NATIONAL OPEN UNIVERSITY OF NIGERIA

FACULTY OF HEALTH SCIENCES

DEPARTMENT OF PUBLIC HEALTH

COURSE CODE: PHS805

COURSE TITLE: RESEARCH METHODS IN PUBLIC HEALTH

COURSE GUIDE

PHS 805: RESEARCH METHODS IN PUBLIC HEALTH

Course Developer/Writers: Professor Babatunji A. Omotara
Department of Community Medicine
University of Maiduguri

Dr. Adewale L. Oyeyemi
Department of Physiotherapy
University of Maiduguri

Course Editor: **Dr. Matthew O. Olayiwola**
Department of Public Health
Federal University of Agriculture, Abeokuta

Course Coordinator: **Dr. Florence N. Uchendu**
Department of Public Health
National Open University of Nigeria

Programme Leader **Prof. Grace Okoli**
Faculty of Health Sciences
National Open University of Nigeria

National University of Nigeria

Headquarters

14/16 Ahmadu Bello Way

Victoria Island, Lagos

Abuja Office

5 Dar es Salaam Street

Off Aminu Kano Crescent

Wuse II, Abuja

Email: centralinfo@nou.edu.ng

URL: WWW.nou.edu.ng

Published by

National Open University of Nigeria

Printed 2020

ISBN:

CONTENTS **PAGE**

Introduction	6
What you will Learn in this Course	6
Course Aim	6
Course Objectives	7
Working through this Course	7
Course Materials	7
Course Modules and Units	8
Text Books and References	9
Assessment	12
Tutor-Marked Assignments	13
Final Examination and Grading	13
Summary	13

INTRODUCTION

This course, PHS805: Research Methods in Public Health is a three-credit unit course that provides the necessary information concerning the various research methods employed in the field of public health to unravel the mystery surrounding disease and injury causation with the view of preventing and or eradicating them. Research in public health has evolved through many stages from the pre-industrial era to present day advanced technology. With the complexity associated with technological advancement, the dynamics of disease causation also became more complex, requiring different approach to identifying and ameliorating its effects on human health.

The aetiology of disease causation is very important in public health as it helps to prevent the occurrence of disease in man. As such, this course is to guide you to understand issues involved in research methods in public Health.

WHAT YOU WILL LEARN IN THIS COURSE

This course will familiarise you with a good understanding of the various methods used in the field of Public health by providing necessary information on the various research methods and the formulation of the appropriate study design. It will also introduce students to the application of appropriate statistical analyses used for various public health research designs.

COURSE AIM

The aim of the course is to expose students to the fundamentals of public health research and acquaint them with the various methods employed in the field of public health.

COURSE OBJECTIVES

At the completion of this course, you should be able to:

1. define and explain public health research
2. describe the various types of research and study design in public health
3. describe the rudiments of research methodologies in the field of public health.
4. describe the concept of sampling and the various sampling techniques
5. describe the concept of psychometric properties (validity and reliability) in public health measurements
6. develop appropriate tools (e.g., questionnaire) for the research process
7. identify problems associated with data collection, process of data collection
8. identify statistical analysis methods and their interpretation in public health
9. describe the process of formulation of research topics
10. identify and describe the basic components of research proposal
11. understand how to write the result and discussion of your research and how to prepare and submit a manuscript in a journal

WORKING THROUGH THIS COURSE

This course has been developed to enhance the understanding of all students including those who are new to public health. The course has been written to allow for both self and group study. Students are encouraged to spend good time to study the course and should not hesitate to ask questions for better understanding of the course contents from the course team as needed.

COURSE MATERIALS

The course materials consist of a course guide and the study units.

COURSE MODULE AND UNITS

This two units course comprises of three modules broken down into 9 units. They are listed as below:

Module 1	**Introduction to Public Health Research**
Unit 1	Definition and Concept of Public Health Research
Unit 2	Concepts of Sampling in Public Health Research
Unit 3	Reliability and Validity of Measurements in Public Health Research
Module 2	**Research Methods and Data Analyses in Public Health**
Unit 1	Study Designs in Public Health Research
Unit 2	Quantitative vs Qualitative Research Methods
Unit 3	Data Analysis Methods in Public Health Research
Module 3	**Topic Selection and Components of Research Proposal in Public Health**
Unit 1	Construction and Formulation of Research Topics
Unit 2	Components of a Research Proposal
Unit 3	Beyond Research Proposal

Module 1

In unit 1 you will be taken through the definition and concept of public health research. The unit will also teach you the types of variables and measurement scales in public health research. In unit 2 you will be taken through the definitions and concepts of sampling in public health research. You will also be introduced to the various types of sampling techniques in public health research. In unit 3, you will be introduced to the concepts of psychometric properties of outcome measures in public health. This unit will also introduce you to questionnaire development and design

Module 2

In Unit 1, you will be taken through the concepts of study design in public health, including the distinction between descriptive and analytical research. In unit 2, you will learn the distinctions between quantitative and qualitative research methods. In unit 3, you will be introduced to the various statistical methods used for data analyses in public health research.

Module 3

Unit 1, you will be taken through the various steps in construction and formulation of research topics in public health. In unit 2, you will be introduced to the various components and how to write an effective research proposal for a Master of Public Health Degree. Finally, in unit 3, you will learn how go beyond the research proposal by understanding how to write the results and discussion of your research findings and also how to prepare and submit manuscripts for publication in journals.

TEXT BOOKS AND REFERENCES

The lists of textbooks, articles and website addresses that can be consulted for further reading for the modules in this course are indicated below:

Adeloye, D., Thompson, J.Y., Akanbi, M.A., Azuh, D., Samuel, V., Omoregbe, N. & Ayo, C.K. (2016). The burden of road traffic crashes, injuries and deaths in Africa: a systematic review and meta-analysis. *Bulletin of the World Health Organization,* 94:510–521A

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

Al-Riyami, A. (2008). How to prepare a research proposal. *Oman Medical Journal,* 23(2):66-69

Bamgboye E.A. (2004). *A companion of medical Statistics*. Published by FOLBAM Health Research and Data Management Center, Ibadan. ISSN 978-056-661-9.

Belcher, W.L. (2009). *Writing Your Journal Article in 12 Weeks: a guide to academic publishing success*. Thousand Oaks, CA: SAGE Publications

Bowers, D. (2008). *Medical Statistics from scratch: an introduction for health professionals*. London: John Wiley & Sons Ltd.

Chin, R., & Lee, B.Y. (2008). *Principles and Practice of Clinical Trial Medicine*. Elsevier

De Negri B. & Thomas, E. (2003). "*Making sense of Focus Group Findings: A systematic Participatory Analysis Approach"*. Washington, DC: Academy for Education Development.

DePoy, E., & Gitlin, L.N. (2016). *Introduction to Research: Understanding and Applying Multiple Strategies*. (5th ed). Mosby.

Hoffman, J.I.E. (2015). *Biostatistics for Medical and Biomedical Practitioners*. Elsevier.

International Physical Activity and the Environment Network (IPEN, 2019). Publication Guidelines for the IPEN Studies. Available at: https://www.ipenproject.org/

Kallestinova, E.D. (2011). How to Write Your First Research Paper. *Yale Journal of Biology and Medicine*, 84:181-190. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3178846/

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

McCombes, S. (2019). How to write a discussion section. Available at https://www.scribbr.com/dissertation/discussion/

McDowell, I., & Newell, C. A. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

Mukaka, M.M. (2012). "A guide to appropriate use of Correlation coefficient in medical research." *Malawi Medical Journal*, 24(3): 69–71

Obasola, O.I. & Mabawonku, I.M. (2018). Mothers' perception of maternal and child health information disseminated via different modes of ICT in Nigeria. *Health Information and Libraries Journal*, 35: 309–318

Omotara, B.A., Yahya, S.J., Wudiri, Z., Amodu, M.O., Bimba, J.S. and Unyime, J. (2015). Assessment of the Determinants of Healthy Ageing among the Rural Elderly of North-Eastern Nigeria. *Health*, **7**, 754-764.

Oyeyemi, A.L., Oyeyemi, A.Y., Jidda, Z.A., and Babagana, F. (2013). Prevalence of physical activity among adults in a metropolitan Nigerian city: A cross-sectional study. *Journal of Epidemiology*, 23(3):169-177.

Pajares, F. (2207). Elements of proposal 2007. From http://www.des.emory.edu/mfp/proposal.html. Accessed January 18, 2020.

Portney, L.G., & Watkins, M.P. (2000). *Foundation of Clinical Research: Application to practices*. (2nd ed). Upper Saddle River, New Jersey: Prentice Hall.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Smith, R.V. (1984). *Graduate Research: A guide for students in the sciences*. Michigan: ISI Press

Streiner, D.L., Norman G.R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use*. (5th ed.). Oxford: Oxford University Press.

Yeomens, S.G. (2000). The clinical application of outcomes assessment. Stanford Appleton and Lange.

Wang, G.T., & Park, J. (2016). Student Research and Report Writing: From topic selection to the complete paper. John Wiley & Sons ltd; Malden, USA

Wong P. How to write research proposal. International network on Personal meaning. Available at www.meaning.ca/archives. Accessed January 18, 2020.

www.fhi.org/nr/rdonlyers/..../datacollectorguideenrh.pdf

ASSESSMENT

The assessment for this course will be based on the cumulation of the tutor marked assignments and final examination.

TUTOR-MARKED ASSIGNMNETS (TMAs)

The Tutor-Marked Assignment (TMAs) is the continuous assessment of the course and its account for 30% of the total score.

FINAL EXAMINATION AND GRADING

The final examination will complete the assessment for the course. It will constitute 70 % of the marks/grade for the whole course.

SUMMARY

This course was designed to provide students with the knowledge of public health research and acquaint them with the various research methods and statistical analyses used in the field of public health. We wish you success in this course.

MAIN COURSE

CONTENTS **PAGE**

Module 1	**Introduction to Public Health Research**	**15**
Unit 1	Definition and Concept in Public Health Research	15
Unit 2	Concepts of Sampling in Public Health Research	25
Unit 3	Reliability and Validity of Measurements in Public Health Research	36
Module 2	**Research Methods and Data Analyses in Public Health**	**51**
Unit 1	Study Designs in Public Health Research	51
Unit 2	Quantitative vs Qualitative Research Methods	65
Unit 3	Data Analysis Methods in Public Health Research	76
Module 3	**Topic Selection and Components of Research Proposal in Public Health**	**92**
Unit 1	Construction and Formulation of Research Topics	92
Unit 2	Components of a Research Proposal	97
Unit 3	Beyond Research Proposal	110
Appendix 1	**Work Plan for Research Proposal**	**126**
Appendix 2	**Template for Research Proposal**	**127**

MODULE 1 INTRODUCTION TO PUBLIC HEALTH RESEARCH

Unit 1 Definition and Concept of Public Health Research
Unit 2 Concepts of Sampling in Public Health Research
Unit 3 Reliability and Validity of Measurements in Public Health Research

UNIT 1 DEFINITION AND CONCEPT OF PUBLIC HEALTH RESEARCH

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
  3.1 Definition and Importance of Public Health Research
  3.2 Types of Variables in Public Health Research
  3.3 Measurement Scales in Public Health Research
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will be introduced to the definition and importance, as well as the concepts of public health Research. You will be made to understand the types of

variables available in public health research, and be introduced to the types of measurement scales in public health Research.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. define what is public health research
2. explain the importance of public health research
3. differentiate between variables and data
4. describe the various types of variables and data in public health research
5. describe the various level of measurement and measurement scales in public health research

3.0 MAIN CONTENT

3.1 Definition and Importance of Public Health Research

Public health research can be defined as the investigation of health-related problem in a population or community using scientific reasoning and methods involving collection of information or data obtained through observation of the phenomenon of interest. Hypotheses are formulated and tested by further observation and experimentation. The result of the research will provide information on the relationship between the suspected exposure/factor and the disease/health event. Public health research has provided insight into many health problems in the past that led to their eradication and control as the case may be. Most notable of these are the eradication of smallpox worldwide in 1980, control of cholera in many parts of the world simply by improving sanitation in communities, especially provision of potable water, sanitary disposal of faeces and personal hygiene. Another example of public health research is on the global effort to understand the drivers of poliomyelitis. Currently, there is massive effort to eradicate poliomyelitis because public health research had shown that the polio virus that is responsible for the disease has only human beings as its reservoir. Once the reservoir has been rid of the organism, the disease would be eradicated. Only a few nations in Africa and Asia are yet to achieve

this, unfortunately Nigeria is one of them! In the area of non-communicable disease, numerous public research results have helped to find solution that has led to the control of hypertension, coronary heart disease (an example is the well-known Framingham study in the US), diabetes mellitus and many more. Therefore, public health research has played a major role in the improvement of the quality of life and consequently increased life expectancy in man.

Apart from its usefulness in the prevention and control of health-related events, public health research results have also been useful in the identification of at-risk group in a population so that prudent use can be made of scarce resource that often characterize health care systems of most developing countries. In addition, research also provides better insight into the understanding of the complex interaction between man and his environment. This was initiated by Hippocrates (460-377 BC), the father of modern medicine as indicated in his publications *Epidemic I, Epidemic III and On Air, Waters and Places,* in which he made connection between disease and environmental conditions, especially in relation to water and seasons. Much later, John Snow (1813-1858) observed and recorded important environmental factors related to the course of cholera. He showed that cholera was a waterborne disease that travelled in both surface and underground water supplies. Since then public health research has evolved through many stages to what is known today. Some examples of major public health research in Nigeria includes the Nigeria Demographic Health Survey (2013, 2018), the World Health Organization StepWise Approach to Surveillance (2003), the Nigerian HIV/AIDS Indicators and Impact Survey (2018) where the epidemiology and prevalence of various disease conditions (e.g., malaria, HIV/AIDS, poliomyelitis, hypertension, diabetes, obesity, cancer, infant and child mortality, adult and maternal mortality, etc.,), their behavioural risk factors (e.g., tobacco use/smoking, physical inactivity, unhealthy diets, alcohol consumption, unhealthy sexual practices, poor family planning, risky road driving habits, etc.) and related socioeconomic and environmental factors (e.g., poverty, low education, poor housing conditions, etc.) were explored at the national and subnational levels. There are also multiple examples

of public health research conducted by independent public health researchers in Nigeria. Can you think of some of these examples?

The public health research process is usually initiated by conceptualizing a scenario that may likely describe health problems in specified population. A concept is a general idea or understanding derived from known instances or occurrences. It may be based on observations or experiences. After conceptualization, the next stage is to move to assign or identify variables that can be used to provide empirical intervention. An empirical approach emphasizes direct observation and experimentation, and variables are used to test hypotheses.

3.2 Types of variables in Public Health Research

A variable is an entity that can assume different values with different observations and can be measured or categorized. For example, height, weight, age, sex, colour are variables because they can assume different values or categories. Age for example, can be measured in hours, days, weeks, month s or years depending on the interest at hand. Furthermore, variable can be distinguished into two: dependent and independent. Dependent, outcome or the response variable is usually the variable in which we are interested in identifying the change that occurs due to the variation or presence of another, the explanatory or independent or effect variable. For example, a public health research aims to determine the relative influence of two types of exercise (walking and jogging) on lowering the prevalence of hypertension among older people living in a low-income community. For this public health research, the dependent (outcome or response) variable to measure will be the blood pressure (to determine hypertension), while the independent (explanatory) variables are the rates of walking and jogging among the older people in the community.

When variables are measured, they constitute data. Therefore, data may be thought of as observations, measurements of a phenomenon of interest such as level of cholesterol in the blood or information about disease condition collected from a

patient in the community. Data are usually established by observation, measurement or experiment for select number of variables. Measurements can be made in different categories depending on the phenomenon under study. They can be nominal, ordinal or interval.

Nominal data is the lowest form of data which may also be called qualitative data because they describe the quality of a thing or person.

3.3 Measurement Scales in Public Health Research

Definition: Measurement at its weakest level exists when numbers or other symbols are used simply to classify an object, person, or characteristic. When numbers or other symbols are used to identify the groups to which various objects belong, these numbers or symbols constitute a nominal or classificatory scale. For example, the psychiatric system of diagnostic groups constitutes a nominal scale. When a diagnostician identifies a person as “schizophrenic”, “paranoid” “manic depressive” or “psychoneurotic” he is using a symbol to represent the class of persons to which this person belongs, and thus he is using nominal scaling. Other examples include sex (female (1) and male (2)), marital status (single/never married, married, widowed, separated/divorce as 1,2,3,4), and blood group (A,B,AB,O as 1,2,3,4, respectively).

Since the symbols which designate the various groups on a nominal scale may be interchanged without altering the essential information in the scale, the only kinds of admissible descriptive statistics are those which would be unchanged by such transformation: the mode, frequency counts, etc. Under certain conditions, we can test hypotheses regarding distribution of cases among categories by using the non-parametric statistical test, X^2, or by using a test based on the binomial expansion. These tests are appropriate for nominal data because they focus on frequencies in categories.

Ordinal or Ranking Scale

Definition: It may happen that the objects in one category of a scale are not just different from the objects in other categories of that scale, but that they stand in some kind of relation to them. Typical relations among classes are: higher, more preferred, more difficult, more disturbed, more matured, etc. Such relations may be designated by the carat (>). For example, medical students can be ordered by their years of exposure in medical school, MBBS six> five>four, etc. Other examples of the ordinal scale are the level of education (1=no education, 2=primary school education, 3=secondary school education, 4= tertiary education) and the general health status (1=poor, 2=fair, 3=good, 4=very good). It can be seen from these examples that there is an element or ranking and ordering in the responses with option 4>3, option 3>2 and option 2>1.

Interval Scale

Definition: When a scale has all the characteristics of an ordinal scale, and when in addition the distances between any two numbers on the scale are of known size, then measurement considerably stronger than ordinality has been achieved. In such a case measurement has been achieved in the sense of an interval scale. An interval scale is characterized by a common and constant unit of measurement which assigns a real number to all pairs of objects in the ordered set. In this sort of measurement, the ratio of any two intervals is independent of the unit of measurement and of the zero point. In an interval scale, the zero point and the unit of measurement are arbitrary. E.g., we measure temperature on an interval scale. In fact, two different scales- Centigrade and Fahrenheit are commonly used. The unit of measurement and the zero point in measuring temperature are arbitrary; they are different for the two scales. However, both scales contain the same amount and the same kind of information. This is the case because they are linearly related. That is, a reading on one scale can be transformed to the equivalent reading on the other by the linear transformation

$$F= 9/5C+32$$

Where F= number of degrees on Fahrenheit scale

C= number of degrees on Centigrade scale

It can be shown that the ratios of temperature differences (intervals) are independent of the unit of measurement and of the zero point. For instance, “Freezing” occurs at 0 degrees on the centigrade scale, and “boiling” occurs at 100 degrees. While on the Fahrenheit scale, “freezing” occurs at 32 degrees and “boiling” at 212 degrees. Some other readings of the same temperature on the two scales are:

Centigrade	0	10	30	100
Fahrenheit	32	50	86	212

Notice that the ratio of the differences between temperature readings on one scale is equal to the ratio between the equivalent differences on the other scale. For example, on the centigrade scale the ratio of the differences between 30 and 10, and 10 and 0, is $\frac{30-10}{10-0} = 2$

For the comparable readings on the Fahrenheit scale, the ratio is $\frac{86-50}{50-32} = 2$

The ratio is the same in both cases- 2. In an interval scale, in other words, the ratio of any two intervals is independent of the unit used and of the zero point, both of which are arbitrary.

Ratio Scale

Definition: Measurement on the ratio scale has all the measurement properties of nominal, ordinal and interval scales in addition to having a true zero point. The ratio of any two measurements on the ratio scale is physically meaningful. For example, zero point on height (0 meter= 0 centimeter), weight (0 pound = 0 kilogram), time (0 minutes = 0 seconds), ages (0 years = 0 months = 0 days). Ratio scales are usually metric in that they can be measured with meaningful number. Hence, they can also be

considered as metric variables. There are two types of ratio/metric variables: discrete and continuous.

Discrete and Continuous Variables

The distinction between discrete and continuous variables has proved useful in organizing statistical procedures. A discrete variable is one which inherently contains gaps between successive observable values or a variable such that between any 2 (potentially) observable values there lies at least one (potentially) unobservable value. For example, a count of the number of bacterial colonies growing on the surface of an agar plate is a discrete variable. Whereas counts of 3, 4 and 5 are observable one of 3½ or 4½ is not. Any variable in the form of a count will be discrete, although not all discrete variables will be of this form.

In contrast, a continuous variable has the property that between any 2 (potentially) observable values lies another (potentially) observable value. A continuous variable takes values along a continuum i.e. along a whole interval of values. Length and weights are examples of continuous variables. A man's height might be either 1.0 meter or 1.05 meters, but it could also assume any intermediate value such as 1.001m. An essential attribute of a continuous variable is that, unlike a discrete variable, it cannot be measured exactly. With continuous variable there must inevitably be some measurement error. A continuous variable could be represented exactly only by an infinite decimal, and no one has yet written down all the digits in such a number (I recommend against your attempting this). This implies that when we write down a number for a continuous variable, we are only approximating its actual value by a number that reflects the precision of the measuring instrument used.

4.0 CONCLUSION

Public health research can be defined as the investigation of health-related problem in a population or community using scientific reasoning and methods involving collection of information or data obtained through observation of the phenomenon of interest. A variable is an entity that can assume different values with different

observations and can be measured or categorized. For example, height, weight, age, sex, colour are variables because they can assume different values or categories. When variables are measured, they constitute data. Therefore, data may be thought of as observations, measurements of a phenomenon of interest such as level of cholesterol in the blood or information about disease condition collected from a patient in the community. There are generally four level or scale of measurements from the lowest to the highest: the nominal, ordinal, interval or ratio level/scale. It should be noted that the arithmetic operations of addition and multiplication are not possible on measurements obtained from the nominal and ordinal scales. Only arithmetic operation of addition and subtraction is possible on the interval scale and all arithmetic operations (addition, subtraction, division and multiplication) are possible for measurements on the ratio scale.

5.0 SUMMARY

In this unit, the definition and importance of public health research have been discussed. We have also learnt the distinction between variables and data, and the types of variables in public health research have been highlighted. This unit has also discussed the issue of measurement and explained the four levels of measurement in public health research.

6.0 TUTOR-MARKED ASSIGNMENT

1. Give a concise definition of public health research
2. Explain the importance of public health research to health problems in Nigeria
3. Distinguish between variables and data
4. Discuss describe the various types of variables and data in public health research
5. Discuss the various level of measurement and measurement scales in public health research

7.0 REFERENCES/FURTHER READING

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

Bamgboye E.A. (2004). A companion of medical Statistics. Published by FOLBAM Health Research and Data Management Center, Ibadan. ISSN 978-056-661-9.

Bowers, D. (2008). *Medical Statistics from scratch: an introduction for health professionals*. London: John Wiley & Sons Ltd.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

UNIT 2 CONCEPTS OF SAMPLING IN PUBLIC HEALTH RESEARCH

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
- 3.1 Why Sampling and not the Whole Population?
- 3.2 Definitions of Concepts in Sampling
- 3.3 Various types of Sampling Techniques in Public Health
  - 3.3.1 Probability Sampling
  - 3.3.2 Non probability Sampling

4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will be introduced to the importance and relevance of sampling in public health Research. You will learn about the concepts of sampling and the various sampling techniques available in public health research.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. explain what is sampling, and its importance in public health research
2. define common concepts used in sampling
3. identify the best sampling technique in public health
4. describe the various sampling techniques available in public health research
5. explain what a bias is, and identify the types of bias in public health research

3.0 MAIN CONTENT

3.1 Why Sampling and not the whole Population?

Sampling is the statistical process of selecting a subset (called a "sample") of a population of interest for purposes of making observations and statistical inferences about that population. Public Health research is generally about inferring patterns of behaviours within specific populations. We cannot study entire populations because of feasibility and cost constraints, and hence, we must select a representative sample from the population of interest for observation and analysis. It is extremely important to choose a sample that is truly representative of the wider group (population) so that the inferences derived from the sample can be generalized back to the population of interest. The wider group is known as the target population, for example all premature babies born in Nigeria in 2019. It would be impossible to study every single baby in such a large target group (or every member of any population). So instead, we might wish to take a sample from a more accessible group. For example, all premature babies born in the maternity units of one teaching hospital each in each of the six geopolitical zone of Nigeria. This more restricted group is the study population. Suppose we take as our sample the last 300 babies born in each of these 6 teaching

hospitals. What we find out from this sample we hope will also be true of the study population, and ultimately of the target population. The degree to which this will be the case depends largely on the representativeness of the sample. Some of the reasons for taking sample rather than the whole population are given below:

Because we want to save time and money, we can concentrate on quality rather than quantity of data, and moreover, it is not necessary to include everyone.

By sampling we introduce sampling error. That is, the results from the sample will not be identical to the actual values of the population. error = bias + random error

We can eliminate or reduce some (not all) biases by careful design of the sampling scheme. We can reduce random error by a suitable choice of sample size.

The best approach for selecting a sample is random sampling. Chance determines who will be in the sample. This removes any possibility of bias in selection.

3.2 Definitions of concepts in Sampling

***Sampling unit*:** individual person; household; school; mosquito; bacteria colony; etc.

***Target population*:** the population in which we are interested, e.g. People of Borno state.

***Study population*:** the population from which we take our sample (often more limited than the target population, e.g. infants, women of reproductive age, adolescent male, etc.

***Sampling design*:** the scheme for selecting the sampling units from the study population.

***Sampling frame*:** the list of units from which the sample is to be selected.

3.3 Various types of Sampling Techniques in Public Health

Sampling techniques can be grouped into two broad categories: probability (random) sampling and non-probability sampling. Probability sampling is ideal if generalizability of results is important for our study, but there may be unique circumstances where non-probability sampling can also be justified.

3.3.1 Probability Sampling

Probability sampling is a technique in which every unit in the population has a chance (non-zero probability) of being selected in the sample, and this chance can be accurately determined. All probability sampling have two attributes in common: (1) every unit in the population has a known non-zero probability of being sampled (i.e., every unit in the population has equal chance of being selected to participate in the study), and (2) the sampling procedure involves random selection at some point. The different types of probability sampling techniques include:

A random sample is one drawn from a population of units in such a way that every member of the population has the same probability of selection and different units are selected independently. Random sample plays no favorites, but assigns the same selection probability to every member of the population. In doing so it assures that the population is fully known and defined by the investigator. This avoids biases and over- sights, which can lead to faulty inferences.

The use of the random table is designed to satisfy the following two conditions:

(1) In any predetermined geographic position in the table any one of the 10 digits 0 through 9 has a probability $_{1/10}$ of occurring.

(2) The occurrence of any specific digit in a predetermined geographic position in the table is independent in the probability sense of the occurrence of specific digits in other positions in the table. Note that these two conditions

equal probability and independence correspond to the two conditions in the definition of a random sample.

Various types of sampling methods are recognized. They are:

1. **Simple Random Sampling (SRS)**. This is the best scheme when we have a reasonably small and compact population. For large surveys it is often used at some stage.

We must first draw up a sampling frame listing all the units in the study population, then we randomly select the required number of units from the list. The selection of units may be done using random number tables (exercise)

2. **Stratified random sampling** involves dividing the population to be studied into different strata of similar social, environmental or health condition. A random selection of study units is then taken from each stratum.

An advantage of stratified random sampling is that information about the composition of the population with respect to a number of stratifying variables can be taken into account. If, for example, the age, sex, and ethnic composition of the population are known, then the sample can be selected to conform exactly to this composition by using these variables as the basis for stratification. Proportion is also taken into account.

3. **Cluster sampling** consists of groups or cluster of sampling units enclosed in an easily recognizable boundary. In forming clusters, the study units within a "cluster" do not need to be similar. However, as far as possible all clusters should contain approximately the same number of study units. A random sample is then taken from each cluster.

4. **Systematic Sampling:** in many situations this is as good as SRS and often more convenient. Involves a system by which the sample is chosen whereby a particular sequence is followed, must obtain a sampling frame and calculate the sampling fraction. For instance, if the sampling frame is 2000 and the sample size is 500, the sampling proportion is 2000/500 = 4 so 1 in every 4 persons will be selected to get the total of 500 people.

5. **Multistage Sampling**

Often it is not possible to do a SRS or systematic sample

-if a sampling frame is not available

-if the population is spread out over a wide area

Therefore, many times we need to do the sampling in two or more stages.

Example

For a national survey, make a list of all the states of the federation. Select a random sample of the first- stage units. In each of these units, take a random sample of the second –stage units (e.g. LGAs). In each of the LGAs take a random sample of the third-stage units (e.g. districts). In each of the district select a random sample of the fourth- stage units (villages or towns). In each of the towns or village select a random sample of the fifth-stage units (Households).

In each stage, it may be necessary to select the units with **probability proportional to size (PPS)** that is, the larger units have a greater probability of being selected. To do this we need a list of all the units in the region where the survey is to take place, together with some approximate measure of the number of the variable we are interested in.

Points to Consider in Sampling

1. The key reason for being concerned with sampling is that of validity, that is, the extent to which the interpretations of the results of the study follow from the study itself and the extent to which the results may be generalized to other situations.
2. Sampling is critical to external validity, that is, the extent to which findings of a study can be generalized to people or situations other than those observed in the study. To generalize validly the findings from a sample to some defined population requires that the sample has been drawn from that population according to one of several *probability* sampling plans. By a *probability sample* is meant that the probability of inclusion in the sample of any element in the population must be given *a priori*. All probability samples involve the idea of *random sampling* at some stage. In experimentation, two distinct steps are involved.

   Random selection—participants to be included in the sample have been chosen at random from the same population. Define the population and indicate the sampling plan in detail.

   Random assignment—participants for the sample have been assigned at random to one of the experimental conditions.
3. Another reason for being concerned with sampling is that of *internal validity*—the extent to which the outcomes of a study result from the variables that were manipulated, measured, or selected rather than from other variables not systematically treated.
4. The key word in sampling is *representative*. One must ask oneself, "How representative is the sample of the survey population (the group from which the sample is selected) and how representative is the survey population of the target population (the larger group to which we wish to generalize)?"
5. When a sample is drawn out of convenience (a nonprobability sample), rationale and limitations must be clearly provided.

6. If available, outline the characteristics of the sample (by gender, race/ethnicity, socioeconomic status, or other relevant group membership).

3.3.2 Non- Probability sampling

Nonprobability sampling is a sampling technique in which some units of the population have zero chance of selection or where the probability of selection cannot be accurately determined. Typically, units are selected based on certain non-random criteria, such as quota or convenience. Nonprobability sampling may be subjected to a sampling bias; therefore, information from a sample cannot be generalized back to the population. Types of non-probability sampling techniques include:

1. ***Convenience sampling***. Also called accidental or opportunity sampling, this is a technique in which a sample is drawn from that part of the population that is close to hand, readily available, or convenient. For instance, if you stand outside a shopping center and hand out questionnaire surveys to people or interview them as they walk in, the sample of respondents you will obtain will be a convenience sample. This is a non-probability sample because you are systematically excluding all people who shop at other shopping centers. The opinions that you would get from your chosen sample may reflect the unique characteristics of this shopping center only and therefore may not be representative of the opinions of the shopper population at large. Hence, the scientific generalizability of such observations will be very limited. Other examples of convenience sampling are sampling students registered in a certain class or sampling patients arriving at a certain medical clinic. This type of sampling is most useful for pilot testing, where the goal is instrument testing or measurement validation rather than obtaining generalizable inferences.

2. ***Quota sampling***. In this technique, the population is segmented into mutually-exclusive subgroups (just as in stratified sampling), and then a non-random set of observations is chosen from each subgroup to meet a predefined quota. In proportional quota sampling, the proportion of respondents in each subgroup

should match that of the population. For instance, if the population of Lagos consists of 70% normal weight, 15% overweight, and 13% obese people, and you wish to understand their diets preferences in a sample of 98 people, you can stand outside a shopping center and ask people their diets preferences. But you will have to stop asking overweight people when you have 15 responses from that subgroup (or obese people when you have 13 responses) even as you continue sampling other normal weigh group, so that the body composition of your sample matches that of the general Lagos population.

3. ***Expert sampling***: This is a technique where respondents are chosen in a non-random manner based on their expertise on the phenomenon being studied. For instance, in order to understand the impacts of a new governmental policy such as the Non-communicable disease policy, you can sample a group of public health professionals who are familiar with this policy. The advantage of this approach is that since experts tend to be more familiar with the subject matter than non-experts, opinions from a sample of experts are more credible than a sample that includes both experts and non-experts, although the findings are still not generalizable to the overall population at large.

4. ***Snowball sampling***: In snowball sampling, you start by identifying a few respondents that match the criteria for inclusion in your study, and then ask them to recommend others they know who also meet your selection criteria. For instance, if you wish to survey public health physical activity researchers in Nigeria and you know of only one or two such people, you can start with them and ask them to recommend others who also do physical activity research in public health. Although this method hardly leads to representative samples, it may sometimes be the only way to reach hard-to-reach populations or when no sampling frame is available.

Bias

Bias is an error introduced into an experimental design or study leading to misguided interpretation of the results obtained from such observations.

Types of Bias

(1) Selection Bias /method selection is not random. Is the observed association due to the way subjects were selected for the study?

(2) Response Bias/ information bias: Is the observed association due to errors of measurement or classification of the exposure and/ or the disease?

(3) Observers Bias: This is a bias arising from making wrong observations in the course of a study especially when the data is solely based on subjective observation which is not backed by video or voice recording.

(4) Detection/Instrument Bias: This is a bias resulting from faulty instrument which may have poor sensitivity or detection.

4.0 CONCLUSION

Sampling is the statistical process of selecting a subset (called a "sample") of a population of interest for purposes of making observations and statistical inferences about that population. You have learnt that sampling techniques can be grouped into two broad categories: probability (random) sampling and non-probability sampling. Probability sampling is ideal if generalizability of results is important for your study, but there may be unique circumstances where non-probability sampling can also be justified. It has been explained that the best type of sampling is random sample because it ensures that every units in the population has equal chance of being selected into the study and it eliminates biases. In this unit, it has been explained that the sampling process comprises of several stage. The first stage is defining the target population. A population can be defined as all people or items (unit of analysis) with the characteristics that one wishes to study. The second step in the sampling process is to choose a sampling frame. This is an accessible section of the target population (usually a list with contact information) from where a sample can be drawn.

5.0 SUMMARY

In this unit, you have learnt about the importance of sampling in public health research and the definitions of common terms used in sampling. You also learnt the various sampling techniques in public health research and the relative merits of each technique.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the importance of sampling in public health research
2. Define the common concepts used in sampling
3. Discuss the various sampling techniques in public health research
4. Discuss the concepts of bias in public health research

7.0 REFERENCES/FURTHER READING

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

Bowers, D. (2008). *Medical Statistics from scratch: an introduction for health professionals*. London: John Wiley & Sons Ltd.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Smith, R.V. (1984). *Graduate Research: A guide for students in the sciences*. Michigan: ISI Press

www.https://courses.lumenlearning.com/suny-hccc-research-methods/chapter/chapter-8-sampling/

UNIT 3 RELIABILITY AND VALIDITY OF MEASUREMENTS IN PUBLIC HEALTH RESEARCH

CONTENTS

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Psychometric properties of Outcome Measures

3.2 Definitions and Types of Validity
3.3 Definitions and Types of Reliability
3.4 Questionnaire Development and Design
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will be introduced to concepts of psychometric properties and the various types of validity and reliability in public health measurements. You will also learn how to design and develop questionnaires for public health research.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. explain the concepts of validity and reliability in public health research
2. define validity and reliability
3. explain the various types of validity and reliability
4. describe the process of questionnaire development

3.0 MAIN CONTENT

3.1 Psychometric Properties of Outcome Measures

The World Health Organization defines an outcome measure as a "change in the health of an individual, group of people, or population that is attributable to an intervention or series of interventions." Outcome measures allow us to objectively quantify the quality and attributes (e.g., mortality, readmission, patient experience, etc.) that we are trying to change or improve. There are various types of outcome measures, including: Self-report outcome measures / Patient-reported Outcome

Measures (PROM), Performance-based outcome measures, Observer-reported outcome measures, and Clinician-reported outcome measures. Self-report outcome measures are typically captured in the form of a questionnaire and are the most commonly used in public health research. In providing scientific utility and applicability of outcome measures, validity and reliability are two psychometric properties considered very important in public health. Reliability and validity of outcome measures are important for the interpretation and generalisation of research findings in public health. The term reliability is often used in relation to the "precision" of measurements, while the term validity is often used to describe the "accuracy" of measurements. Outcome measures (tests, an instruments, tools or questionnaires) that are not reliable and valid will not provide meaningful information, but rather will provide “numbers” or “categories” that give false impression of meaningfulness.

3.2 Definitions and Types of Validity

Validity is commonly defined as the extent to which an outcome measure (e.g., a questionnaire) measures what it is intended to measure. There are various forms of validity in any measurements, ranging from “face” to “criterion-related” validity.

i. **Face Validity** is the extent to which a test appears to measure what it is intended to measure. Face validity is when the instrument or tool appears to be measuring what it is supposed to measure with the content of test matching instructional objectives. For example, the World Health Organization Quality of Life Questionnaire (WHOQOL) will be considered to have good and acceptable face validity because it was developed by experts with knowledge of quality of life and all the items were approved by the experts to gauge the concepts of quality of life.

ii. **Content Validity** refers to the degree to which the terms in a questionnaire cover the relevant issues. Content validity is the degree to which a test includes all the items necessary to represent the concept being measured. The content validity of a test may vary widely depending on the question the test is being used to ask and

the population involved. For example, the WHOQOL will be considered to have acceptable content validity because the instruments appear to capture all the relevant domains of quality of life. That is, WHOQOL has contents that focus on the physical, psychological, social and environmental domains of quality of life.

Neither face nor content validity can be examined experimentally, and both are considered lower levels of validity. The higher forms of validity are criterion and construct validity and can both be objectively examined.

iii. **Construct Validity** reflects the ability of a test to measure the underlying concept of interest to the researcher. Construct validity is tested for when a new construct is established and there is no existing scale measuring the new construct or when some key aspects are omitted in the existing construct. There is no simple way to establish the construct validity of an outcome measure. However, Construct validity seeks the implications between a theoretical concept and a specific measuring device. It includes constructs like concepts, ideas, theories, etc. An example is a community health worker assessing the effectiveness of behavioural control therapy (BCT) on chronic pain among depressed patients at the primary health center. At every BCT session, she asks the patients to rate their level of pain on a 10-points rating scale (from zero no pain to 10 worst pain ever). The construct validity of the pain rating scale will test whether the instrument is actually measuring chronic pain from depression and not numbness, anxiety or discomfort.

iv. **Convergent and Discriminant** validity can be used to support the construct validity of a test. *Convergent validity* is demonstrated when scores on the test being examined are highly correlated to scores on a test thought to measure similar or related concepts. For example, scores on a gait index should correlate to scores from an activity limitation measure because the concepts of gait and activity limitation are related. *Discriminant validity* is demonstrated when scores on the test being examined are not correlated to scores on a test meant to measure a very different construct.

v. **Criterion related validity:** It is the correlation of a scale with some other measure of the trait or disorder under studies, ideally, a "gold standard" that has been used and accepted in the field. The criterion validity of an outcome measure is tested by comparing the result of the outcome measure or target test/measurement to a gold standard or criterion test/ measurement. If the target test measures what it is intended to measure, then its results should agree with the results of the gold standard criterion test. For example, a public health researcher interested in assessing the level of depression in a large population decided to create a 19-item shorter version of an existing well-established 42-item questionnaire on depression. This new questionnaire was created in order to reduce participants burden and to generate high completion rate among the participants in his study. However, to ensure that he has a valid new questionnaire, he will need to compare his new questionnaire against a method or instrument (measure) that is already well established. This well-established measurement procedure acts as the **criterion** against which the **criterion validity** of the new measurement procedure is assessed. When the new 19-item questionnaire on depression compares favourably with the well-established 42-item questionnaire, the new questionnaire will be considered to have acceptable evidence of construct validity. However, it should be noted that construct validity can be examined by giving both tests/measurements at the same time **(Concurrent validity)** or by giving the target test/measurement first to determine whether it predicts the findings of the gold standard test/measurement administered at a later time **(Predictive validity)**.

3.3 Definitions and Types of Reliability

Reliability is the proportion of observed variation in scores across repeated measurements that reflects actual variation in health levels and concerned with error in measurements. It can also be defined technically as the degree to which random error in a test is reduced. Reliability is characterized by a measure of the degree of consistency in the results obtained following repeated testing. It is also concerned with error in measurement. Several synonyms have been used for reliability; these include

precision, stability, reproducibility, consistency and predictability. There are various methods to assess the reliability of an instrument.

i. **Intra-rater (or intra-observer) reliability; also known as test-retest reliability:** This describes the agreement between results when the instrument is used by the same observer on two or more occasions (under the same conditions and in the same test population). Test-retest reliability measures stability over time in repeated applications of the test. It can be defined as the consistency in scores obtained on an instrument on two occasions separated by some interval of time. Self- report measures that require individuals to respond to a series of written questions should be particularly examined for test –retest reliability. Test-retest reliability is examined by having individual complete the measure on more than one occasion with the assumption that no real change will have occurred between sessions. For example, the one-week test-retest reliability of the new 19-item depression questionnaire will be determined by having the participants in the study to complete the questionnaires on two separate occasions seven days apart. Test-retest reliability can be measured statistically using the intraclass correlation coefficients.

ii. **Inter-rater (or inter-observer) reliability:** this measures the degree of agreement between the results when two or more observers administer the instrument on the same subject under the same conditions. For example, the inter-rater reliability for a blood pressure measuring instrument (sphygmomanometer) would be determined by having two different research assistants with the same training on blood pressure measurement use the same sphygmomanometer to measure blood pressure of same participants. The values of measurements from each of the two research assistants are then compared to determine the level of agreement between both raters. You will notice that the participants and the measuring instrument (sphygmomanometer) are constant but only the raters are different; hence the need to determine inter-rater reliability when confronted with this kind of public health research scenario. Inter-rater reliability can be measured

using the Cohen's kappa (k) statistic. Kappa indicates how well two sets of (categorical) measurements compare. Kappa values range from -1 to 1, where values ≤0 indicate no agreement other than that which would be expected by chance, and 1 is perfect agreement. Values above 0.6 are generally deemed to represent moderate agreement. Limitations of Cohen's kappa are that it can underestimate agreement for rare outcomes, and that it requires the two raters to be independent.

iii. **Internal consistency reliability:** Internal consistency reliability is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results. It describes the degree of agreement, or consistency, between different parts of a single instrument. It is mostly a statistical procedure rather than actual measurement. Internal consistency can be determined using Cronbach's alpha (α). Cronbach's alpha values range from minus infinity to one, with one indicates perfect internal consistency, and a negative value suggests that there is greater within-subject variability than there is between subjects. Cronbach's alpha values above 0.7 are generally deemed acceptable.

3.4 Questionnaire Development and Design

When we gather information to describe or explain a situation, much care is needed to ensure that the data collected reflect the real situation as closely as possible.

Error in Data = Response Error + Processing Error

Where the response error is the difference between the 'true' answer and
What is written on the data collection form questionnaire?

Response error can arise as a result of each of the following:

Questionnaire faults

Interviewer errors

Error in subject's response

The following notes are aimed to act as a checklist for ways of minimizing the response error.

I. Questionnaire design
II. Composing the questions

General Points:

- Keep the questions short
- Must be clear and unambiguous – don't use technical jargon
- Only ask one thing at a time
- Use simple language
- Avoid leading questions, negative questions and hypothetical questions
- Write the questions exactly as they are to be read out.

Open and closed questions:

There are three main options-

1) respondents answer in any form they like and interviewer tries to report what they say;
2) As above, but the interviewer then fits the response into Pre-coded categories;
3) Respondents given l limited choice of responses.

Usually types (2) and (3) are used in large-scale survey; pre-coding the responses lends to easier handling at a later stage.

- Make sure all the common responses are included. A pilot study is usually essential for this;
- Always have a category for "other" responses, and include space On the form to record what the response was;

A "don't know" and "not applicable" category may also be necessary.

Cultural relevance:

- How far back are you expecting your respondents to recall accurately? A suitable period will depend on the event you are talking about.

Bias due to wording of questions:

- Try to keep the wording of the questions "neutral" so that respondents don't just give the reply that they think is expected of them- especially for questions on attitudes.

Sensitive questions:

- Is your question acceptable?
- Is it practical to expect to be able to get this information?
- Might an indirect approach using several questions help? There are several techniques available for ensuring confidentiality, but they may not be feasible in all situations.
- Put sensitive questions at the end of the questionnaire.

Composing the questionnaire

Introduction:

- Interviewer should begin with a brief introduction explaining the purpose of the survey. It is useful to write this at the beginning.

Identification:

- **Title of questionnaire.**
  If you are doing repeated surveys of the same population, very important to include distinguishing survey number.
- Each respondent should be clearly and uniquely identified.

Name and identification (ID) number are the minimum an ID number should be reproduced on each sheet of the questionnaire.

General Layout:

- Questionnaire must be easy to read and to use.
- Use good quality paper- it may have to stand up to rough handling.
- Choose paper size carefully, e.g. A4. Small pieces of paper are much harder to sort through and not so convenient on a clipboard.
- Use one side of the paper only.
- Lay out of the typing is extremely important. Leave plenty of space for the answers to each of the questions.

Ordering of Questions:

- The flow of questions should be logical, is to deal with all the questions on one issue before going on to the next one.
- Usually want to ask a few questions to "classify" the individual e.g. by age, sex, ethnic group, and occupation. Put these near the beginning, unless you think they are sensitive.
- Put sensitive questions at the end. Then if the interview is not completed you still have some of the data.

Instructions to interviewers:

- Sometimes necessary to print instructions on the questionnaire to guide the interviewer. Especially important where there are "branches or jumps". E.g. if the answer is "No", go to question 8.
- Transitional statements: these are explanations to be read out by the interviewer, e.g. to explain that you are moving on to a new subject or to define a term you are using. They should be written exactly as they are to be read.
- Distinguish instructions from the things which are to be read out, e.g. by printing in Italics or capitals.

Coding

Coding means that the answers to the questions are classified into a defined set of categories. It can be done at various stages:

1) By the respondent- the interviewer asks him/her to select from a list of responses:
2) By the interviewer- the respondent answers freely and the interviewer ticks or circles the appropriate response on the questionnaire:
3) After the interview (either by the interviewer or back in the office)- the response is recorded verbatim and coding is done later.

When deciding who is to do the coding need to consider:

- How it will affect the smooth running of the interview;
- How it will affect the accuracy of the data.

For some questions the categories are relatively obvious, e.g. "Have you ever smoked cigarette?" Can only get responses "Yes", "No" and perhaps "Refuses to answer". For other questions, especially if opinions are sought, coding can be much more problematic. If pre-coding is to be done, it is essential to do a pilot study to find out the kinds of response that you are likely to get. If this is not possible, best to record the response verbatim and decide on the codes after the study is complete. In this case:

- When the forms are ready for coding, examine a sample of the questionnaires and decide on appropriate codes. Do this before starting the actual coding process.
- Avoid changing the codes after starting the coding (Sometimes it may be OK to add an extra code if a completely new response turns up).

NB: The term "coding" also refers to the action of attaching "codes" (either letters or numbers) to the responses once they have been categorised. This stage is essential when the data are to be analyzed by computer.

Recording the data for Computer analysis

There are some further points to be considered in relations to storage of the data from the questionnaire in a computer.

- It is important to make clear which parts of the form are to be entered in the computer. Providing boxes for each answer usually does this.
- One box is needed for every digit or letter. Hence, for each response you need to work out the maximum number of digits which may be needed, e.g. for height, what is the largest height that you are likely to record?
- It is useful to reserve the right-hand side of the form for "coding boxes". The questions, codes written responses etc, are all put on the left-hand side and the coding boxes on the right are reserved for the numbers (and letters) which are to be entered into the computer.
- The first few coding boxes will usually contain a code for the type of questionnaire, followed by identification codes for the respondent (e.g. village code, household code, and individual code).
- There may be some specific requirements for the way the data are to be recorded, depending on the software that it is intended to use. Consultation with data processing staff and looking ahead to the requirements of the proposed analysis will help the design.
- Some computer packages cannot distinguish between a blank and a zero. This needs to be considered when deciding on numerical codes e.g. code "missing values" as 9 rather than leaving a blank.
- Except where questions are not applicable and have hence been skipped, no blanks should be left since they might be ambiguous.

Multiple response questions

Sometimes the respondent can give more than one response, which can cause a problem when trying to code. There are two ways of dealing with this type of question

1) Have one coding box for each possible response. E.g. for "where do you obtain your water?" You might have boxes for "well", "river", and "taps" and codes 1=Yes, 2=No. Then for an individual answering "I get most of my water from the well, but also some from the river", you would put "I" in the well and river boxes, and "2" in the box for taps.
2) Alternatively, determine the maximum number of responses an individual is likely to give. E.g. if you have decided on six categories of response for a particular question but you think no respondent is likely to give more than three of them, you could simply reserve three coding boxes and record the numerical codes to the responses given.

In most cases (1) is more convenient when it comes to analyzing the data, unless you have sophisticated statistical software.

Examples of bad questions:

1 Does your child have an Upper Respiratory Tract Infection?
2 Do you go to the health centres to get treatment for back pain?
3 Do you think the hospital and health centres give a good service?
4 More should be spent on village health-care service, and less on expensive hospitals. Do you agree?
5 There is no point in having low-level health workers in the village unless there is a good back-up hospital for difficult cases. Do you agree?
6 Are you not able to sleep at night?
7 If health centres were built in the village, would you go there instead of to the traditional health?

Stages of data entry and management by microcomputer

1. -Data are usually entered onto the computer by using a database such as SPSS, STATA, Epi Info™, SAS, etc
2. -Often data from one questionnaire are stored in one record
   I record per household,
   or I record per person

Records are stored in a data file.

3. -Data should be entered twice by two different operators and stored in two separate data files. The data files should be compared (by a suitable computer program) and differences investigated. Good at detecting typing errors. This is known as verification.
4. –Range and consistency checks should be carried out on the data file.
5. -Files are often stored on memory disk or hard disk. Backup copies should be made, both during the data entry period and when it is finished.
   One copy should be stored in a different physical location.
6. -Often new data files have to be created from the original one- e.g. when creating new variables, it is a good idea to keep copies of both the original and the new data files.

4.0 CONCLUSION

Outcome measures allow us to objectively quantify the quality and attributes that we are trying to change or improve. There are various types of outcome measures. Self-report outcome measures are typically captured in the form of a questionnaire and are the most commonly used in public health research. Reliability and validity of outcome measures are important for the interpretation and generalisation of research findings in public health. The term reliability is often used in relation to the "precision" of measurements, while the term validity is often used to describe the "accuracy" of measurements. Outcome measures (tests, an instruments, tools or questionnaires) that are not reliable and valid will not provide meaningful information, but rather will provide “numbers” or “categories” that give false impression of meaningfulness.

5.0 SUMMARY

In this unit, you have been introduced to the concept of outcome measures in public health and their psychometric properties. You have learnt that the two most important and commonly reported psychometric properties of measurement tools/instruments in public health are reliability and validity. We have outlined the definitions and described various types of reliability and validity in public health research. You have also learnt the details of how to design and develop questionnaires for public health research.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the concept of psychometric properties of outcome measures in public health research
2. Define and describe the various types of validity
3. Define and describes the various types of reliability
4. Describe the various steps in the development of a new questionnaire

7.0 REFERENCES/FURTHER READING

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

De Negri B. & Thomas, E. (2003). "*Making sense of Focus Group Findings: A systematic Participatory Analysis Approach"*. Washington, DC: Academy for Education Development.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

McDowell, I., & Newell, C. A. (1996). *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

Portney, L.G., & Watkins, M.P. (2000). *Foundation of Clinical Research: Application to practices*. (2nd ed). Upper Saddle River, New Jersey: Prentice Hall.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Smith, R.V. (1984). *Graduate Research: A guide for students in the sciences.* Michigan: ISI Press

Streiner, D.L., Norman G.R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use*. (5th ed.). Oxford: Oxford University Press.

Yeomens, S.G. (2000). The clinical application of outcomes assessment. Stanford Appleton and Lange.

www.fhi.org/nr/rdonlyers/..../datacollectorguideenrh.pdf

Module 2 **Research Methods and Data Analyses in Public Health**

Unit 1 Study Designs in Public Health Research
Unit 2 Quantitative vs Qualitative Research Methods
Unit 3 Data Analysis Methods in Public Health Research

UNIT 1 STUDY DESIGNS IN PUBLIC HEALTH

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
  3.1 Introduction to Study Design in Public Health Research
  3.2 Descriptive (Observational non-analytic) Research
  3.3 Analytical (interventional/experimental) Research
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

Since the various health related events in public health are diverse, so are the various research methods employed to prevent and reduce the burden of disease and other health related events. This unit will introduce you to the concept of research design in public health. You also learn the various types of research in public health and be able to differentiate between various types of public health research.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. explain study design in public health research
2. identify different classification of research designs in public health
3. discuss descriptive research design and its classifications
4. discuss analytic research design and its classifications
5. explain randomized control clinical trial and its features

3.0 MAIN CONTENT

3.1 Introduction to Study Design in Public Health Research

Study design refers to the method used to collect epidemiological data. While the generation of hypotheses may come from anecdotal observations, the testing of those hypotheses must be done by making controlled observations, free of systematic bias. Statistical techniques to be valid must be applied to data obtained from well-designed studies. Otherwise, solid knowledge is not advanced. An epidemiological study could be designed in several ways so as to collect new data. Two principles should always be followed: the study should be comparative and we should seek to avoid all potential sources of bias.

Basically, two main classes of study type may be identified: **Descriptive (observational non-analytic) and Analytical** (**Observational, Intervention/Experimental)**. By far the vast majority of epidemiological studies are observational, meaning that data are collected simply 'to see what is happening', as shown in the classical study of Doll & Hill (1950) on lung cancer and cigarette smoking. Here **nature** determines who is exposed to the factor of interest and who is not exposed. These studies demonstrate association. Association may imply causation, or it may not. By contrast intervention study is an experiment: that is, things are made to happen. Intervention studies are considered as the gold standard as far as etiological investigations are concerned.

3.2 Descriptive (Observational non-analytic) Research

Descriptive design (may be referred to as Observational non analytic) involves the collection of data that will assist in providing useful information about the exposure and health related status without necessarily providing concrete etiologic clue or inference. It will however, provide information about the pattern and frequency of the health related-state or events in that population. This provides a useful template for analytical design. According to Merrill (2008) descriptive study assists us in:

1. Providing information about a disease or condition.
2. Providing clues to identify a new disease or adverse health effect.
3. Identifying the extent of public health problem
4. Obtaining a description of the public health problem that can be easily communicated.
5. Identifying the population at greatest risk.
6. Assisting in planning resource allocation.
7. Identifying avenues for future research that can provide insights about an etiologic relationship between an exposure and outcome.

Descriptive studies include case/case-series studies, cross sectional studies and ecological studies. Each design is suitable for different research setting and conditions.

Case study aims at providing complete understanding about a problem or situation. It can be described as a snapshot of a description of a problem regarding a group or individual. Its strengths lie in the fact that it provides in depth description of the health-related event or state, provides a clue to identifying a new disease and potential areas of research. However, conclusions are limited to individual, group or context of the study and most importantly cannot be used to establish cause-effect relationship. However, a health care worker may see a series of patients (cases) with similar but unusual symptoms or outcomes, find something interesting and write it up as a study. This is a case-series. Case series studies usually point to a need for further

investigations. Bowers (2008) wrote about a classical an example from practice on a case series:

"In 1981, a drug technician at the Center for Disease Control in the USA, noticed an unusually high number of requests for the drug pentamidine, used to treat Pneumocystic carini pneumonia (PCP). This led to a scientific report, in effect a case-series study, of PCP occurring unusually in five gay men in Los Angeles. At the same time a similar outbreak of Kaposi's Sarcoma (previously rare except in elderly men) in a small number of young gay men in New York, also began to raise questions. These events signaled the arrival of HIV in the USA."

Cross sectional studies provide the opportunity to measure all the variables at one time. It does not permit temporal relation between exposure and outcome but allows control over study population and measurements of several associations between variables at the same time. It has potential bias such as, poor recall and higher proportion of long survivors. Examples of cross-sectional studies are surveys in which the distribution of a disease, disability, pathological condition, immunological condition, nutritional status, fitness, or intelligence, etc., is assessed. This design may also be used in health systems research to describe 'prevalence' by certain characteristics – pattern of health service utilization and compliance – or in opinion surveys. A common cross-sectional study procedure used in family planning and in other services is the KAP survey (survey of knowledge, attitudes and practice) (WHO, 2001).

Ecologic study involves making comparison between populations or groups rather than among individuals. It takes advantage of existing data, relatively easy to conduct, gives quick results and can be used to evaluate policies, programs or regulations implemented at ecological level. It also allows estimation of effects not easily measurable on individuals However; exposure and disease/injury outcomes are not measured on the same individuals. Three types of Ecologic measures were identified by Morgenstern (1998) as follows: aggregate, environmental and global. ***Aggregate***

measures give the summaries of observations based on individuals within a group, while ***environmental measure*** deals with the physical characteristics of a place such as pollution level, hours of dust storm, harmattan haze or mean temperature. Global measures on the other hand deals with attributes of groups for which no analogue at the individual level exists, for example, population density, number of health facilities, policies, laws etc. The type of ecological analysis performed depends on whether the unit of analysis of variable(s) in the study is on the individual or group level. If data are collected on the individual level, a value for each variable is measured for each individual in the study. If data are collected on the ecologic level, a value for each variable is measured for the group in the study. Research may involve complete ecologic analysis, partial ecologic analysis or multilevel analysis. Ecological study is becoming more widespread because it is not often feasible to measure accurately individual exposure levels for a large number of people because of resource and time constraints.

3.3 Analytical (Interventional/experimental) Research

They are distinct from the descriptive design in that they utilize a comparison group that has been explicitly collected. Their usefulness is largely in the fact that they are used to identify environmental causes of events or health-related events or states. Analytic study designs in environmental epidemiology and epidemiology in general are based on the comparison of individuals who are classified according to exposure and injury/disease status. Here the unit of analysis is the individual. They fit into two major categories, observational (case-control, cohort) and experimental. In observational study the researcher evaluates the strength of the relationship or association between an exposure and health related state of event. The observed variables are beyond the control or influence of the investigator. On the other hands in experimental studies some of the participants in the study are deliberately manipulated for the purpose of studying an intervention effect. However, if the intervention being assessed is an environmental exposure being investigated for its adverse health effects, it would be unethical to assign the intervention. Instead a dose effect relationship can be investigated by categorizing the environmental exposure by time and dose. Nature

often produces natural experiment through peoples' behavior and life style. A natural experiment is an unplanned type of experimental study where the level of exposure to a presumed cause differ from a population in a way that it is unaffected by extraneous factors so that the situation resembles a planned experience. John Snow's investigation on cholera epidemic in 1854 is a classic example of natural experiment. The result of the investigation showed that the rate of infection was 8.4 times greater among those who obtained water from the contaminated source than those who obtained water from another alternative source.

Perhaps the most important utility of analytical studies is that it enables us to determine the level of risk associated with developing the disease in the presence of the environmental factor which enables the establishment of exposure limits and standard. Risk assessment is a tool to integrate exposure and health effects in order to identify the health hazards in humans.

Various methods are used in the analytic designs. They include experimental, cohort, case-control and case-crossover. Each of them has its specific application. Experimental study design examines the relations between intervention and outcome variables in a group of cohorts that has been followed over time. It may be sometimes referred to as intervention study. This type of design produces the strongest evidence for causal association, may produce faster answers to research questions than cohort study and may be the most appropriate design to answer certain research question. However, it may involve ethical barriers, outcomes which are too rare, cost and time intensive and high attrition (loss to follow-up).

i. **Cohort or Prospective study** involves the following up of people for a period of time in order to describe the incidence or natural history of a health outcome among **two groups**; exposed and unexposed. The procedures of a prospective study are relatively simple. A sample of the population is selected then information is gathered regarding those who have and those who do not have the characteristics in question, e.g. smoking-smokers and non-smokers. Those with

the characteristics are known as the experimental or exposed group, those without the characteristics are the control or non-exposed group. The population is then followed over a period of time to see how many of those with the characteristic develop; or die from the disease in comparison with those who do not have the characteristics.

Characteristic	With Disease	Without Disease	Total
With (exposed) Without (not exposed)	a c	b d	a + b c +d
Total	a + c	b + d	N

Compare the incidence of the disease among those who have the characteristic in relation to the incidence of disease among those who do not have the characteristics, that is, we

Compare a/ a+b with c/ c+d

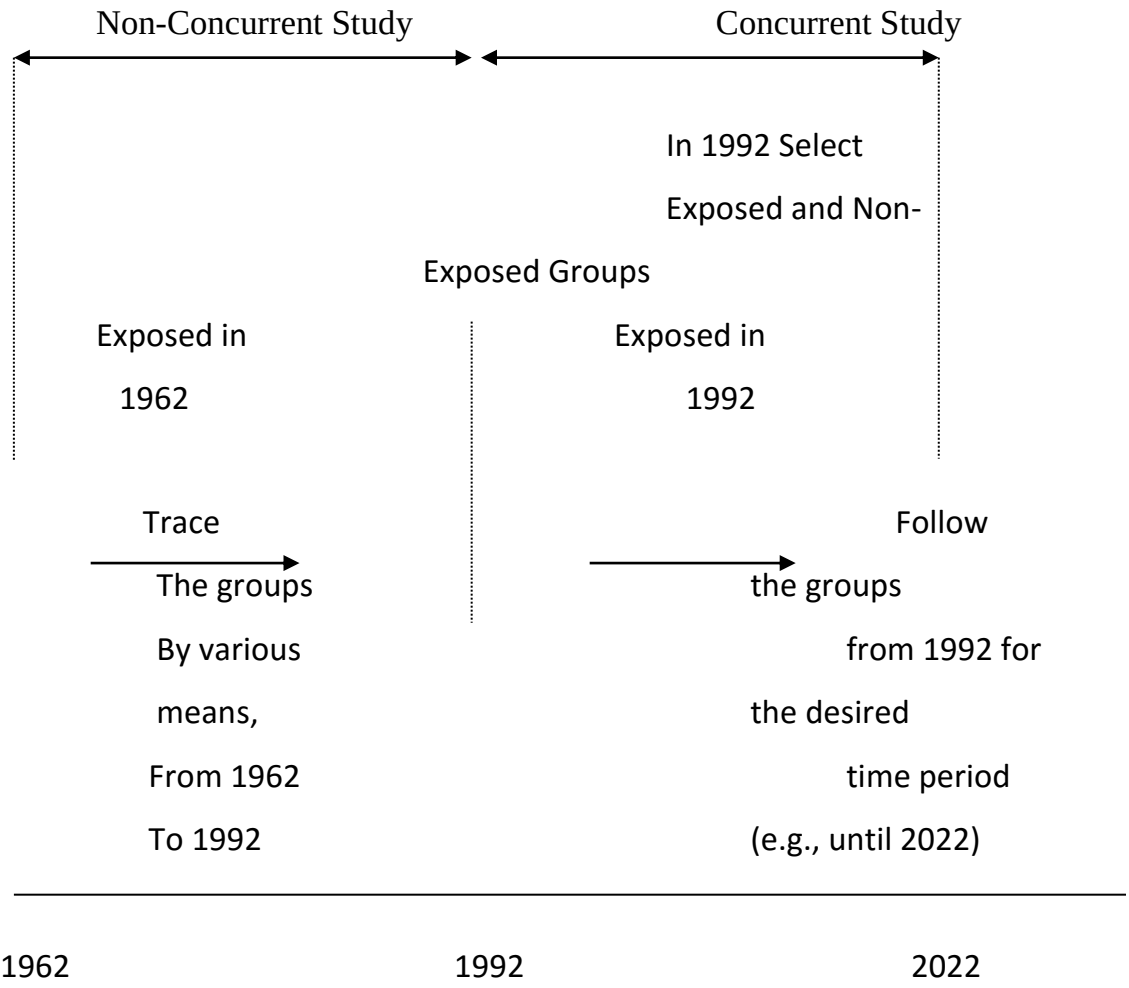
There are usually two types of prospective studies:

A- Concurrent where the investigator begins with a group of individuals and follow them for a number of years.

B- Non-concurrent- the investigator goes back in time selects his study groups and traces these groups over time, usually to the present, by a variety of tracing methods.

(a) 10,000 to 20,000 people followed to obtain 50 deaths from CHD.

(b) 1,000,000 people must be followed to obtain 50 deaths from leukemia.



Diagrammatic representation of concurrent and non-concurrent prospective studies

Analysis of data from prospective data

a) Direct estimate of risk (incidence)
b) Relative risk
c) Attributable risk.

The major advantage of this method is that incidence rate of the disease or health event can be generated from which the risk of developing the disease can be calculated for preventive purpose. It can also be used to establish time sequence of

events as well as avoid bias in measuring predictors from knowing the outcome. In addition, several outcomes can be assessed. However large samples are often required for follow-up and therefore not feasible for rare diseases. It requires large sums of funds and time, only one risk factor can be examined and has high attrition rate.

ii. **Case-control or Retrospective study** involves a backward look at association between exposure and health related outcomes among **two groups**; cases (those with the health-related event) and controls (those without the health-related event).

Selection of cases and controls

Various methods have been used to select cases and controls for retrospective studies some only infrequently. In developing guidelines for the selections of controls, any factors already known or strongly suspected to be related to the disease should be taken into consideration if unbiased statistical tests of significance of the specific characteristics being studied are desired**.**

Cases and controls may be selected as follows:

1. All cases diagnosed in the community (in hospitals, other medical facilities, include private clinic). For controls; probability sample of the general population in a community.
2. All cases diagnosed in all hospitals in community, for control; non-cases in a sample of the general population or subgroup of a sample of general population.
3. All cases diagnosed in all hospitals in the community, for control sample of patients in all hospitals in the community who do not have the disease or related diseases being studied.
4. All cases diagnosed in a single hospital, for control; sample of patients in same hospital where cases were selected.
5. All cases diagnosed in one or more hospitals. For control; sample of individuals who are residents in same street or neighborhood of cases.

6. Cases selected by any of the above methods. Control; spouses, siblings or associates of cases accident victims.

Framework of a retrospective study

Characteristic	**Cases**	**Control**	**Total**
With	a	b	a+b
Without	c	d	c+d
Total	**a+c**	**b+d**	**a+b+c+d=N**

If $\frac{a}{a+c}$ is statistically significantly greater than $\frac{b}{b+d}$

an association can be said to exist between the disease and the characteristic.

A more realistic example will illustrate this better. The following is a result of a Retrospective study on Cigarette smoking and cancer of the bladder.

Smoking habit	**With cancer of bladder**	**Control**	**Total**
Smokers	192	156	348
Non-smokers	129	181	310
Total	**321**	**337**	**658**

The percentage of Cigarette smokers among the bladder cancer patients is 60 percent

$\frac{192}{321}$ compared to 46 percent $\frac{156}{337}$ among the controls.

It may be desirable to match cases and controls for factors such as age and sex whose association with the diseases under study is already known or has been observed in available mortality statistics, morbidity surveys, or other sources. In addition, when cases and controls are matched on any selected factor, the influence of that factor on the disease can no longer be studied. This emphasizes the need to exercise caution in the number of variables or factors selected for matching even when feasible.

It is effective for rare health related events, requires less time when compared with cohort study. It cannot however yield direct risk of developing the disease; rather the odds ratio is provided. It is limited to one outcome alone, less effective than the cohort study, potential recall and interviewer bias and has a potential survival bias.

iii. **Case-crossover** study is becoming increasingly common in public health where the relationship between environmental exposure and development of disease is sought. It involves comparing the exposure status of a case immediately before its occurrence with that of the same case at a prior time. The argument here is that if precipitating events exist, they should occur more frequently immediately prior to the onset of disease rather than during a period more distant from the disease onset. The case-crossover study design is especially appropriate where individual exposures are intermittent, the disease occurs abruptly and the incubation period for detection is short, and the induction period is short. Individuals serve as their own controls, with the analytic unit being time-where the time just before the acute event is the "case" time compared with some other time, referred to as the "control" time. It does not however control for confounding for time related factors.

iv. **Randomized controlled clinical trial** is a prospective experimental design to compare one or more interventions against a control group in order to determine the effectiveness of the interventions. A clinical trial may compare the value of a drug versus a placebo. A placebo is an inert therapy that looks like the drug that is being tested. It may compare a new therapy with a currently standard therapy, surgical with medical intervention, two methods of teaching reading, or two methods of psychotherapy.

Features of randomized clinical trials

1. There is a group of patients which are designated patients. All criteria must be set forth and met before a potential candidate can be eligible for the study. Any exclusion must be specified (Set inclusion and exclusion criteria).
2. Any reasons for excluding a potential patient from participating in the study must be specified prior to starting the study. Otherwise, unintentional bias may enter.
3. Once a patient is eligible, he/she is randomly assigned to the experimental or control group. Random assignment is not 'haphazard' assignment but rather it means that each person has an equal chance of being an experimental or control patient. It is usually accomplished by using a table of random numbers (see section on sampling).
4. Clinical trials may be double blind (see note on blind studies).
5. While clinical trials often compare a drug or treatment with placebo, they may also compare two treatments with each other or a treatment and 'usual care'. Trials that compare an intervention with 'usual care', obviously cannot be blinded, for the assessment of effect (measurement of weight or blood pressure, or some hypothesized effect of weight loss) should be done in a blinded fashion, with the assessor not knowing which group the participant has been assigned to.
6. It is important to match the patients with the control (they should be similar in most aspects e.g. age, sex etc.) so that the difference in outcome can be attributed to differences in the treatment and not to different characteristics of the two groups. Randomization helps to achieve this comparability.
7. Note that new drugs must have undergone phases I and II which determines toxicity and safety and efficacy, respectively. These studies are done on small numbers of volunteers. Phase III trials are large clinical trials, large enough to provide an answer to the question of whether the drug tested is better than placebo or than a comparison drug.

CONCLUSION

In this unit we have outlined the various types of study design in public health research. It has been emphasized that two main classes of study type are: Descriptive (observational non-analytic) and Analytical (Observational, Intervention/Experimental). By far the vast majority of epidemiological studies are observational, meaning that data are collected simply 'to see what is happening'. Analytical studies utilize a comparison group that has been explicitly collected. Their usefulness is largely in the fact that they are used to identify environmental causes of events or health-related events or states. Analytic study designs fit into two major categories, observational (case-control, cohort) and experimental. In observational study the researcher evaluates the strength of the relationship or association between an exposure and health related state of event. The observed variables are beyond the control or influence of the investigator. On the other hands in experimental studies some of the participants in the study are deliberately manipulated for the purpose of studying an intervention effect.

4.0 SUMMARY

In this unit, you have been acquainted with the various study design in public health research. You have learnt that public health research can be broadly divided into two categories, vis descriptive design and analytical design. You have understood the various sub-categories of study designs that constitute the two broad categories of public health research, when to choose which research design is suitable for specific research questions. This module has also introduced you to qualitative research method and taught you the various steps/details of how to conduct observation, in-depth/semi structured interview and focus group discussion.

5.0 TUTOR-MARKED ASSIGNMENT

1. Discuss the concepts of study design in public health research
2. Differentiate between descriptive and analytical studies
3. Explain the principles of cohort (prospective) and case-control (retrospective) studies

4. Outline the general idea of randomized control clinical trial

7.0 REFERENCES/FURTHER READING

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

Bowers, D. (2008). *Medical Statistics from scratch: an introduction for health professionals*. London: John Wiley & Sons Ltd.

Hoffman, J.I.E. (2015). *Biostatistics for Medical and Biomedical Practitioners*. Elsevier.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Smith, R.V. (1984). *Graduate Research: A guide for students in the sciences*. Michigan: ISI Press

www.fhi.org/nr/rdonlyers/..../datacollectorguideenrh.pdf

UNIT 2 QUANTITATIVE AND QUALITATIVE RESEARCH METHODS

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
    3.1 Introduction to Quantitative and Qualitative Methods
    3.2 Qualitative Research Methods
    3.3 Types of Qualitative Research Methods
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

This unit will introduce you to the distinctions between quantitative and qualitative research methods. You will also learn the various types of qualitative research methods in public health and how to conduct focus group interview in public health research.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. differentiate between qualitative and quantitative research methods
2. explain the types of qualitative research methods
3. discuss the various steps involved in focus group
4. identify the strengths and weaknesses of focus group

MAIN CONTENT

3.1 Introduction to Quantitative and Qualitative Methods

The two basic approaches used in research methods in public health can be categorized as qualitative and quantitative. While the psychometricians try to measure it, the experimentalists try to control it, the interviewers ask questions about it, observers watch it, statisticians count it, evaluators value and the qualitative inquirers find meaning to it. The two methods are complimentary in the sense that they both look at the research problem from different points with the view of proffering a solution. The qualitative approach provides information on why the health problem using small population, exploring cultural influence by obtaining insiders views using somewhat unstructured methods, while the quantitative approach focuses on how many based on large populations obtaining outsiders views using constructive methods.

In Public Health quantitative study is the systematic empirical investigation of observable phenomena via statistical, mathematical and computerizational techniques. The objective of quantitative method is to collect data on phenomena that are suspected to be related to the causation or influence the occurrence of disease or health related event in human population. The method involves collection of variables which are measurable, in other words, variables that can assume different values.

3.2 Qualitative Research Methods

Qualitative research is useful when there is need to explore and explain behaviours-explains rather than describes. When the subject matter is unfamiliar or insufficiently researched especially at the mapping level and when suitable vocabulary is not available to communicate with respondents. Qualitative research asks why, how and under what circumstances events occur, seeks depth of understanding, views social phenomena holistically, explores and discovers and provides insight into the meanings of decisions and actions. The qualitative researcher builds a complex, holistic picture, analyses words, reports detailed views of informants, conducts the study in a natural

setting and is an instrument in the research setting. Qualitative methods give participants the opportunity to respond in their own words, rather than forcing them to choose from fixed responses which makes the responses to be more meaningful, culturally relevant and salient to the participant. In this case, the responses may not have anticipated by the researcher and are rich and explanatory in nature. In addition, it allows the researcher to probe beyond what would have been given in a quantitative setting.

3.3 Types of Qualitative Research Methods

1. Observational

This is appropriate for naturally occurring behaviours in their usual context. It involves observing people as they engage in activities in their own environment that would probably occur in much the same way if you were not present. Observing what is actually done (what they do, with who, when) and understanding the physical, social and economic context within the behaviour occurs.

Three steps are involved in conducting an observation:

- ✓ **Plan for the observation**

  To ensure that one stays focused during the observations you must thoroughly understand the study and its objectives. Explanation of the purpose of the observation should be clearly done. In doing so there is need to be discrete enough not to disrupt normal activity. For example, it would be incorrect to say 'I would like to see how you bathe your baby' rather 'I would like to understand more how people in this community care for their baby'.

  Plan how you will take notes and be unobtrusive during the observation. The purpose is to observe in a natural condition. You must make sure you know how to operate all equipment and practice until you are comfortable and confident, rehearsal is helpful.

If you are recording the observation, you must make sure the camera is fully charged and has enough recording memory.

- ✓ **Recruitment of participants and building rapport**

  Participants should be recruited in conformity with the objectives of the study. Determine the appropriate time for the observation that is convenient to the participants, especially when the practice you are interested in observing will be performed. The participants need to feel comfortable in your presence. To facilitate this, you should be non-judgmental, open and friendly. Spend some time in the environment before commencing the observation and engage in friendly conversation. Display how the instrument works, if possible, do a test run. It is usually desirable to record more than one observation with the same group/individual as the case may be.

- ✓ **Conducting Observation in an Unobtrusive manner**

  The most important principle of an observation is to be discrete and to not get in the way. Do not affect the natural flow of the activity. Position yourself in such a way that is unobtrusive but not too far away so that you can concentrate on the activity. Take notes discretely. It is useful to have a prepared guideline for the observation in line with the objectives of the study. Ask questions to clarify what happened and why after the observation. Be as objective as possible in making your note, filter out personal biases.

The strengths and limitations of an observation are as follows:

Strengths

- Can help understand what really goes on: What people say they do and what they actually do is often different
- Gives a better understanding of the whole picture: interaction, process, timing, context

- Uncovers unknown issues: In interviews we do not always ask the right questions
- Provides first-hand experience of the activity

Weaknesses

- The presence of the observer can affect the behaviour
- Time consuming and usually relies on very small samples
- Observer may be subjective and observes with their own biases

2. Narratives

This is appropriate for personal histories and experiences. They refer to events that occurred to a person over time and show how one event influences another. In addition, they allow us to examine the whole picture in order to provide a complete understanding of the situation and they are usually collected using a semi- structured guide.

3. In-depth/semi structured interview

This is Appropriate for personal perspectives and experiences. They gain a picture of the participants perspective and feelings and gives an insight into how they interpret the world. They put the respondent in the position of the expert and the in the position of student. In order to obtain relevant, truth and detailed information the respondent must; understand the purpose of the interview, be interested in the interview, feel comfortable and able to open up, be encouraged to talk and elaborate but not deviate on the topic. Respondents must understand the purpose of the interview. To ensure this you must explain at the beginning that you want to know their thoughts and opinions in their own words, that there are no correct answer to the questions and that you want to learn from them and mean it. Ask or encourage them to interrupt during the interview with anything they think is important. In selecting the respondents make sure you select those who have information to give and are willing to give it. You need to be enthusiastic, humorous and show keen interest. In addition, show your commitment be arriving on time at the venue, be familiar with the

material/documents. Keep things that may distract the respondents out of his view and keep promises. Observe non-verbal responses, be aware of your tone of voice, facial expression and body language and those of the participants.

Some tips for proper body language include: sitting squarely facing the person you are listening to, leaning slightly to demonstrate interest in what they are saying, maintaining a relaxed and open position to show you are at ease with them-arms should not be crossed, nodding the head, saying hmmmm, or repeating what you heard e.g. 'so you said you ran after him'. Respondents must feel comfortable and able to open up. This can be achieved by assuring them of confidentiality and anonymity, explaining the procedure that will be undertaking, e.g. note taking, recording of the interview), choosing a neutral location which is sufficiently quiet devoid of any disturbance, being unthreatening, assuming a friendly position, non-judgmental, wearing of appropriate clothing that is culturally acceptable, appropriate greetings, use of words that give the respondent 'permission' to talk. Do not the afraid of silence as respondents need time to reflect, gather their thoughts and prepare to say something. Try not to comment on everything that is said and avoid question and answer session, discourage monosyllable responses by probing when necessary. Remember that the list of questions is only a guide, there may be need to divert or ask questions in a different way to enable the respondent have better understanding of what is required of him. Allow natural flow from one interview to flow to the next. Try to carry out on every interview by asking yourself the following questions:

- Does this response confirm something I already know?
- If not, am I understanding the response? Exactly how does it differ?
- Can I probe to see why there is a difference in the response?
- In this case you are much more like the investigator than the data collector.

It is important that the interview unobtrusively control the rhythm of the interview, control the time allocated to each topic and subtly move the discussion from topic to

topic, interrupt gracefully and sympathize where necessary and reorienting the discussion when it goes off track by saying ‘interesting point but how about----------‘

4. Focus groups

This is appropriate for understanding cultural norms or an overview of issues among a group. It involves a group of 6-10 people guided by a facilitator during which group members talk freely and spontaneously about a certain topic. It is a method where participants influence each other through their presence and by reacting to what others say. It is not a question and answer interview, rather a forum where group members discuss topics among themselves. The steps involved in the process include:

- ✓ Recruitment of participants
- ✓ Arrangement of the venue to encourage interaction
- ✓ Conduction of the session
- ✓ Writing up and analysing the results

Recruitment of the participants

Between 6 and 10 persons are selected per group who should have similar background such as age, sex, socioeconomic status, etc., as this will facilitate discussion. ‘Key informants’ in the community are usually relied upon to make the selection who should be informed of the purpose of the study and familiar with the process of FGD. They should select talkative people with a range of views, invite the participants a day or two in advance to prevent disappointment and explain the general purpose of the group to potential participants.

Arrangement of the Venue

The venue should be arranged in a way to encourage interaction. It should be neutral, sufficiently quiet, and easy to get to, not too hot and where there will be no

disturbance. The chairs should be arranged in a circle, separate friends to avoid side conversations.

Conduction of the Session

Two people are needed to conduct the session; the facilitator or moderator and the note /recorder. The moderator/ facilitator should preferably be similar to the participants e.g., same sex, roughly in the same age group.

The Moderator/ Facilitator

The moderator/ facilitator introduces the session and introduces themselves and the recorder/note taker. Asks participants to introduce themselves and tries to put them at ease. Informs them the purpose of the FGD and how the information will be used. He assures them of confidentiality, asks and obtain their permission to use recorder and take notes. He encourages discussion by making the participants to be enthusiastic, lively, humorous and show interest.

Punctuality is important as you do not want to keep participants waiting, so arrive on time and be familiar with the material so that the session will have a natural flow. All conditions stated earlier under the narrative should be followed.

It is important for the facilitator to avoid a question and answer session by asking for clarification 'can you tell me more about--------'. Does not comment on everything that is said, if there is a pause, waits to see what happens. He should limit his participation when discussions begin. It is also useful to use one person's response to involve another person 'Amina said……. But how about you? Does anyone have a different experience? By this the discussions are expanded and those who may not be participating are encouraged to get involved. Ideas should be linked with similar ones for clarity. Finally, the facilitator gives time for general discussions 'are there any other issues….' This gives room for bringing in what might have been missed for the guide questions which were only guides in the first place, thereby enriching the information. It may be necessary to rephrase the question if the respondents are not

responding well to it. Also, it may make sense to change the order of the question in the direction of discussion by participants. Explore interesting and relevant issue if they come up.

Unobtrusively control the rhythm of the group by subtly controlling the time allocated to each topic of interest, smoothly move from topic to topic. If participants change topic allow them to continue in case useful information emerges, them summarize the main points and reorient the discussion. Discourage dominant participants by avoiding eye contact, moving away slightly, thanking the participant and changing the topic. Encourage reluctant participants by using their name requesting their opinion, making more frequent eye contact and thanking them when they talk.

The facilitator takes time at the end of the meeting to summarize, and thank the participants and let them know their ideas were useful. He also listens for additional comments and discussions that occur after the meeting including during the refreshment sessions.

The Note taker/Recorder

He keeps a record of the date, time, place, names and characteristics of participants and description of group dynamics. It is useful to map the respondents in the sitting arrangement. The recorder should find a good place for the tape recorder before the focus group and do a short test recording. There should be enough supplies and extras. It is important to listen to the tape immediately after the interview. If it malfunctioned sit with the facilitator and add more detail to the field note. In addition, it can help resolve conflict situations that facilitator finds difficult to handle.

Period for the Session

FGD session should be between 60 to 90 minutes. If the session is less than 60 minutes, it may be difficult to fully explore the discussion topic or subject matter. On the other hand, if the time exceeds 90 minutes the discussion can become

unproductive as participants may lose focus of the topic or get unnecessary carried away.

Writing up of the proceeding

After the session the facilitator, recorder and supervisor meet to evaluate how the focus group went and to discuss key findings, unexpected findings, and with other FGDs. As soon as possible transcribe the FGD in full.

Strength and Weaknesses of FGD

It helps us to understand social norms, group opinion and te range of perspectives and allows us to collect a large amount of information in a short time.

However, it only measures social norms, not good for sensitive issues and when in-depth experience is required.

4.0 CONCLUSION

The two basic approaches used in in public health research can be categorized as qualitative and quantitative methods. The objective of quantitative method is to collect data on phenomena that are suspected to be related to the causation or influence the occurrence of disease or health related event in human population. Qualitative method is useful when there is need to explore and explain behaviours rather than describes. Qualitative methods give participants the opportunity to respond in their own words, rather than forcing them to choose from fixed responses which makes the responses to be more meaningful, culturally relevant and salient to the participant.

5.0 SUMMARY

In this unit, you have been introduced to the distinction between quantitative and qualitative research methods. You have also learned about the various types of

qualitative research method and taught about you the various steps/details of how to conduct observation, in-depth/semi structured interview and focus group discussion.

6.0 TUTOR-MARKED ASSIGNMENT

1. Differentiate between qualitative and quantitative research methods
2. Explain the types of qualitative research methods
3. Explain the steps involve in conducting focus groups discussion

7.0 REFERENCES/FURTHER READING

Adindu, A. (2011). *Qualitative health research*. Calabar, Nigeria: DERHES Publications.

De Negri B. & Thomas, E. (2003). "*Making sense of Focus Group Findings: A systematic Participatory Analysis Approach"*. Washington, DC: Academy for Education Development.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

UNIT 3 DATA ANALYSIS METHODS IN PUBLIC HEALTH RESEARCH

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
    3.1 Statistical Methods for Experimental Studies
    3.2 Statistical Methods for Cohort (Prospective) Designed Studies
    3.3 Statistical Methods for Case-control (Retrospective) Studies
    3.4 Statistical Methods for Cross-sectional Designed Studies
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

The types of statistical techniques used in public health research are determined by the study design and the nature of the data collected. In this unit, you will learn about the common types of statistical methods used to analyse and present public health data. You will also be introduced to the principled of statistical inference and parametric statistics.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. explain the concept of statistical inference
2. identify the common parametric statistics and their non-parametric equivalents used for analyzing experimental research design

3. give some examples of research situations where the risk or odd ratio is calculated
4. explain and interpret results cohort, case-control and experimental studies

3.0 MAIN CONTENT

3.1 Statistical Methods for Experimental Studies

According to the WHO Guide to Health Research (2001), an experimental design is the best study design to prove causation. Testing of hypotheses is best done by experiment, where all the factors other than those under consideration can be controlled. However, statistical inference is needed to aid the generalization of the hypothesis tested in an experimental designed research.

Principles of Statistical Inference

One of the goals of public health is to solve a health problem in the population or identify a phenomenon of interest about a population. In order to do this, researchers would take a sample and hope to generate their findings, first to the study population and ultimately to the target population. This process of generalizing from a sample to a population is called statistical inference or inferential statistics. There are two broad statistical methods used to make statistical inference; the parametric and non-parametric statistics.

Parametric Statistics

Hoffman (2015) described *Parametric statistics* as mathematical formulas that are used to test hypotheses on the basis of three assumptions:

1. The sample must be derived from a population with normal distribution. That is, the data must be derived from a population in which the characteristic to be studied is distributed normally (appearing as a bell shape or normal curve).
2. The variances within the groups to be studied must be homogeneous. Homogeneity is displayed by the scores in one group having approximately the same degree of variability as the scores in another group.

3. The data must be measured at the interval or ratio level and not at the ordinal or nominal level.

According to DePoy and Gitlin (2016), parametric statistics exist in different forms and can be used to test the extent to which numerous sample structures are reflected in the population. For example, some parametric statistics can be used to test differences between only two groups, whereas others can be used to test differences among many groups. Some parametric statistics can be used to test main effects (i.e., the direct effect of one variable on another), whereas other parametric statistics have the capacity to test both main and interactive effects (i.e., the combined effects that several variables have on another variable). Furthermore, some parametric statistical action processes test group differences only one time, whereas others test differences over time. Most researchers attempt to use parametric tests when possible because they are the most robust of the inferential statistics. This mean that parametric statistics are the most likely statistics to detect a significant effect or increase power and decrease the chance of Type II errors.

Some examples of commonly used parametric statistics in public health research include the t-test, one-way analysis of variance (ANOVA), Pearson's Product–Moment Correlation Coefficient, and linear regression.

i. t-Test

The t-test is the most basic and common type of parametric statistics. It is used to compare two sample means on one variable. That is, it can be used only when the means of two groups are compared. Consider the following example where the effect of a new antihypertensive drug was tested in a randomized control clinical trial of 100 hypertension patients attending a Nigerian teaching hospital. The patients were randomly assigned into two groups (experimental and control) of 50 patients each. The patients in experimental group were treated for 12 weeks on the new antihypertensive medication, while the patients in the control group received a placebo treatment for the same period of time. The blood pressure of the participants in both groups was recorded before the commencement of the trial (pre-test) and at the end of

the trial at the 12th week (post-test). The t-test provides the best statistical method to answer this research question.

DePoy and Gitlin (2016) in their book emphasized that the t-tests must be calculated only with interval-level or ratio data and should be selected only if the researcher believes that the assumptions for the use of parametric statistics have not been violated. The t-test yields a *t* value that is reported as "$t = x$, $p = 0.05$"; *x* is the calculated *t* value, and *p* is the level of significance set by the researcher.

As described by Hoffman (2015) and DePoy and Gitlin (2016), there are two types of t-tests. One type is for independent or uncorrelated data; often called the independent t-test. The other type is for dependent or correlated data; often called the paired t-est. To understand the difference between these two types of t-tests, let return to the example of the two-group randomized design to test an experimental intervention of a new antihypertensive drug. One of the first statistical tests to be performed is to determine whether the blood pressure of the experimental and control group patients is different at pre-test (before the commencement of the trial). The pre-test blood pressure data of experimental and control group patients reflect two independent samples. Therefore, the *independent t-test* for independent samples could be used to compare the difference between these two groups. However, if we want to compare the pre-test blood pressure scores to post-test blood pressure scores for only the experimental patients, in this case, we would compare scores from the same subjects at two points in time. The scores are likely to be more similar because they are drawn from the same group and are apt to be highly correlated. Therefore, the *paired t-test* for dependent data will be used, which considers the correlated nature of the data.

However, parametric tests can lose efficiency when the distributions are severely non-normal because of skewing, outliers, or grossly unequal variances. In such situation, they can be replaced by several robust tests that are referred to as distribution free or non-parametric tests. The main nonparametric test to replace the independent t-test is the Mann-Whitney U test, and the major replacement for the paired t-test is the

Wilcoxon-signed rank test. However, for studies with more than two groups, the researcher must select other statistical procedures.

ii. One-Way Analysis of Variance

The "one-way" ANOVA, or "single-factor" ANOVA, serves the same purpose as the t-test. It is designed to compare sample group means to determine whether a significant difference can be inferred in the population. However, one-way ANOVA, also referred to as the "F-test," can manage two or more groups. It is an extension of the t-test for a two-or-more-groups situation. The null hypothesis for an ANOVA, as in the t-test, states that there is no difference between the means of two or more populations (Hoffman, 2015; DePoy and Gitlin 2016).

The procedure is also similar to the t-test. The original raw data are put into a formula to obtain a calculated value. Computing the one-way ANOVA yields an *F* value that may be reported as "$F(a,b) = x$, $p = 0.05$"; x is computed *F* value, a is group degrees of freedom, b is sample degrees of freedom, and p is level of significance. "Degrees of freedom" refers to the "number of values, which are free to vary" in a data set (Hoffman, 2015; DePoy and Gitlin 2016).

There are many variations of ANOVA. Some test relationships when variables have multiple levels (two-way ANOVA), some test related group (i.e., same group of participants), not independent one, over and over again (Repeated Measures ANOVA), and some examine complex relationships among multiple levels of variables (MANOVA). The main non-parametric test to replace the one-way ANOVA is the Kruskal-Wallis one-way ANOVA, while the Friedman Test is non-parametric equivalence of the Repeated Measures ANOVA.

iii. Pearson's Product–Moment Correlation Coefficient

This is a very commonly used parametric statistic that is used to measure the linear relationship between two continuous variables (ordinal or ratio data). It is measured by a statistic called the correlation coefficient and is denoted as r for a sample statistic.

As described by Mukaka (2012), the correlation coefficient is a dimensionless quantity that takes a value in the range −1 to +1. A correlation coefficient of zero indicates that no linear relationship exists between two continuous variables, and a correlation coefficient of −1 or +1 indicates a perfect linear relationship. The strength of relationship can be anywhere between −1 and +1. The stronger the correlation, the closer the correlation coefficient comes to ±1. If the coefficient is a positive number, the variables are directly related (i.e., as the values of one variable tend to increase, the values of the other also do so). If, on the other hand, the coefficient is a negative number, the variables are inversely related (i.e., as the values of one variable tend to increase, the values of the other tend to decrease). However, when the variables being studied are not normally distributed, the spearman correlation coefficient is used as the non-parametric equivalent of the Pearson's Product–Moment Correlation Coefficient. According to Mukaka (2012), the rule of thumb for interpreting the size of correlation coefficients is shown below:

Size of correlation	Interpretation
0.90 to 1.00 (–0.90 to –1.00)	Very high positive (negative) correlation
0.70 to 0.90 (–0.70 to – 0.90)	High positive (negative) correlation
0.50 to 0.70 (– 0.50 to –0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (– 0.30 to –0.50)	Low positive (negative) correlation)
0.00 to 0.30 (–0.00 to –0.30)	Negligible correlation

iv. Linear Regression

Linear regression is similar to correlation because it is also used to measure the extent to which there is a linear relationship between two variables. A main difference between the two is that linear regression makes a distinction between independent and dependent variables but correlation makes no such distinction. In particular, the main purpose of linear regression is to “predict’ the value of the dependent variable based on the values of the independent variables. In addition to the other main assumptions for parametric statistics, for linear regression to be conducted the dependent variable

must be a continuous data (interval or ratio scale) and the independent variable(s) can be measured on either a categorical or continuous measurement scale. There are two types of linear regression, simple linear regression and multiple linear regression. In both types, there is only one single dependent variable. The different between the two is the number of independent variables. In simple linear regression, a single independent variable is used to predict the value of a dependent variable, while in multiple linear regression two or more independent variables are used to predict the value of a dependent variable.

The first step to investigate the relationship between two continuous variables is by drawing a scatter diagram. The dependent variable is plotted on the Y axis while the independent variable is plotted on the X axis. The linear regression can then be used to capture the linearity between the dependent variable and independent variable by investigating the properties of the straight line on the scatter diagram. This can be presented in a model (equation). This model can be written in a more general form in terms of two variables *Y* and *X* as:

$Y = b_o + b_1 X$

Y is the dependent variable and X is the independent variable. The term b_o is known as the constant coefficient of intersection (it is where the line cuts the *Y* axis). The term b_1 is known as the slope coefficient, and will be positive if the line slopes upward from left to right, and negative if the line slopes down from left to right. The values of b_1 indicate the amount of by which the dependent variable would change if the value of independent variable increase by 1 unit.

The non-parametric equivalent of linear regression is logistic regression. This can either be binary logistic regression or multivariate logistic regression.

3.2 Statistical Methods for Cohort (Prospective) Studies

The commonly used statistical tests for the analysis and presentation of data in cohort studies have been highlighted in the section on study design to include (1) direct estimate of risk (incidence), (2) relative risk, and (3) attributable risk. However, it is important to understand that the basic measure of risk in in public health is the probability of a disease or any outcome of interest. According to the WHO (2001), two measures of probability are commonly used in public health; the prevalence and incidence. While prevalence measures the probability of having a disease, incidence measures the probability of having a disease.

These can be expressed formally as:

Point prevalence = number of people with the disease (outcome) in a population at a specific point in time / total population at risk at that time.

Period prevalence = number of people with the disease in a population during a specific period / total population at risk during that period.

In both cases, the numerator is the number of existing cases. However, the prevalence is the measure that is used for analysis in a cross-sectional study, while incidence is the measure of analysis in cohort study.

Incidence has the number of new cases in the numerator. There are two ways of measuring incidence, depending on what denominator is used: the cumulative incidence and the incidence density. Both provide estimates of probabilities of acquiring the disease, but the unit of measurement is different in the two methods. Cumulative incidence estimates the probability of acquiring the disease per person, and the incidence density is the estimate of probability of acquiring the disease per person-time.

When the probabilities of disease in two groups are compared, as in the case of a cohort study, where the probability of disease among the exposed group is compared

with the probability of disease among the unexposed, a relative measure is used. The ratio of the two probabilities is called the relative risk (RR).

RR = incidence among exposed / incidence among the unexposed

Either of the two measures of incidence may be used. For example, suppose a cohort study of 400 smokers and 600 non-smokers documented the incidence of hypertension over a period of 10 years. The following table summarizes the data at the end of the study.

	Hypertension		
Smoking	Yes	No	Total
Yes	120	280	400
No	30	570	600
Total	*	*	1000

Source: WHO, 2001

The probability of hypertension among smokers, P(H|S), also denoted as I_e (incidence among exposed) = 120/400 = 0.30.

The probability of hypertension among non-smokers, P(H|nonS), also denoted as I_o (incidence among unexposed) = 30/600 = 0.05.

RR = I_e / I_o = 0.3/0.05 = 6.0.

An RR of more than 1 indicates the factor to be positively associated with the disease (exposure increases the chance of the disease) and an RR of less than 1 indicates a protective factor (exposure decreases the chance of disease).

Another measure that is commonly derived from the probabilities of disease in the two groups, is the attributable risk (AR): the excess risk for the exposed group compared with the unexposed group. This is simply the difference between the two probabilities:

$AR = I_e - I_o = 0.30 - 0.05 = 0.25$

Twenty-five per cent of the new cases of hypertension among the exposed group can be attributed to smoking.

3.3 Statistical Methods for Case-control (retrospective) Studies

In case-control studies, clearly the incidence is not measurable, and hence the relative risk is not estimable. To calculate the risk that those who had the outcome had the risk, you need to know two things: the total number of those who have had the condition (e.g., hypertension), and the number of these who had been exposed to the risk (e.g., smoking). You then divide the latter by the former. In a cohort study, the groups would normally be selected on this basis- whether they have been exposed to the risk or not. So, one group would contain individuals exposed to the risk and the other those not exposed.

But in a case-control study, the groups are not selected on the basis of whether people have been exposed to the risk or not, but on the basis of whether they have some condition (e.g., hypertension) or not. So, you have one group composed of individuals who have had hypertension, and one group without hypertension, but both groups would contain individuals who were and were not exposed to the risk (smoking). Moreover, the number of cases and controls may not be equal in a case-control study. For example, the number of controls can double the number of control and vice-versa. This means that the column totals, which would otherwise be needed for risk calculation, are meaningless.

In case-control studies, a measure related to probability constitutes the ‘odds’ of an event, and a good approximation of the relative risk is the odds ratio (OR).

Researchers can compare the odds that those with a condition (e.g., hypertension) will have been exposed to the risk factor (e.g., smoking), with the odds that those who do not have the condition will have been exposed. The odds ratio is the division of the former by the latter. Thus, in a case-control study, the odds ratio (OR) is used as a measure of association of the disease and the risk factor.

In the example described for cohort study, if the same results were obtained from a case-control study of 150 cases of hypertension and 850 people without hypertension, the table would appear as follows:

	Hypertension		
Smoking	Yes	No	Total
Yes	(a) 120	(b) 280	(a + b)
No	(c) 30	(d) 570	(c + d)
Total	(a + c) 150	(b + d) 850	1000

Source: WHO, 2001

OR = 120x570 / 30x280 = 8.14

The odds ratio calculation can be generalized with the help of the 2 x 2 table as above.

- The odds of exposure to the risk factors among those with the condition = a / c
- The odds of exposure to the risk among those without the condition = b / d
- Therefore; odds ratio = a/c ÷ b/d= ad / bc

In the example given above, the odds that those with a hypertension had smoked = 120/30 = 4.000; and the odds that those without hypertension had smoked = 280/570 = 0.491. Dividing the former by the latter, you get the odds ratio = 4.000/0.491 = 8.14.

This result suggests that those with hypertension are more than 8 times likely to have smoked compared to those without hypertension.

3.4 Statistical Methods for Cross-sectional Studies

In the case of cross-sectional studies, the population sampled is the total population. Therefore, both the prevalence of disease and the prevalence of the risk factor can be estimated. Here, all the elements of the 2x2 table are valid measurements, and allow for the calculation of the appropriate probabilities. Note, however, that the probabilities are not 'risk of acquiring the disease', but rather the prevalence measure.

All the measures stated above can be computed from the 2x2 table. The RR and OR would be calculated in the same way, and other quantities such as the AR can also be calculated. If the prevalence and incidence are similar, these measures may have the same interpretations. More importantly, testing of hypotheses regarding the various probabilities would be valid in this type of design, and would provide the basis for further refinement of the risk estimates in studies with better designs (cohort, quasi-experimental or experimental).

In the table of observations, all the cells would now have valid numbers. The above table, if it had arisen from a cross-sectional study, would appear as:

	Hypertension		
Smoking	Yes	No	Total
Yes	120	280	400
No	30	570	600
Total	150	850	1000

Source: WHO, 2001

Statistical Variations in the Measure – Confidence Interval Estimation

Notice that all the above measures (RR, AR, OR) are point estimates of the appropriate statistical methods of association for the various study designs. Since public health studies are usually based on samples, it is important to identify the 'random error' associated with these estimates, i.e. what is the possible range of values within which the true measure lies. One needs to develop the probability distribution of these measures, and from this, calculate an appropriate confidence interval. But to get a confidence interval it is important to introduce an important concept in statistical inference – the 'standard error'. Suffice it to say that we can usually calculate a 'standard error' of the estimates and, using this, obtain the confidence intervals using the normal approximation (i.e. estimate +/- 2 standard error would give approximately a 95% confidence interval for the risk measure). The standard error can easily be estimated with the equation: $s.e.(\bar{x}) = s/\sqrt{n}$. Here s is the sample standard deviation and is n the sample size.

For example, if we took a sample size n = 100 from a population, and measured systolic blood pressure, and obtained a sample mean of 135 mmHg and a sample standard deviation of 3 mmHg, then the estimated standard error would be:

$s.e.(\bar{x}) = 3/\sqrt{100}$ = 3/10 = 0.33 mmHg

With the standard error, it is easy to calculate the confidence interval for the population mean because we can be 95 percent confident that any sample mean is going to be within plus or minus two standard errors of the population mean.

Population mean = sample mean ± 2 x standard error

Therefore, the 95% confidence interval for the population mean of systolic blood pressure is:

(135 – 2 x 0.33 to 135 + 2 x 0.33) or (134.34 to 135.66)

This can be interpreted as follow: we can be 95% confident that the population mean of blood pressure is between 134.34 to 135.66.

Similar to the population mean, the standard error of the sample proportion can be estimated by the following equation:

$$s.e. = (p)\sqrt{p\ (1-p)/\ n}$$

Where p is the sample proportion, and n is the sample size. The 95 percent confidence interval for the population proportion is equal to the sample proportion plus or minus 1.96 standard errors.

{[p – 1.96 x s.e (p)] to [p + 1.96 x s.e (p)]}

If the confidence interval for the ratio (risk or odds ratio) of two population parameters does not contain the value of 1, then you can be 95% confident that any difference in the size of the two measures is statistically significant.

4.0 CONCLUSION

In this unit, you have learnt that *Parametric statistics* are used to test hypotheses on the basis of three assumptions. It was highlighted some parametric statistics can be used to test differences, linear relationships or associations between only two groups, whereas others can be used to test differences, relationships or associations among many groups. Most researchers attempt to use parametric tests when possible because they are the most robust of the inferential statistics. Some examples of commonly used parametric statistics in public health research include the t-test, one-way analysis of variance (ANOVA), Pearson's Product–Moment Correlation Coefficient, and linear regression. This unit also emphasized that the commonly used statistical tests for the analysis and presentation of data in cohort studies include (1) direct estimate of risk (incidence), (2) relative risk, and (3) attributable risk.

5.0 SUMMARY

In this unit, you have been introduced to the concept of statistical inference in experimental studies. You have learnt about the types and assumptions involved in parametric statistics and when to choose non-parametric statistic equivalence in public health research. You have also learnt about the various statistical methods used to analyze and present public health research, and how to choose which statistical methods for different research design.

6.0 TUTOR-MARKED ASSIGNMENT

1. Give a concise definition of statistical inference
2. List the assumptions that must be met before parametric statistics can be used in data analysis
3. Describe the difference between independent t-test and paired t-test
4. Identify the data situation in which the One-Way Analysis of Variance can be used
5. List the parametric statistics and their non-parametric equivalent that can be used to explore relationships and associations between variables
6. Give some examples of research situation where the risk ratios and odd ratios are used for statistical analysis

7.0 REFERENCES/FURTHER READING

Bowers, D. (2008). *Medical Statistics from scratch: an introduction for health professionals*. London: John Wiley & Sons Ltd.

Chin, R., & Lee, B.Y. (2008). *Principles and Practice of Clinical Trial Medicine*. Elsevier

DePoy, E., & Gitlin, L.N. (2016). *Introduction to Research: Understanding and Applying Multiple Strategies*. (5th ed). Mosby.

Hoffman, J.I.E. (2015). *Biostatistics for Medical and Biomedical Practitioners*. Elsevier.

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology.* (3rd ed.). Oxford University Press

Merril, R.M. (2008). *Environmental Epidemiology: Principles and methods*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

Mukaka, M.M. (2012). "A guide to appropriate use of Correlation coefficient in medical research." *Malawi Medical Journal*, 24(3): 69–71

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Smith, R.V. (1984). *Graduate Research: A guide for students in the sciences*. Michigan: ISI Press

MODULE 3 TOPIC SELECTION AND COMPONENTS OF RESEARCH PROPOSAL IN PUBLIC HEALTH

Unit 1 Construction and Formulation of Research Topics
Unit 2 Components of a Research Proposal
Unit 3 Beyond Research Proposal

UNIT 1 CONSTRUCTION AND FORMULATION OF RESEARCH TOPICS

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
  3.1 Construction of Research Topics in Public Health
  3.2 Steps in the Formulation of Research Topics
4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will learn the process of construction of a good research topic. You will also learn about the various steps in the formulation of research topics in public health.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. explain how to construct a research topic
2. describe the various steps in the formulation of a research topic

3.0 MAIN CONTENT

3.1 Construction of Research Topics

The first step in any research process is to think of a topic for the research. However, choosing the right research topic can be somewhat challenging. A good research topic must be narrowed and focused enough to be interesting, but broad enough to capture adequate information. It is important for the student to select a good public health topic because this will set you up in a good direction, keep you interested in the project, and lead to a successful research project.

3.2 Steps in the Formulation of Research Topics

According to Wang and Park (2016), students are almost confronted with four basic questions when selecting topics for their research project. These questions as fully described by Wang and Parker (2016) are presented below:

1. Where can I start to find a good topic?

   A good topic is a topic that you are interested in. This will include the public health issue or problem you are interested in and for which you will like to find solution. You can begin by forming a list of themes or questions you want to know more about. The goal is to crystallize and clarify your ideas about the issues and problems you are interested about in, and develop them into a

research topic. If your ideas are vague or difficult to put into a research topic, the following strategies or resources for inspiration are relevant to help you crystallize your ideas or formulate them into concrete research topics:

a. Talking to people (lecturers, colleagues, etc)
b. Searching the internet
c. Consulting reference books, statistics, and other library resources
d. Reading scholarly journal articles
e. Reading current events and recent policy debates

2. How do I narrow down my topic?

   It is important that topics should be narrow enough for the student to investigate within the given time (e.g., academic sessions) and broad enough to satisfy the requirements for a master's degree programme in Public Health. Neuman (2011) suggested the followings as way to narrow down a research topic idea:

   a. Replicate a previous research project exactly or with slight variation
   b. Explore unexpected findings discovered in previous research
   c. Follow suggestions an author gives for future research at the end of an article
   d. Extend an existing explanation or theory to a new topic or setting
   e. Challenge findings or attempt to refute a relationship
   f. Specify intervening process and consider linking relations

3. What topic is appropriate for me?

   Selecting a research topic that you can handle and that you will be excited about is the most important step in the research process. Therefore, you should spend sufficient time collecting background information, discussing your ideas with others, and writing down your thoughts. You may consider the following factors when considering what topic to find appropriate:

   a. A topic for which you are excited about
   b. A topic for which you are prepared

c. A topic related to your experience or employment
d. A feasible topic: those that can be done with the available resources to you. A research topic is feasible if you can answer the following question in the affirmative
    - Do I have access to the study population?
    - Can I draw a robust sample of this population?
    - Can I complete my research within the given time frame?
    - Do I have the financial resources to carry out the field work of the research?
    - Do I have sufficient skills and knowledge to complete this project?
e. A topic you can build upon
f. A topic with a broader audience
g. A topic similar to your Professor or Supervisor's research

4. How do I know the topic I selected is a "good topic"?

   To know if the research topic you have selected is good, you must ask yourself the following questions to evaluate your topic:

   a. Am I really interested in the topic?
   b. Am I familiar with the topic? If not, am I prepared to do extra research on it?
   c. Is there adequate information available to me to research this topic?
   d. Will I have access to my study population? Can I get permission to conduct this research?
   e. Does this topic offer future opportunities for research?
   f. Do I have personal experience related to this research?
   g. Will my future employment benefit from this research?
   h. Are there professors or tutors in my department who have expertise in this area? Will they be able to help me?

4.0 CONCLUSION

In this unit, we have described that a good research topic must be narrowed and focused enough to be interesting, but broad enough to capture adequate information. It is important for the student to select a good public health topic because this will set you up in a good direction, keep you interested in the project, and lead to a successful research project. This unit has also described that the four basic questions for students to fully consider when selecting research topics for their research project are: (1) Where can I start to find a good topic? (2) How do I narrow down my topic? (3) What topic is appropriate for me? And (3) How do I know the topic I selected is a "good topic"?

5.0 SUMMARY

In this unit, you have been introduced to how to select a research topic. You have also learnt about the various process involved in the formulation of a good research topic.

6.0 TUTOR-MARKED ASSIGNMENT

1. Describe the process in the formulation of a research topic
2. Highlight the important questions that you need to answer when selecting topics for your research project

7.0 REFERENCES/FURTHER READING

Al-Riyami, A. (2008). How to prepare a research proposal. *Oman Medical Journal*, 23(2):66-69

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology*. (3rd ed.). Oxford University Press

Pajares, F. (2207). Elements of proposal 2007. From http://www.des.emory.edu/mfp/proposal.html. Accessed January 18, 2020.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Wang, G.T., & Park, J. (2016). *Student Research and Report Writing: From topic selection to the complete paper*. John Wiley & Sons ltd; Malden, USA

Wong P. How to write research proposal. International network on Personal meaning. Available at www.meaning.ca/archives. Accessed January 18, 2020.

UNIT 2 COMPONENTS OF A RESEARCH PROPOSAL

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
- 3.1 Chapter One of the Research Proposal
- 3.2 Chapter Two of the Research Proposal
- 3.3 Chapter Three of the Research Proposal

4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment
7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will learn how to identify and describe the basic components of research proposal, including how to write the introduction, research objectives, statement of problem, justification of study, literature review and the methodology.

You will also review a prototype template for writing research proposal and a Gantt Chart that can be adopted when preparing your research proposal for the MPH degree.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. describe the various components of research proposal
2. understand how to write a research proposal for your project

3.0 MAIN CONTENT

3.1 Chapter One of the Research Proposal

Al-Riyami (2008) and Pajares (2007) in their review articles highlighted and summarized the elements of research proposal as fully described and reproduced below:

1. Title: It should be concise and descriptive. It must be informative and catchy. An effective title not only pricks the readers interest, but also predisposes him/her favourably towards the proposal. Often titles are stated in terms of a functional relationship, because such titles clearly indicate the independent and dependent variables. The title may need to be revised after completion of writing of the protocol to reflect more closely the sense of the study.

2. Abstract: It is a brief summary of approximately 300 words. It should include the main research question, the rationale for the study, the hypothesis (if any) and the method. Descriptions of the method may include the design, procedures, the sample and any instruments that will be used. It should stand on its own, and not refer the reader to points in the project description.

3. Introduction: The introduction provides the readers with the background information. Its purpose is to establish a framework for the research, so that readers can understand how it relates to other research. It should answer the question of why the research needs to be done and what will be its relevance. It puts the proposal in context.

4. Statement of the Problem: This describes the context for the study and should be stated in precise and clear terms. A problem might be defined as the issue that exists in the literature, theory, or practice that leads to a need for the study. A good problem statement answers the question "Why does this research need to be conducted". If a researcher is unable to answer this clearly and succinctly, then the statement of the problem will come off as ambiguous and diffuse. The statement of the problem is important because (1) it is the essential basis for the construction of a research proposal (research objectives, hypotheses, methodology, work plan and budget etc.), (2) it is an integral part of selecting a research topic, (3) it will guide and put into sharper focus the research design being considered for solving the problem, and (4) it allows the investigator to describe the problem systematically, to reflect on its importance, and to point out why the proposed research on the problem should be undertaken.

5. Objectives of the Study: Research objectives are the goals to be achieved by conducting the research. They may be stated as 'general or broad' and 'specific'. The general or broad objective of the research is what is to be accomplished by the research project, for example, to determine whether or not a new vaccine should be incorporated in a public health program. The specific objectives relate to the specific research questions the investigator wants to answer through the proposed study and may be presented as primary and secondary objectives, for example, primary: To determine the degree of protection that is attributable to the new vaccine in a study population by comparing the vaccinated and unvaccinated groups. The key points to keep in mind when preparing the objectives of the study are:

a. Incorporate a sentence that begins with “The objective(s) of this study is/are …..”
b. Clearly identify and define the central concepts or ideas of the study.
c. Identify the specific method of inquiry to be used.
d. Identify the unit of analysis in the study.
e. Objectives must SMART; that is, Specific, Measurable, Achievable, Realistic and Timely.

6. Research Questions and/or Hypotheses: A research question poses a relationship between two or more variables but phrases the relationship as a question. A hypothesis on the other hand, represents a declarative statement of the relations between two or more variables. In other words, the hypothesis translates the problem statement into a precise, unambiguous prediction of expected outcomes. Deciding whether to use questions or hypotheses depends on factors such as the objective of the study, the nature of the design and methodology. Questions are relevant in both qualitative and quantitative research, but hypotheses are only used in quantitative research. Hypotheses are not meant to be haphazard guesses, but should reflect the depth of knowledge, imagination and experience of the investigator. In the process of formulating the hypotheses, all variables relevant to the study must be identified. For example: “Health education involving active participation by mothers will produce more positive changes in child feeding than health education based on lectures”. Here the independent variable is types of health education and the dependent variable is changes in child feeding. The practice of using hypotheses stemmed from theoretical research and have distinct advantage in statistical testing, as researchers tend to be conservative and cautious in the statements of their conclusions. In general, the research questions and hypotheses guide the researchers to interpret the outcomes and results of the study.

7. Significance of the Study: The significance of a study indicates how the research will refine, revise or extend existing knowledge in the area under investigation.

Such refinement, revisions, or extensions may have either practical, theoretical, or methodological significance. So why crafting the significance of the study, the researchers need to think about the implications of their study, that is, how results of the study may affect research in the field of study, theory, practice, educational interventions, policy, etc. When thinking about the significance of the study, the following questions need to be considered:

a. What will results mean to the field of research study?
b. What suggestions for subsequent research will arise from the findings?
c. What will results mean to practice?
d. Will results influence programs, methods, and/or interventions?
e. Will results contribute to the solutions to problems?
f. Will results influence policy decisions?
g. What will be improved or changed as a result of the proposed research?
h. How will results of the study be implemented and what innovations will come about from them?

8. Delimitations and Limitations of the Study
    a. A *delimitation* addresses how a study will be narrowed in scope, that is, how it is bounded. This is the section to explain the things that would not be done in the research and why the researcher chose not to do them – the literature that would not be reviewed (and why not), the population that would not be studied (and why not), the population that would not be studied (and why not), the methodological procedures that would not be used (and reason for not using them). The delimitation of the research should be limited to things that the readers might reasonably expect the researcher to do but that the researcher has decided not to do for clearly explained reasons.
    b. A *limitation* identifies potential weaknesses of the study. This section focuses on the things that may constitute threat to the validity of the research but that are impossible to avoid or minimize. Specifically, the researchers need to think about their analyses, the nature of their

instruments, self-report, the sample, research design etc. when writing on the limitations of their study.

3.1 Chapter Two of the Research Proposal

9. Literature Review: The review of the literature provides the background and context for the research problem. It should establish the need for the research and indicate that the researcher is familiar with and knowledgeable about the research area. In a proposal, the literature review is generally brief and to the point. The literature selected should be pertinent and relevant. Select and reference only the more appropriate citations and make key points clearly and succinctly. It is important to avoid statements that imply that little has been done in the area or that what has been done is too extensive to permit easy summary. Statements of this sort are usually taken as indications that the researcher is not really familiar with the literature. It is good to include a subsection that outline the *search strategy*—the procedures you used and sources you investigated (e.g., databases, journals, text books, experts in the field) to compile your literature review. The literature review accomplishes several important things:
    a. It shares with the reader the results of other studies that are closely related to the study being reported or envisaged.
    b. It relates a study to the larger, ongoing dialogue in the literature about a topic, filling in gaps and extending prior studies.
    c. It provides a framework for establishing the importance of the study, as well as a benchmark for comparing the results of a study with other findings.
    d. It "frames" the problem earlier identified.
    e. It helps to identify gaps that need to be filled
    f. Demonstrate to the reader that you have a comprehensive grasp of the field and are aware of important recent substantive and methodological developments.

3.3 Chapter Three of the Research Proposal

10. Methodology: The methods or procedure is the heart of the research proposal. The guiding principle for writing the Methods section is that it should contain sufficient information for the reader to determine whether the methodology is sound. A good proposal should contain sufficient details for another qualified researcher to implement the study. It is important to indicate the methodological steps that would be taken to answer every question or to test every hypothesis illustrated in the Questions/hypotheses section. It is vital to consult a biostatistician during the planning stage of the research, to resolve the methodological issues before submitting the proposal. This section should include:

    a. **Research Design:** The selection of the research strategy is the core of research design and is probably the single most important decision the investigator has to make. The choice of the strategy, whether descriptive, analytical, experimental, observational, interventional or a combination of these depend on a number of considerations, but this choice must be explained in relation to the study objectives.

    b. **Research setting**: The research setting includes all important facets of the study, such as the place and time of the study, and the population to be studied (sampling frame).

    c. **Research subjects or participants**: Depending on the type of study design, the following questions should be answered regarding the subjects or participants to be recruited into the study:
        - What are the criteria for inclusion or selection?
        - What are the criteria for exclusion?
        - What is the sampling procedure that will be used so as to ensure representativeness and reliability of the sample and to minimize sampling errors?

- Will there be use of controls in the study? Controls or comparison groups are used in scientific research in order to increase the validity of the conclusions. Control groups are necessary in all analytical epidemiological studies, in experimental studies of drug trials, in research on effects of intervention programmes and disease control measures and in many other investigations. Some descriptive studies (studies of existing data, surveys) may not require control groups.
- What are the criteria for discontinuation?

d. **Sample size**: The proposal should provide information and justification (basis on which the sample size is calculated) about sample size in the methodology section. Calculation of sample size has been made easy by computer software programmes, but the principles underlying the estimation should be well understood.

e. **Instrumentations/ Measures:** Instruments are the tools by which the data are collected. The instruments proposed to use (e.g., questionnaires, scales, interview protocols, observation grids) should be clearly outlined in the proposal. If instruments have previously been used, identify previous studies and findings related to reliability and validity of the instruments. If instruments have not previously been used, outline procedures you will follow to develop and test their reliability and validity. In the latter case, a pilot study is essential. For validated questionnaires/interview schedules, reference to published work should be given and the instrument appended to the proposal. For a new questionnaire which is being designed specifically for your study the details about preparing, precoding and pretesting of questionnaire should be furnished and the document appended to the proposal. Descriptions of other methods of observations like medical examination, laboratory tests and screening procedures is necessary- for established procedures, reference of published work cited but for new or modified procedure, an adequate description is necessary with justification

for the same. It is important to include an appendix with a copy of the instruments to be used or the interview protocol to be followed. Also include sample items in the description of the instrument. For a mailed survey, identify steps to be taken in administering and following up the survey to obtain a high response rate.

f. **Data Collection:** in this section, the general plan for collecting the data for the research should be outlined. This will include survey administration procedures, interview or observation procedures. This should also include an explicit statement covering the field controls to be employed and outline of the time schedule that is expected to be followed. A description of the protocol of data collection should be clear. For example, in a study on blood pressure measurement: time of participant arrival, rest for 5 to10 minutes, which apparatus (standard calibrated) to be used, in which room to take measurement, measurement in sitting or lying down position, how many measurements, measurement in which arm first (whether this is going to be randomized), details of cuff and its placement, who will take the measurement. This minimizes the possibility of confusion, delays and errors.

g. **Data Analysis:** This section should include the design of the analysis form, plans for processing and coding the data and the choice of the statistical method to be applied to each data. It is important to state clearly the analytic tools to be used for the data analysis (e.g., SPSS, SAS, SYSTAT, EPI Info, etc.) and the level of statistical significance to be tested.

h. **Ethical issues**: Ethical considerations apply to all types of public health research. Before the proposal is submitted to the Ethics Committee or the Institution Review Board (IRB) for approval, two important documents mentioned below (where appropriate) must be appended to the proposal. In

additions, there is another vital issue of Conflict of Interest, wherein the researchers should furnish a statement regarding the same.

1. The Informed consent form (informed decision-making): A consent form, where appropriate, must be developed and attached to the proposal. It should be written in the prospective subjects' mother tongue and in simple language which can be easily understood by the subject. The use of medical terminology should be avoided as far as possible. Special care is needed when subjects are illiterate. It should explain why the study is being done and why the subject has been asked to participate. It should describe, in sequence, what will happen in the course of the study, giving enough detail for the subject to gain a clear idea of what to expect. It should clarify whether or not the study procedures offer any benefits to the subject or to others, and explain the nature, likelihood and treatment of anticipated discomfort or adverse effects, including psychological and social risks, if any. Where relevant, a comparison with risks posed by standard drugs or treatment must be included. If the risks are unknown or a comparative risk cannot be given it should be so stated. It should indicate that the subject has the right to withdraw from the study at any time without, in any way, affecting his/her further medical care. It should assure the participant of confidentiality of the findings.

2. In the case of minors, an assent must be obtained which signifies the child's affirmative agreement to participate in the research. It is an act signifying understanding (that the minor has not reached full legal age). Mere failure to object by the child should not, absent affirmative agreement, be construed as assent. In such a case the parent/guardian signs the assent form for the child.

3. Ethics checklist: The proposal must describe the measures that will be undertaken to ensure that the proposed research is carried out in accordance with the Declaration of Helsinki on Ethical Principles for research involving Human Subjects. It must answer the following questions:
    - Is the research design adequate to provide answers to the research question? It is unethical to expose subjects to research that will have no value.
    - Is the method of selection of research subjects justified? The use of vulnerable subjects as research participants needs special justification. Vulnerable subjects include those in prison, minors and persons with mental disability. It is important to mention that the population in which the study is conducted will benefit from any potential outcome of the research and the research is not being conducted solely for the benefit of some other population. Justification is needed for any inducement, financial or otherwise, for the participants to be enrolled in the study.
    - Are the interventions justified, in terms of risk/benefit ratio? Risks are not limited to physical harm. Psychological and social risks must also be considered.
    - For observations made, have measures been taken to ensure confidentiality?

i. **Work Plan:** A work plan is an overview of tasks/proposed activities and a time frame for the entire research project. You put weeks, days or months at one side, and the tasks at the other. You draw fat lines to indicate the period the task will be performed to give a timeline for your research study. An example of a work plan for a Master of public health dissertation is shown in appendix 1 below.

j. **References:** The proposal should end with relevant references on the topic. The reference style (e.g., APA or Harvard or Vancouver) should be used consistently in the text and in the reference list. Only references cited in the text should be included in the reference list. For web-based search, include the date of access for the cited website, for example: add the sentence "accessed on January 25, 2020".

k. **Appendixes**: The need for appropriate documentation dictates the inclusion of appropriate appendixes in the proposal. The followings are appropriate appendixes to be included in the proposal: (1) interview protocols, (2) sample of informed consent forms, (3) cover letters sent to appropriate stakeholders, (4) official letters for permission to conduct research. Regarding original scales or questionnaires, if the instrument is copyrighted then permission in writing to reproduce the instrument from the copyright holder or proof of purchase of the instrument must be submitted. An example of a research proposal template that can be adapted for your study is shown in appendix 2 at the end of this module.

4.0 CONCLUSION

Research proposal in Public Health are written in three chapters. Chapter One which is the introduction entails writing the background, the statements of research problem, objectives of the study, the research questions and/or hypotheses, significance of the study and the delimitations and limitations of the proposed research. Chapter Two of the proposal which is the Literature Review provides the pertinent and relevant literature in a way that indicate that the researcher is familiar with and knowledgeable about the research area. It is good to include a subsection that outline the *search strategy*—the procedures you used and sources you investigated (e.g., databases, journals, text books, experts in the field) to compile your literature review. Chapter three of the proposal is the methodology. This section focusses on the research design, settings and participants. It also describes the instruments that will be used, the full

details of the process for data collection and the data analysis methods that will be utilized. The methodology section should also include sub-section on ethics and the work plan (Gantt chart for the project).

5.0 SUMMARY

In this unit, you have learnt how to identify and describe the basic components of research proposal, including how to write the introduction, research objectives, statement of problem, justification of study, literature review and the methodology.

6.0 TUTOR-MARKED ASSIGNMENT

1. Itemize and describe the basic components of the research proposal
2. Enumerate the most important questions when recruiting participants into your study

7.0 REFERENCES/FURTHER READING

Al-Riyami, A. (2008). How to prepare a research proposal. *Oman Medical Journal,* 23(2):66-69

Lilienfeld, D.E. & Stolley, P.D. (1994). *Foundations of Epidemiology.* (3rd ed.). Oxford University Press

Pajares, F. (2007). Elements of proposal 2007. From http://www.des.emory.edu/mfp/proposal.html. Accessed January 18, 2020.

World Health Organization. (2001). *Health Research Methodology: A guide for training in research methods.* (2nd ed.) Manilla: WHO.

Wang, G.T., & Park, J. (2016). *Student Research and Report Writing: From topic selection to the complete paper.* John Wiley & Sons ltd; Malden, USA

Wong P. How to write research proposal. International network on Personal meaning. Available at www.meaning.ca/archives. Accessed January 18, 2020.

UNIT 3 BEYOND RESEARCH PROPOSAL

CONTENTS

1.0 Introduction
2.0 Objectives
3.0 Main Content
- 3.1 Writing Results of your Dissertation
- 3.2 Writing the Discussion of your Research Findings
- 3.3 Manuscript Preparation and Publication in a Journal

4.0 Conclusion
5.0 Summary
6.0 Tutor-Marked Assignment

7.0 References/Further Reading

1.0 INTRODUCTION

In this unit, you will learn how to write the results and discussion sections of your dissertation. This unit will introduce you to the various important elements to consider when writing the discussion of your results/findings. You will also be introduced to how to prepare your research for manuscript publications in scientific journals.

2.0 OBJECTIVES

At the end of this unit, you will be able to:

1. understand how to write the results section of your dissertation
2. describe the various elements to consider when writing the discussion of the findings of your research
3. describe and understand the various steps involved in manuscript preparation
4. learn how to choose the right scientific journals to publish your research

3.0 MAIN CONTENT

3.1 Writing Results of your Dissertation

The results are the findings that emanate from the research after subjecting the data collected to statistical analyses. In your dissertation, the results can be presented as a standalone section in chapter four. Sometimes, depending on the format, the results and the discussion can be combined together in chapter four. The important focus of the results section is to highlight and present what was found in your research in a logical and unbiased manner without inferring meaning to the outcomes. The results of your dissertation should be comprehensive enough to include all findings and outcomes of the analyses that were conducted in your chapter three and related to your objectives. This section is usually written in past tense. Findings from the descriptive statistics should first be presented before the findings from the inferential statistics

deployed to test the hypotheses of the research. This is to allow the readers to have an understanding of the sociodemographic characteristics of the participants involved in your research. For the descriptive results, it is conventional to write about the sociodemographic distribution of the subjects that participated in the study as, an example: *the average age of the participants in the study was 45±6.5 years. The age of the participants ranged from 25 – 83 years. While 25% of the participant were female, 75% were male.*

Omotara and colleagues (2015) in their public health research on assessment of determinants of healthy ageing among the rural elderly of North-Eastern Nigeria provide better examples on how to present the descriptive results of a public health research. The authors wrote the descriptive section of their results as:

A total of 1600 elderly aged 60 years and above were recruited for the study out of which 53% were 60 to 64 years, 72.4% males and 85.6% were Muslim. Majority 81.6% were still married, 48.9% were farmers while 73.6% had no formal education. Majority 85.3% earn less than $100 (N16000) monthly, 73.6% did not have medical check-up in the last 12 months preceding the study while only 26.4% had monthly, quarterly and annual medical check-up.

The results presentation in the Omotara and colleagues (2015) study also provide a classical example of how the information presented in the tables of results can be described in the text of the results. For example, the authors described in details the results presented in their table as shown below:

Table 2 shows the prevalence of chronic medical conditions as reported by the respondents. Only 308 (19.3%) reported that they were hypertensive out of which 221 (71.8%) were on antihypertensive medication and 107 (6.7%) reported to be diabetic out of which 89 (83.2%) were on medication. On examination, 270 (16.9%) were found to be hypertensive and 188 (11.8%) were diabetic. Of the 270 diagnosed to be hypertensive 169 (62.6%) were newly diagnosed while the remaining were those who

had hypertension but was not controlled. Of the 188 diagnosed to be diabetic 142 (75.5%) were newly diagnosed while the remaining were those whose diabetes were not controlled. There was no significant gender disparity in both conditions. Slightly more than half 53.4% reported to have chronic joint pains which shows no significant gender disparity, however, 23.4% of them use walking aid. Only 27.6% had urinary problems with a female/male ratio of 1 to 3. Less than half (46.8%) reported to have eye problems. On examination 2.5% were found to have corneal opacity, 4.9% cataract, 1% ptosis, 0.4% ectropium, 0.6% entropium and others 0.7%.

However, it should be noted that while inferential statistics like t-test and Chi-square are also commonly used to test the differences in the distribution of the outcome of interest by demographic characteristics, inferential tests of regression statistics and correlation coefficients are usually used to report the associations and relationships between the outcomes of interest. For example, Oyeyemi et al (2013) in a study to determine the prevalence of physical inactivity among adults in Maiduguri, found differences in physical inactivity prevalence according to the participants sociodemograhic characteristics. The authors wrote the results for the differences in their findings as:

The prevalence of physical activity was 68.6%, and about one-third of participants (31.4%) were physically inactive. There was no significant difference ($p>0.05$) between the proportion of men (68.0%) and women (69.3%) who were physically active but the proportion of physically active men decreased from 75.6% in the youngest group to 53.6% in the oldest group, while the proportion of physically active women increased from 68.1% in the youngest group to 69.5% in the oldest group. Participants with less than a secondary school education were significantly($p<0.05$) more physically active (76.7%) than those with a secondary school education (66.5%) and those with more than a secondary school education (60.9%). Physical activity was significantly ($p<0.05$) greater among those who had blue collar jobs (artisan, trader) than among those with white collar jobs and those who were not employed (homemaker, student, retired, or unable to work). Participants who did not own a car

(77.9%) were significantly (p<0.05) more physically active than those who did (57.6%). The highest prevalence rates of physical activity were among participants who were divorced/separated (80.4%), those without a car (77.4%), those with a blue-collar job (76.6%), those with less than a secondary education (76.6%), and those with the lowest income (72.6%).

The authors went further to report the association between prevalence of physical activity and sociodemographic variables as:

Married participants were about 52% more likely to be physically active (OR = 1.52, CI = 1.02–4.73) than those who were single or never married. Participants with blue collar jobs were more than twice as likely to be physically active (OR = 2.19, CI = 1.16–4.12) as those who had white collar occupations or were unemployed. On the other hand, participants were less likely to be physically active if they owned a car (OR = 0.38, CI =0.17–0.86) or had a monthly income greater than 90 000 naira (OR = 0.54, CI = 0.10–0.95).

In the Omotara and colleagues (2015) study, the results of stepwise regression analysis on association between various combinations of lifestyle and psychosocial factors as determinants of healthy ageing were written as:

Diet in combination with alcohol consumption (R^2 = 0.044, p = 0.001), cigarette smoking (R^2 = 0.057, p = 0.022) and physical activity were significantly associated with healthy ageing while no significance was observed in combination with duration of sleep (R^2 = 0.002, p = 0.958). Combination of cigarette smoking with physical activity was not significantly associated with healthy ageing while combination of duration of sleep with cigarette smoking and physical activity were significantly associated with healthy ageing. Combination of three life styles were significantly related to healthy ageing except the combination of diet, alcohol and cigarette smoking (R^2 = 0.001, p = 0.633). Combination of four and five life style factors were all significantly associated with healthy ageing while combination of the various

psychosocial support were also significantly associated with healthy aging among our study population.

3.2 Writing Discussion of your Research Findings

The discussion section needs to follow from your results and relates back to your literature review. This is mostly the place to interpret and explain your results and critically evaluate the importance of your research. McCombes (2019) described the key elements involved in writing the discussion chapter of a dissertation to include focus on interpretations (what do the results mean), implications (why do the results matter), limitations (what can't the results tell us) and recommendations (what practical actions or scientific studies should follow).

Hess (2004) fully described the key important elements to include in the discussion section and things to avoid when writing the discussion. These are reproduced and summarized with appropriate practical examples below:

a. State the Major Findings of the Study

According to Hess (2004), the discussion should begin with a statement of the major findings of the study. The chapter should start by restating your research problems and then concisely summarizing the major findings. However, it should not include data or reference to the study design. This statement of the major findings of the study should normally no be more than one paragraph. Several examples illustrate the point. In a research by Obasola and Mabawonkwu (2018), the discussion begins with the sentence as: *"Our results showed that mothers' perception of maternal and child health information was largely positive"*. This clearly states the most important finding of that study. Oyeyemi et al (2013) also started their discussion section with the statement: *"To our knowledge, no published study has assessed the prevalence of physical activity and the way in which sociodemographic factors relate to physical activity levels in a subnational sample of Nigerian adults. This study revealed that about 68% of participants met WHO recommendations for sufficient physical*

activity". Adeloye and colleagues began the discussion section of their paper with the sentence, *"Our study reflects the difficulties that many experts have noted in describing the extent of road traffic crashes, injuries and deaths in Africa, for which modelling based on scarce and variable information, may not necessarily provide a reliable estimate"*. These provide good examples of a direct, declarative, and succinct proclamation of the study results.

b. Explain the Meaning of the Findings and Why the Findings Are Important

While the meaning of the results might seem obvious to you, they might not be so clear to the readers. Thus, it's important to spell out the significance of the results for the reader and show exactly how they answer your research questions. Hess (2004) opined that one of the purposes of the discussion is to explain the meaning of the findings and why they are important, without appearing arrogant, condescending, or patronizing. After reading the discussion section, the reader must be able to think that your results "makes perfect sense without over interpretation and unwarranted speculation". For example, Oyeyemi et al (2013) in their study explained the meaning of their findings of why physical activity tended to decrease with increasing age among active men but tended to increase with increasing age among active women in the following way:

"While it is difficult to explain these discrepant findings, our results may reflect the very high level of physical activity among divorced/separated (80.4%) adults, which mostly comprised older women. Perhaps Nigerian women who were divorced/separated and thus older had more work responsibilities, in addition to their traditional domestic duties (house/yard chores). If so, they would have had more job-related physical activity as compared with married or single women, who were younger. Also, older women in our study had lower incomes and education level and were less likely to own a car. The routine of walking daily to the market or doing shopping among these older women might explain why physical activity increased with increasing age among physically active Nigerian women in this study."

c. Relate the Findings to Those of Similar Studies

According to Hess (2004), no study is so novel and with such a restricted focus that it has no relation to other previously published papers. Therefore, the discussion section should relate your study findings to those of other studies. Questions raised by previous studies may have served as the motivation for your study. The findings of other studies may support your findings, which strengthens the importance of your study results. Obasola and Mabawonkwu (2018) discussed their study results in the context of a previous study by others as:

"Our finding that the mothers were more comfortable receiving maternal and child health information by phone and radio corroborates that of Palmer (2010) in India and was also in tandem with Bowen (2010) and Murthy (2010) in Ghana and Tanzania, respectively. But it differed from what was found in a study (Castle et al., 2011) in the northern part of Nigeria where the study indicated a negative disposition by the sampled mothers to the use of mobile phones for receiving maternal and child health information owing to poor ICT skills."

It is also important to point out how your study differs from other similar studies. An example can again be drawn from the study of Oyeyemi et al (2013) that explained the reasons why their findings are different from other studies as:

"The lower prevalence rate of physical activity in our study compared to those reported for 18 of the 22 African countries that participated in the recent WHO Stepwise approach to chronic risk factor survey might be due to differences between the measurement tools used in the two studies. The international physical activity questionnaire (IPAQ) used in our study was designed for population surveillance of physical activity but has been reported to overestimate physical activity prevalence as compared with the Global Physical Activity Questionnaire (GPAQ) used in the study of 22 African countries. However, the GPAQ was not used in the present study because, unlike the IPAQ, it has not been validated and tested in Nigeria."

d. State the Practical and Policy Relevance of the Findings

The major goal of public health research is to produce findings that are relevant to practice and policy. Thus, it is important for the discussion section of your dissertation to highlight the relevance of your research to public health practice and government policy. You must not over inflate the importance of the findings in a way that are not supported by your data. Thus, you must not introduce new results that are not based on your data because of your expectation for practice and policy. The main questions to ask when writing this section of your discussion is: "how does the findings of my research affect or influence public health practice", what does the findings contribute to public health practice", and "what can the government learn from the findings to use to improve policy decisions" The paper by Omotara et al (2015) gives an example of a public health study for which the practical and policy relevance was clearly stated:

"The fact that majority of the elderly in our study did not have medical examination in the past 12 months is an indication of the poor and inadequate health care system which the three tiers of government (local, state and federal) need to seriously address. Our results also indicate that there is a need for improvement in the health seeking behavior of the community. These two should be done simultaneously, that is: improvement in the health care system and creation for awareness on the importance of regular medical check-up for elderly. The need for the Nigerian government to take the issue of care of the elderly with all the seriousness it deserves has been sufficiently made by the growing number of elderly due to reduced fertility and increasing life expectancy. Older people's lives are characterized by growing inadequacies in customary family supports, social exclusion and non-existent social security targeted at them, thus being vulnerable to poverty and diseases".

e. Acknowledge the Study's Limitations

Even the best research has some limitations, and acknowledging these is a very important part of the research process. According to McCombes (2019), limitations

are not about listing your errors, but about providing an accurate picture of what can and cannot be concluded from your study. Limitations might be due to your overall research design, methodological approach, or unanticipated obstacles that emerged during the research process. You should only mention limitations that are directly relevant to your research objectives, and evaluate how much impact they had on achieving the aims of the research. For example, if your sample size was small or limited to a specific group of people, note that this limits its generalizability. If you encountered problems when gathering or analyzing data, describe these and explain how they influenced the results. After noting the limitations, you can reiterate why the results are nonetheless valid for the purpose of answering your research questions (McCombes, 2019). The paper of Oyeyemi et al (2013) provided a comprehensive example of how the limitations of a study should be acknowledged. The main limitations of their study were written as:

"*This study has some limitations that should be considered when interpreting the results. Although the sample was selected to be representative, the response rate was comparatively low, which could compromise the external validity and generalizability of the findings. The use of a self-report measure of physical activity, with the potential for information bias, is another limitation of this study. Overreporting due to social desirability leading to overestimation of the prevalence of physical activity has been reported for the IPAQ. Moreover, the validity of the short version of the IPAQ in adequately capturing patterns of physical activity has been challenged. The short IPAQ does not differentiate between activity contexts. For example, individuals may engage in walking activity solely for leisure or transportation or engage in physical activity as part of house chores, sports, or for job commuting or transportation purposes. However, despite the limitations, in view of its standardized survey methodology and measures and appropriate statistical methods, this study provides a valuable snapshot of physical activity prevalence and its patterns among Nigerian adults with different sociodemographic characteristics. The findings have implications for identifying the sociodemographic groups in Nigerian society that need to be targeted for effective interventions promoting physical activity.*"

f. Make Suggestions for Further Research

This should be the place to make your recommendations. Based on the discussion of your results, you can make recommendations for practical implementation or further research. According to McCombes (2019), suggestions for further research can lead directly from the limitations, and concrete ideas should be given for how future work can build on areas that your own research was unable to address. The paper of Obasola and Mabawonkwu (2018) provides a clear example of how to write the suggestions for further research:

"Future ICT based projects/research should adopt more than one ICT tool and more consideration should be given to the development of more maternal and child health information products in local languages to ensure the effectiveness of the use of ICT for maternal and child health information dissemination".

g. Give the "Take-Home Message" in the Form of a Conclusion

According to Hess (2004), the take "take-home message" is what you want the reader to remember from your study, and this should be the first sentence of your conclusions section. The conclusions section may also provide suggestions for practice change, if appropriate. An example of a well-written conclusion comes from a study by Adeloye et al (2016), who wrote:

"*In conclusion, our study suggests that the burden of road traffic injuries in Africa is high and there is an underestimation of road traffic fatalities. Improved road traffic injury surveillance across African countries may be useful in identifying relevant data gaps and developing contextually feasible prevention strategies in these settings"*

3.3 Manuscript Preparation and Publication in a Journal

The ultimate goal of the research process is to have your study disseminated to other researchers and the wider community. The most popular method of dissemination of your research findings is through publication in peer reviewed scientific journals.

Thus, it is a significant accomplishment for you as a researcher to have your dissertation published in a public health journal. For scientific writing required for publication, one of the stumbling blocks is the beginning of the process and writting the first draft. However, this should not be a major problem for you since you already have a full dissertation by now. The major challenge for you would be how to summarize the dissertation into an acceptable publishable format and how to decide on the best journal to submit your manuscript. Belcher (2009) and Kallestinova (2011) proposed the guidelines that can be followed to prepare well-structured and comprehensive manuscripts for publication. Based on these strategies and the publication guidelines developed by the International Physical Activity and Environment Network (IPEN, 2019), the following tips on how to prepare and submit a manuscript for publication may guide you through the process:

1. Book at least 1-2-hour blocks of writing time in your daily work schedule and consider them as non-cancellable appointments.
2. Create an outline that will help generate ideas. The outline should be similar to a template of the manuscript you want to write and should initially form the structure of the paper. Write down in any order anything important about the paper. For example, what is the topic of my manuscript? Why is this topic important? What is the major finding of the manuscript I want to write? Are these findings important? And why are they important? Draw figures and tables, what are the implications of this research? what is the problem? Why are you writing this paper? Is this research important to the field? What is already known? What are the objectives of the paper? Who were the participants? How was the study conducted? Does it add to the body of knowledge? What were the significant results? Are there any supporting results or trends? What's new? What's interesting? Is there anything challenging? What are the limitations and strengths? Is this paper providing solutions to a particular problem? Then structure your ideas under introduction, methods, results, discussion and conclusion.

3. After you have developed your outline, the next step is to discuss your ideas for your manuscript with your supervisor(s) or mentor and your coauthors. Getting feedback during early stages of your draft can save a lot of time. Talking through ideas allows people to conceptualize and organize thoughts to find their direction without wasting time on unnecessary writing. This is also the best time to identify the top 3 journals for submission on the basis of the results of the paper.
4. After you get enough feedback and decide on the journal you will submit to, you then begin the process of real writing which includes reworking the contents of what is already written in your dissertation. You will need to copy your outline into a separate file and expand on each of the points by adding data and elaborating on the details. When you create the first draft of the manuscript, do not succumb to the temptation of editing. Do not worry about the language. Get your ideas down and then go back to fine-tune the language, logic and sentence structure later.
5. Take a second look at the first draft and ensure that the Introduction and objectives of the manuscript match with the methods, results and discussion. The question to ask and what to look for in the manuscript at this time is whether there is a flow? Is it well organized and presented? You have to ensure that information is not being repeated unnecessarily, especially repeating the results in the discussion section. You must also ensure that your discussion is representative of the results.
6. Then share the first draft with your supervisor(s) and co-authors and get as much feedback as you can. You should be open to all feedback as the goal is to improve the manuscript so that it can be accepted in high impact journal.
7. In order to increase the chances of acceptance of your paper you should choose a journal that its scope and aim align well with the research focus of your paper. As a start, you should have a look at your reference lists for the articles that you cited to see the journals where similar articles have been published. You can also make use of journal selection tools such as JANE.

JANE is located at http://jane.biosemantics.org/ where you type the title of your manuscript, click on 'Find journals', 'Find authors' or 'Find Articles'. JANE will then compare your document to millions of documents in Medline to find the best matching journals, authors or articles.

8. Ensure that you adhere to the guidelines of the journal. Read the guidelines carefully for the cover letter, title page, manuscript formatting, figures, tables and references
9. While writing the manuscript is the main goal, do not underestimate the power of a strong cover letter. Here you must highlight the importance of your research and showcase the relevance of your findings. The Editor-in-Chief of the journal may decide on the basis of your cover letter whether to send your paper out for review or not.

4.0 CONCLUSION

The results section of your dissertation is where to present the findings that emanate from your research after data analyses. The results should be presented in a logical and unbiassed manner highlighting only what was found in the research without inferring any interpretation to the findings. The results section should usually be written in the past tense. The discussion section should be the place to interpret and explain your results and critically evaluate the importance of your research. While writing the discussion of your research you should focus on the interpretations of the findings (what do the results mean), implications of the findings (why do the results matter), limitations of the findings (what can't the results tell us) and recommendations from the findings (what practical actions or scientific studies should follow). The followings are important elements that should be included when you are writing your discussion: State the study's major findings, Explain the meaning and importance of the findings, Relate the findings to those of similar studies, State the practice and policy relevance of the findings. Acknowledge the study's limitations, Make suggestions for further research, and Give the "Take-Home Message" in the Form of a Conclusion. However, the ultimate goal is to have your research publish in a journal that is accessible to other researchers and the community at large. To prepare

and submit your manuscript for publication in a journal, you need to follow some guidelines and rules to increase your chance of success. These rules include: (1) Create regular time blocks for writing as appointments in your calendar and keep these appointments, (2) Create a detailed outline and discuss it with your supervisor(s) or mentor and co-authors, (3) Interest your reader in the introduction section by signaling all its elements and stating the novelty of the work, (4) Be clear, concise, and objective in writing your introduction and methods and in describing your Results and writing the discussion of the findings, (5) Receive feedback on the draft manuscript from your supervisor(s) and co-authors and then revise accordingly, (7) Choose and submit only in a journal that its scope and aim fit with the research focus of your manuscript. You will need to be conversant with the 'note to contributing authors' of the selected journal preparatory to writing the paper.

5.0 SUMMARY

In this unit, you have learnt how to write the results section of your dissertation. You have also learnt how to write the discussion of your research findings, including understanding the various components to consider when writing the discussion of your research. You have also been introduced to the various steps involved in manuscript preparation and how to choose the right scientific journals to publish your research.

6.0 TUTOR-MARKED ASSIGNMENT

1. Write an example of how you will describe the results of the age and sex distribution of the members of your class attending the public health course on research methods.
2. Itemize and describe the important elements you will consider when writing the discussion section of your research.
3. Describe the steps you will follow to prepare and submit your manuscript for publication in a journal.

7.0 REFERENCES/FURTHER READING

Adeloye, D., Thompson, J.Y., Akanbi,M.A., Azuh, D., Samuel, V., Omoregbe, N. & Ayo, C.K. (2016). The burden of road traffic crashes, injuries and deaths in Africa: a systematic review and meta-analysis. *Bulletin of the World Health Organization,* 94:510–521A

Belcher, W.L. (2009). Writing Your Journal Article in 12 Weeks: a guide to academic publishing success.Thousand Oaks, CA: SAGE Publications

International Physical Activity and the Environment Network (IPEN, 2019). Publication Guidelines for the IPEN Studies. Available at: https://www.ipenproject.org/

Kallestinova, E.D. (2011). How to Write Your First Research Paper. *Yale Journal of Biology and Medicine*, 84:181-190. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3178846/

Obasola, O.I. & Mabawonku, I.M. (2018). Mothers' perception of maternal and child health information disseminated via different modes of ICT in Nigeria. *Health Information and Libraries Journal*, 35: 309–318

McCombes, S. (2019). How to write a discussion section. Available at https://www.scribbr.com/dissertation/discussion/

Omotara, B.A., Yahya, S.J., Wudiri, Z., Amodu, M.O., Bimba, J.S. and Unyime, J. (2015) Assessment of the Determinants of Healthy Ageing among the Rural Elderly of North-Eastern Nigeria. *Health*, **7**, 754-764.

Oyeyemi, A.L., Oyeyemi, A.Y., Jidda, Z.A., and Babagana, F. (2013) Prevalence of physical activity among adults in a metropolitan Nigerian city: A cross-sectional study. *Journal of Epidemiology*, 23(3):169-177.

APPENDIX 1: WORK PLAN FOR RESEARCH PROJECT

WORK PLAN/ GANTT CHART

Activity	**Time Schedule**						
	January-May 2020	June 2020	July – October 2020	November-2020	December-January 2021	February-March 2021	April-June 2021
Proposal writing, submission and defense							
Preparing and testing of tools							
Data collection (field Work)							
Data Analysis							
Writing chapter 4 & 5, edit, revise							
Dissertation presentation and corrections							
Submission of Dissertation							

APPENDIX 2: TEMPLATE FOR RESEARCH PROPOSAL

CHAPTER ONE: INTRODUCTION

1.1 Introduction

[Start Here, between 2 and 3 pages]

1.2 Statement of the Problem

[Start Here]

1.3 Objectives of the Study

[The objectives of the study will be to:]

1.4 Research Questions

[The following research questions will be answered:]

1.5 Hypotheses

[The following major and sub hypotheses will be tested in this study:]

1.5.1 Major Hypothesis

[Start Here]

1.5.2 Sub Hypotheses

[Start Here]

1.6 Significance of the Study

[Start Here]

1.7 Delimitation (scope) of the Study

[Start Here]

1.8 Possible Limitations of the Study

[Start Here]

1.9 Operational Definition of Terms (if any?)

[Start Here]

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

[Start Here. State the aim of the study. Indicate the search engines, databases and search terms used for the literature review here. Also give the number of relevant papers retrieved during the search]

2.2 Core Literature and Sub-sections

[This section and its subsections should contain literature that is related to the objectives and research questions of the study. Further sub-sections should be categorized as 2.3.1, 2.3.2, 2.3.3 etc]

2.3 Secondary Literature and Sub-sections

[This section and its subsections should contain literature which are not related to the objectives and research questions of the study but are important to understanding the broad scope of the research. Further sub-sections should be categorized as 2.4.1, 2.4.2, 2.4.3 etc]

2.4 Other relevant Literature and Sub-sections (if any?)

Further sub-sections should be categorized as 2.5.1, 2.5.2, 2.5.3 etc

2.5 Summary of the Reviewed Literature and Uniqueness of the Study

[Start Here]

CHAPTER THREE: METHODOLOGY

3.1 Subject/Participants

[Start Here: Describe in details the characteristics of the participants to be used in the study (e.g., basic sociodemographics, source population and town]

3.1.1 Inclusion Criteria

[The following participants will be eligible to participate in the study:]

3.2 Materials

3.2.1 Instruments

[The following instruments will be used for data collection:]

3.3 Methods

3.3.1 Research Design

[Start Here]

3.3.2 Study Site

[Start Here]

3.3.3 Sampling Technique and Sample Size

[Start Here]

3.3.4 Procedure

[Start Here]

3.3.5 Ethical Issue

[In addition to mentioning the institution responsible for ethical clearance and how you will obtain the ethical approval, briefly state how your study will ensure issues of participants' anonymity, confidentiality, malfeasance and beneficence, and data protection]

3.3.6 Study Duration

[Start Here]

3.4 Data Analysis

[Start Here. State the analytic method and the statistical software for the data analyses]

3.5 Expected Outcomes

[It is expected that the outcome of this study will provide:]

3.6 Timeline

[Here, include a table or chart showing the expected timeline for each stage of the project]

REFERENCES

Use the Harvard Reference Style in the list and in the text

APPENDIX

[Start Here]