This stems from the fact that the investigator should not expose patients to a non-efficacious agent, nor should he/she withhold a superior drug from patients.

- b. **Informed consent:** A clear and comprehensive “informed consent” form should be signed by participants for them to be included in a clinical trial. Such an informed consent should be reviewed and approved by the IRB.
- c. **Stopping rules:** Clear criteria and conditions should be outlined at the beginning of the clinical trial that specifies the conditions under which a clinical trial should be stopped, either for evidence of superiority of the therapeutic agent under study or its inferiority. Such a decision should be made by an independent body, which is usually assigned at the beginning of the study (Data Monitoring and Safety Committee).

5. Randomization

Randomization is the most important aspect of a clinical trial. It is used to allocate the therapeutic agent under study to patients, which is used independently of the researchers’ preference. In technical terms, it is done where each member of the sample included in the study has the same probability of receiving either the therapeutic agent under consideration or the conventional one. Although some clinical trials do not follow randomization in the assignment of the therapeutic agent under consideration to patients, randomized clinical trials are superior to those of non-randomized ones.

Randomization is carried out to control for confounding variables by ensuring equal distribution of confounders. It could be done by using computer programs. Randomization does not always generate groups that are balanced in terms of the confounding variables. Verification of the success of the randomization could be done by comparing the groups in terms of their baseline characteristics (which is usually the first table in any clinical trial publication).

There are different types of randomization, mainly simple randomization, block randomization, stratified randomization, etc.

6. Blinding

Blinding in clinical trials is a concept which means that some individuals are masked for the knowledge of the allocated agent to the patients.^[49] Mainly there are three levels of blinding:

- a. **Single blinding:** This means that the patients are blinded to which drug they are allocated. The use of placebo is specifically important when blinding is carried out in a clinical trial. This is done to avoid bias due to psychological benefit of the knowledge of the allocated agent by the patient.
- b. **Double blinding:** This means that both patients and investigators are blinded to which drug the patients are allocated. The investigators' knowledge about the allocated agent might bias their judgment in terms of outcome assessment, especially when the outcome is subjective.
- c. **Triple blinding:** This imposes blinding to the patients, investigators, and statisticians, who in turn might introduce bias to the results of the study based on their knowledge of the allocated agent. Blinding the statistician is the least important in the conduct of a clinical trial.

7. Analyses

The analyses of a clinical trial is similar to that in a cohort study, where relative risk could be calculated. There are two types of analyses that are usually carried out to address different objectives of the study. To illustrate this, assume a hypothetical example of an intervention to compare two medications prescribed to relief headache. In this example, 100 patients are randomized to drug A, and another 100 randomized to drug B. At the end of the study, 5 patients who were randomized to drug A did not comply to the medication (i.e. did not take the medication) and another 2 patients switched to the other medication (they took drug B). On the other hand, among those randomized to drug B, 3 patients did not take the medications and 3 patients switched to the other drug (they took drug A). This is a common finding in clinical trials, and accordingly researchers could carry out two types of analyses:[50]

a. Effectiveness analyses

Effectiveness analyses are carried out to assess the effect of treatment on the outcome in the “real world”, which include poor compliers and those who switch to other medications. The following is a two by two table to illustrate the analyses. This type of analyses is called “intent to treat” analyses.

		Relief of headache		Total
		Yes	No	
Intervention	Drug A			100
	Drug B			100
	Total			200

b. Efficacy analyses

Efficacy analyses are carried out to assess the efficacy of treatment on the outcome among those who actually took the medication. Accordingly, the investigator adjusts the analyses to take into consideration those who switched and those who did not comply. Following is the same example where the number in the drug A group and drug B group is modified based on the information provided about compliance and switching.

		Relief of headache		Total
		Yes	No	
Intervention	Drug A			96
	Drug B			96
	Total			192

Example: As an example, a clinical trial was carried out to examine the effect of intensive insulin therapy on mortality in medical surgical intensive care unit patients, in King Abdul Aziz Medical City.^[51] Patients were randomly allocated to receive either intensive insulin therapy or conventional insulin therapy. Following are the results:

		Treatment group		Total
		Intensive Insulin	Conventional insulin	
Hospital Mortality	Yes	72	83	155
	No	194	184	378
	Total	266	267	523

The incidence of ICU mortality in the intervention group was 27.1% compared to 32.2% among the conventional group, which gives a RR of 0.84. The conclusion from this analysis would be that those on intensive insulin therapy were less likely to die in the hospital as compared to those on conventional insulin therapy.

8. Strengths and limitations

Among the strengths of clinical trials are the following:

- Can directly measure disease incidence, since there is a follow-up inherent in the study design.

- b. At the time of sample selection, the outcome has not occurred yet, and thus temporal relationship can be established between exposure and outcome, thus building evidence for causality.
- c. Appropriate for assessing the association between one exposure and different outcomes.
- d. Reduced biases (selection, information and confounding).

As for the limitations of clinical trials, they include the following:

- a. May be impractical or unethical, based on the available knowledge about the intervention.
- b. Expensive and take long periods of time.
- c. Compliance, as some patients might not comply properly to the intervention under study.
- d. Loss to follow-up, especially when the duration of follow-up is very long, which might introduce selection bias.
- e. Generalizability of the study results might be limited, especially when researchers use restrictive inclusion/exclusion criteria.

Section 16: Manuscript writing

1. Overview

Writing and publishing the results of a study is the most important step a researcher has to accomplish. Unless the results are published, the value of the whole research project will remain minimal. This step is important for the researcher as it provides documentation of the work done and the findings found. For the medical community, it is crucial that research findings are published as it provides public knowledge about the research project which may enhance medical knowledge.

A scientific manuscript is an original text of an author's work which is submitted to a publisher. The structure of a manuscript follows strict criteria that abides by the scientific method. It includes different sections that provides the reader with the information needed to understand what was done and what results found. The structure depends on the journal to which the article is being submitted. Different journals have different guidelines, which should be reviewed and followed when preparing the manuscript to be submitted.^[52,53]

2. Manuscript preparation

Preparing a manuscript is a time consuming and critical step in the research process. The scientific manuscript is made up of the following sections (a template of a manuscript is provided in Appendix II):

a. Title Page

The title page should have the following information:

- **Title of the manuscript:** It should be simple, concise and informative. Authors should include the keywords in the title that will make electronic retrieval of the article easier, as well as understanding the main features of the manuscript easily screened.
- **Authors' names and institutional affiliations:** This is to indicate information about the researchers who carried out the study, as well as their affiliation.
- **Contact information for corresponding authors:** The name, mailing address, telephone and e-mail address of the author responsible for correspondence about the manuscript.
- **Source(s) of funding:** It includes the sources from which funding was provided to carry out the study, if any.

- **Conflict-of-Interest:** To prevent potential conflicts of interest from being overlooked or misplaced, this information needs to be part of the manuscript.

b. Abstract and key words

An abstract summarizes the major aspects of the entire research project. The length of the abstract should be kept to about 200-300 words maximum (a typical standard length for journals.) Although there are various abstract forms-which depend on each journal; generally, an abstract should include the following parts:

- Study aims or objectives
- Methods, such as study design, inclusion/exclusion criteria, and data collected
- Results found which highlights the main findings of the study
- Conclusion which summarizes the main conclusion of the study

The abstract helps readers decide whether they want to read the rest of the paper. It may be the only part they can obtain via electronic literature searches; therefore, enough information need be included to make it comprehensive by itself (without referring to the full paper). Although it is the first section of the paper, the abstract is usually written last since it summarizes the paper. Once the abstract is completed, it has to be checked in reference to the full paper to make sure that the information in it is in agreement with what is written in the paper.^[54]

c. Introduction

The introduction provides a context or background for the study (that is the nature of the problem and its significance). It contains the study background (revealing the importance of the problem under study, supported by data from the available literature), the study objective and questions that were answered by the study.

The structure of the Introduction can be thought of as an inverted triangle - the broadest part at the top representing the most general information and focusing down to the specific problem you studied. Organize the information to present the more general aspects of the topic early in the Introduction, then narrow toward the more specific information that provides context, finally arriving at your statement of purpose and rational. It should be able to ‘drive’ the readers to understand and agree that the idea of the study performed is necessary and reasonable.

d. Methods

The Methods section should include only information that was available at the time the plan or protocol for the study was being written. In this section, the following should be explained:

- The study design
- Time and place of the study
- Inclusion and exclusion criteria
- Subject selection (sampling)
- Information collected throughout the study
- Informed-consent (if applicable)
- Statistical analyses used to reach the results

It is important to note that both the introduction and the methods sections have been prepared earlier for the proposal, nevertheless, the researcher has to update the literature review, as well as to modify the methods section according to what methods actually were used for carrying out the study.

e. Results

The study results must be presented in a logical sequence in the text. Overview of the study sample should be provided, followed by results pertaining to the objective of the study. It serves as a summary of the tables, but it should not replicate them, rather it should highlight the most important findings.

f. Discussion

This section emphasizes the new and important findings in light of what is found in the literature. Results should not be repeated in this section, rather implications of the results should be addressed. Specifically, this section should include possible mechanisms or explanations for the findings, comparison of the results with other relevant studies, indication to the strengths and limitations of the study, as well as implications for future research and for clinical practice.

g. Conclusion

This section summarizes the potential significance of the paper and final implications on similar research or clinical practice concluded from the study.

h. References

It is a list of published scientific material (papers, books, etc.) that were used in the manuscript.

i. Tables and graphs

These are the summary of the results obtained from the study under consideration, and should be clearly presented with titles and footnotes.

After the manuscript has been written according to the guideline provided by the journal, it could be read few times to make sure it flows well, and is free from typos, incomplete sentences, and grammatical errors. At that time, it could be given to some colleagues for review and final comments. Once the researcher is comfortable with the manuscript as it stands, submission to a scientific journal is next.

3. Manuscript submission

Choosing the appropriate journal for publishing research papers is a tedious process. If the research paper is submitted to a journal whose target audience and scope is not relevant, there is little chance of it getting published. It is essential that the researcher chooses the most appropriate journal to submit the research paper to.

Following are some of the factors to be considered during journal selection:

- a. If the journal is peer-reviewed or not
- b. Scope of the journal
- c. Does it publish original research or review papers
- d. Impact factor of the journal
- e. Average duration of review process
- f. Does it have online submission system?
- g. Cost of publication (page charges etc.)

A cover letter should be prepared and submitted along with the manuscript. It should include the following:

- a. A statement that the manuscript is original work of the author and has not been published before in another journal or other media

- b. Statements about any sponsor that potentially may cause any conflict of interest
- c. A statement that the manuscript has been read and approved by all the authors
- d. The name, address, and telephone number of the corresponding author, who is responsible for communicating with the journal

The next step is to submit the manuscript to the selected journal, following the step-by-step submission process, which varies from journal to journal. Most of the scientific journals nowadays accept electronic submission of manuscripts. Authors should consult the journal's Instructions for Authors for detailed description of the submission process. Many journals provide a pre-submission checklist to help the author ensure that all the components of the submission have been included.^[55]

4. Submission response

The time between submission and receiving the response from the journal could be as short as a day and as long as few months. There are three possible responses to the submission of the paper:^[56]

a. Acceptance as is

Acceptance as is: indicates that the editorial office, accepts the manuscript without any further modifications, which is rarely the case. Accordingly, the next step is to finalize the process by signing the “copyright transfer agreement” that is specific to each journal.

b. Conditional acceptance

Conditional acceptance: is an indication that the manuscript has good chances of being accepted for publication. Usually, lists of comments are sent to the author. Some of these comments could be minor (such as typos) or major (such as further analyses). The next step for the authors is to address each and every comment received with as much details as possible, by either abiding by the suggestions or by not taking them into consideration. In either case, the authors should justify their action. At this point, the modified draft of the manuscript, along with a point by point letter addressing the comments should be sent to the journal, through the online system. It might take few rounds of revisions before the manuscript is accepted.

c. Rejection

Rejection which indicates that the manuscript is rejected for publication in that journal, for reasons that might be based on the topic, the quality, the conclusion, or just that it does not fall within the scope of the journal. At this point, the authors should look for another journal to which they should submit their manuscript.

Although the publication process might be time and effort consuming, it is very important and it gets easier with time, as the researcher publishes more papers. One last thing to keep in mind is that most research projects are publishable in scientific journals, but the type and impact of the journal in which it will be published might differ.

Section 17: Evidence based medicine (EBM)

1. What is EBM

EBM is the integration of the current best research evidence with clinical expertise and patient's values, preferences and circumstances.^[57] By best research evidence we mean most updated and relevant clinical research. By clinical expertise we mean the clinical skills and past experience to identify each patient's unique health state and diagnosis and the specific risks and benefits of potential interventions. By patient values we mean the specific preferences, concerns, personal values, and expectations of each patient.^[58]

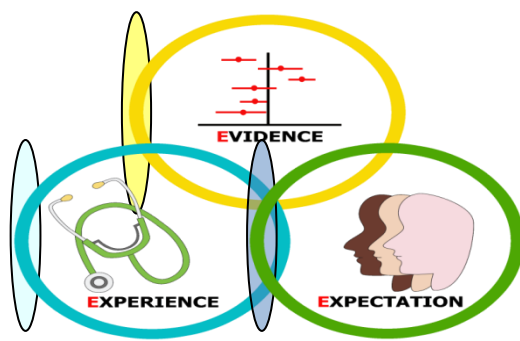


Figure: Three E's - EBM components

2. Why we need EBM

In the past, physicians relied on their own experience or that of other health care workers to take decisions regarding patients' treatment. Currently this approach is inadequate and poor as health care workers rapidly find themselves unable to cope with the influx of a huge variety of new information, from the irrelevant to the very important. Therefore, the need for the evidence-based decision came from:^[59]

- a. Our daily need for valid information about diagnosis, prognosis, therapy and prevention.
- b. The inadequacy of traditional sources for this information because they might be out-of-date (textbooks), potentially wrong (colleagues), or too overwhelming in their volume (medical journals).

- c. The disparity between our diagnostic skills/clinical judgment, which increases with experience, and our up-to-date knowledge which declines with time.
- d. Our inability to afford more than a few minutes per patient for finding and assimilating this evidence or to set aside more than few hours per week for general reading and study.

3. EBM principles

EBM involves two fundamental principles:[59]

- a. Evidence alone is never the sole basis for decisions: Benefits and risks, costs and alternative strategies, as well as the patients' values are all factors that must be taken into consideration alongside with evidence when taking the decision.
- b. EBM has a hierarchy of strength of evidence for treatment decisions: The hierarchy of evidence is a spectrum of potential sources beginning with those most likely to provide the evidence to those with the least likely. Thus, the physician must begin with the highest available evidence from the hierarchy.

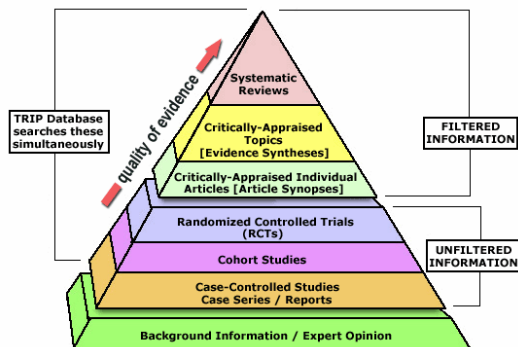


Figure: Hierarchy of Evidence

4. EBM steps

Practicing EBM is primarily based on five well defined steps (5 A's), which can be broadly categorized as the five A's:[58,60,61]

- a. **Asking Focused Questions:** The question should be directly relevant to the problem at hand and should be phrased to facilitate searching for a precise

answer. To achieve these aims, the question must be divided into four components, which are called 'PICO'.

- **P:** The Patient and/or Problem being addressed. Ex: Young children with otitis media
- **I:** The Intervention or exposure being considered. Ex: Treatment with Amoxicillin
- **C:** The Comparison intervention when relevant. Ex: Compared with placebo
- **O:** The clinical Outcomes of interest. Ex: Results in faster improvement?

The overall question becomes: Does Amoxicillin lead to faster improvement in otitis media among young children compared to placebo?

- b. **Acquiring the Evidence:** In this step we make a systematic retrieval of the best evidence available. Choosing the best resource to search is an important decision.

Following is a list of some valuable resources for practicing EBM:[62]

- **Summaries of the primary evidence:** ACP Journal Club, Clinical Evidence
 - **Databases:** PubMed, Cochrane Library
 - **Electronic textbooks and libraries:** AccessMedicine, ACPMedicine,
 - **Meta-Search Engines:** SUMSearch, TRIP Answers
- c. **Appraisal of evidence:** It is the process of assessing and evaluating the evidence for its internal validity, its clinical relevance, and applicability. Appraisal of evidence depends on the following 4 pillars -RVRA:
- **Relevance:** It focuses on the relevance of the literature to the question asked
 - **Validity:** Are the results of the study valid?
 - **Results:** What are the overall results? How precise are they?
 - **Applicability:** Are the results Applicable in and useful for my patients
- d. **Application of the best evidence in practice:** Based on the findings of the above, a clinical decision is to be made.
- e. **Assessing and evaluating the performance:** Evaluating our effectiveness and efficiency in executing Steps 1 - 4 and seeking ways to improve them both for next time.
-



Figure: EBM steps, the 5 A's

5. Limitations of EBM

Although EBM is regarded as the best standard of conventional clinical practice there are a number of limitations for its use:^[58,60]

- a. Lack of good evidence for many clinical questions
- b. Results may not be relevant for all treatment situations
- c. Lag in time between when the research studies is conducted, when its results are published when these are properly applied

6. Linking research to EBM

Finally, to be able to practice medicine based on the best evidence, the health care practitioner has to be equipped with both clinical and research knowledge. More specifically, research knowledge includes understanding of research process, ethical consideration, study design, statistical analyses, and results interpretation.

References

1. Marczyk GR. *Essentials of Research Design and Methodology*. John Wiley & Sons, Inc; 2005.
 2. Shuttleworth M. *Definition of Research*. 2008 [cited 14 August, 2011; Available from: <http://explorable.com/definition-of-research>.
 3. Baylor U. *Conducting research for building strong interpersonal relationships*. Dr. Keith Sanford; Available from: <http://www.baylor.edu/>.
 4. Vicki L, Plano Clark JWC. *Understanding Research: A Consumer's Guide*. Prentice Hall; 2009.
 5. Haynes Brian R. *Forming research questions*. J Clin Epidemiol, 2006. 59(9): 881-6.
 6. Aslam S, Emmanuel P. *Formulating a researchable question: A critical step for facilitating good clinical research*. Indian J Sex Transm Dis, 2010. 31(1): 47-50.
 7. Stone P. *Deciding upon and refining a research question*. Palliat Med, 2002. 16(3): 265-7.
 8. Farrugia P, et al. *Practical tips for surgical research: Research questions, hypotheses and objectives*. Can J Surg, 2010. 53(4): 278-81.
 9. Lipowski EE. *Developing great research questions*. Am J Health Syst Pharm, 2008. 65(17): 1667-70.
 10. LoBiondo-Wood G. *Nursing Research: Methods and Critical Appraisal for Evidence-Based Practice*. 8th ed. Mosby; 2014
 11. Fink A. *Conducting Research Literature Reviews*. Sage Publications, Inc.; 2005
 12. Levy Y, Ellis TJ. *A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research*. Informing Science Journal, 2006. 9: 181-211.
 13. Steward B. *Writing a literature review*. Br J Occup Ther, 2004. 67: 495-500.
 14. DuBois D. *Practical Strategies and Guidelines for Conducting Literature Reviews in Research*. 2012; Available from: http://www.ihrp.uic.edu/files/DuBois_litreviews_june2012.pdf.
 15. Ebrahim NA. *Approach to Conduct an Effective Literature Review*. 2012; Available from: <http://works.bepress.com/aleebrahim/65>.
 16. University of Connecticut. *Literature Review: The What, Why and How-to Guide*. 2013; Available from: <http://classguides.lib.uconn.edu/content.php?pid=239974&sid=1981395>.
 17. Featherstone R. *How to conduct a literature review*. 2011; Available from: <http://www.slideshare.net/featherr/how-to-conduct-a-literature-review>.
-

18. Saunders M, Lewis P, Thornhill A. *Research Methods for Business Students*. Prentice Hal; 2009.
19. Northeastern University. *Using EndNote*. 2010 Last updated March 2010; Available from: <http://www.help.neu.edu/>.
20. Kentucky University. *EndNote X7 (17) Tutorial*. 2012 Updated 07-05-2013; Available from: <http://libraries.uky.edu/>.
21. Cornell University. *EndNote Tutorial*. 2013; Available from: <http://mannlib.cornell.edu/>.
22. Thomas Reuters. *ENDNOTE*. 2013 [cited 2013; Available from: <http://www.endnote.com/endemo.asp>.
23. University of Illinois at Urbana-Champaign. *Writing a Research Proposal*. 2012; Available from: <http://www.library.illinois.edu/learn/research/proposal.html>.
24. Al-Riyami A. *How to prepare a Research Proposal*. Oman Med J, 2008. 23(2): 66-9.
25. Hollins CJ Martin, Fleming V. *A 15-step model for writing a research proposal* British Journal of Midwifery, 2010. 18(12): 791-8.
26. Skills, T.U.o.M.A. *Writing a Research Proposal: A guide for Science and Engineering students* 2013; Available from: <http://services.unimelb.edu.au/academicskills>.
27. Institutional Review Board, American University of Beirut. *Guidelines for Submitting Research Proposals*. 2005; Available from: <http://www.aub.edu.lb/irb/Pages/ResourceDocuments.aspx>.
28. Vijayananthan A, Nawawi O. *The importance of Good Clinical Practice guidelines and its role in clinical trials*. Biomed Imaging Interv J, 2008. 4(1): e5.
29. Skolnick BE. *Ethical and institutional review board issues*. Adv Neurol, 1998. 76: 253-62.
30. Bernard LO. *Ethical Issues in Clinical Research*. Lippincott Williams & Wilkins; 2010.
31. Kapp MB. *Ethical and legal issues in research involving human subjects: do you want a piece of me?* J Clin Pathol, 2006. 59(4): 335-9.
32. Kim WO. *Institutional review board (IRB) and ethical issues in clinical research*. Korean J Anesthesiol, 2012. 62(1): 3-12.
33. Wolf LE, Walden JF, Lo B. *Human subjects issues and IRB review in practice-based research*. Ann Fam Med, 2005. 3 Suppl 1: S30-7.
34. Cook A, Netuveli ACG, Sheikh A. *Basic Skills in Statistics: A Guide for Healthcare Professionals*. London, Class Publishing; 2004.

35. IBM Corp. *IBM SPSS Statistics for Windows Version 21.0*. Released 2012; Available from: <http://www-01.ibm.com/support/docview.wss?uid=swg21476197>.
36. Willis J. *Data Analysis and Presentation Skills: An Introduction for the Life and Medical Sciences*. London: John Wiley & Sons, Ltd; 2004.
37. Campbell MJ. *Medical Statistics: a commonsense approach*. 3rd ed. Chichester: John Wiley & Sons; 1999.
38. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall; 1991.
39. Bowers D. *Statistics from scratch: an introduction for health care professionals*. Chichester, John Wiley & Sons; 1996.
40. Chow SC. *Sample size calculations of clinical research*. Marcel Dekker, Inc.; 2003.
41. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Lippincott Williams & Wilkins; 2008.
42. Alqurashi KA, Aljabri KS, Bokhari SA. *Prevalence of diabetes mellitus in a Saudi community*. Ann Saudi Med, 2011. 31(1): 19-23.
43. Khashoggi TY. *Epidemiology of intrauterine fetal deaths in Saudi Arabia: kkuh experience*. Biomedical Research 2005. 16(1): 59-64.
44. Saeed AA, et al. *Prevalence, Awareness, Treatment, and Control of Hypertension among Saudi Adult Population: A National Survey*. Int J Hypertens, Epub 2011 Sep 6.
45. Al-Kadri HM, et al. *Maternal and neonatal risk factors for early-onset group B streptococcal disease: a case control study*. Int J Womens Health, 2013. 5: 729-35.
46. Ury HK. *Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data*. Biometrics, 1975. 31(3): 643-9.
47. El-Gilany AH, Hammad S. *Body mass index and obstetric outcomes in pregnant in Saudi Arabia: a prospective cohort study*. Ann Saudi Med, 2010. 30(5): 376-80.
48. Al-Khalifah R, et al. *Neonatal short-term outcomes of gestational diabetes mellitus in saudi mothers: a retrospective cohort study*. J Clin Neonatol, 2012. 1(1): 29-33.
49. Machin D, Day S, Green S. *Textbook of clinical trials*. John Wiley & Sons, Ltd; 2004.
50. Geller N. *Advances in clinical trial biostatistics*. Marcel Dekker, Inc; 2004.
51. Arabi YM, et al. *Intensive versus conventional insulin therapy: a randomized controlled trial in medical and surgical critically ill patients*. Crit Care Med, 2008. 36(12): 3190-7.

52. Wenzel V, Dunser MW, Lindner KH. *How do I write an original article? An introduction for beginners*. Anaesthesist, 2007. 56(8): 828-36.
53. Jenicek M. *How to read, understand, and write 'Discussion' sections in medical articles. An exercise in critical thinking*. Med Sci Monit, 2006. 12(6): SR28-36.
54. Alexandrov AV. *How to write a research paper*. Cerebrovasc Dis, 2004. 18(2): 135-8.
55. *Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication*. J Pharmacol Pharmacother, 2010. 1(1): 42-58.
56. Hoppin FG, Jr. *How I review an original scientific article*. Am J Respir Crit Care Med, 2002. 166(8): 1019-23.
57. Timmermans S, Mauck A. *The promises and pitfalls of evidence-based medicine*. Health Aff (Millwood), 2005. 24(1): 18-28.
58. Centre for Evidence-based Medicine-University of Oxford. *What is EBM?* 2011; Available from: <http://www.cebm.net/?o=1914>.
59. Guyatt GH. *Users' Guides to the Medical Literature*. American Medical Association Press; 2001.
60. Wikipedia-The Free Encyclopedia. *Evidence-based medicine*. 2013; Available from: http://en.wikipedia.org/wiki/Evidence-based_medicine.
61. The Lamar Soutter Library-University of Massachusetts Medical School. *Components of EBM*. 2003; Available from: <http://library.umassmed.edu/EBM/components.html>.
62. The Duke University School of Medicine. *Introduction to Evidence-Based Practice*. 2013; Available from: <http://guides.mcclibrary.duke.edu/content.php?pid=431451&sid=3530449>.

Appendices

I – Research proposal

1. Title of Proposal: (Instructions: Fill all appropriate boxes apply to your project) 			
2. Type of Project: (please check all applicable options) ☐ Chart Review ☐ Diagnostic ☐ Qualitative Research ☐ Quantitative Research ☐ Human ☐ Laboratory ☐ Therapeutic ☐ Basic Science ☐ Other _____			
3. Starting Date: 	4. Duration: 	5. Total Fund Requested (US dollars): 	
6. Principal Investigator (PI): Name: _____ Affiliation & Address: _____ Title/Position: _____ Tel. No.: _____			
7. Co-Investigator:	Title/position:	Department:	Signature:
7.1 Principal Investigator's Assurance: The undersigned agrees to accept responsibility for the scientific and technical conduct of the proposed research and submission of progress reports if this application is approved. Advisor's Name & Signature: _____ Date: _____			

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

[illegible]

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

11.1 Study Area/Setting: (Instructions: Describe the area or setting where the study will be conducted. This description will cover the details relevant to the study topic)

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.[illegible]

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

[illegible]

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

14. Budget: (Please use the attached documents for the price list of equipment used, if applicable)			
Budget Breakdown	Unit Cost (US dollars)	Total Cost (US dollars)	Remarks
Personnel			
<i>Total</i>			
Supplies and Equipment			
<i>Total</i>			
Patients Cost			
<i>Total</i>			
Others (please, specify and justify briefly)			
<i>Total</i>			
GRAND TOTAL			

15. Time frame for the study: (Instructions: Please use this form as a template for timeline of project)												
Tasks	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Progress report												

II – Manuscript template

Title Page

Authors

Affiliations

Corresponding author

Conflict of interest:

Keywords:

Abstract

Background:

Objectives:

Methods:

Results:

Conclusion:

Introduction

1. Background information about the topic
2. What is known
3. Gap in knowledge

Conclude the Introduction section with a statement about the objective(s) of the study

Methods

Study design:

Setting:

Study population:

1. Inclusion criteria
2. Exclusion criteria

Sampling:

1. Selection of patients

Data collection:

1. Method of data collection
2. Data collected
 - a. Information
 - i. what
 - ii. when
 - b. Questionnaire used

Endpoints:

Ethical considerations:

1. IRB approval
2. Consent forms
3. Confidentiality

Statistical analyses:

1. Program used
2. Data cleaning
3. Data management
4. Data analyses
 - a. Univariate
 - b. Bivariate
 - c. Multivariate

Results

Overall description of participants (demographic characteristics and socioeconomic profile)

Summary of findings

Multivariate analyses

Discussion

Summary of study and findings

1. Overall methodology and objective
2. Overall finding

Comparison with literature (comparing results to similar studies)

Strengths and limitations

Future recommendation

Conclusion (final statement about the overall study)

Tables and Graphs

Table 1:

Table 2:

Table 3:

Table 4:

References



SCFHS©2014

The Saudi Commission for Health Specialties

www.scfhs.org

ISBN: 978-9960-9832-9-5