Research for Emergency Medicine Residents

Third Edition

David E. Hogan DO MPH FACEP

Research for Emergency Medicine Residents

3rd Edition

This text represents the third version of the Research for Emergency Medicine Residents to be completed since 1998 when the first edition was created. It has been considerably expanded from earlier editions and now includes chapters on the basic statistics and mathematical management of research data. This book is intended to assist the resident in getting started towards completion of their individual research project while in their residency program. It is intended to be used in conjunction with a research advisor or coordinator and in no way provides all the information needed to perform research in its entirety. Although considerable information is provided here, nothing can substitute for the advice and guidance of an experienced researcher to assist the resident in the completion of their task. Use this text to learn and perform the basic tasks associated with getting a research project started in conjunction with your research advisor.

It has been my honor and pleasure to work with residents for almost 30 years, and I have learned much more from them, then they from me. Let us continue to move ahead in our learning, and make a difference.

David E. Hogan – Oklahoma City, Oklahoma – October 2010.

David E. Hogan DO MPH FACEP

Adjunct Professor
Oklahoma State University
Research Coordinator and Director of Prehospital and Disaster Medicine Fellowship Program
Integris Southwest Medical Center Emergency Medicine Residency Program
Oklahoma City, Oklahoma
Shogan4@Cox.Net

Introduction to Research 14
WHY DO RESEARCH DURING RESIDENCY? 14
READING THE LITERATURE 14
SELECTING ARTICLES TO READ 15
THE GENERAL STRUCTURE OF RESEARCH PROTOCOLS 15
ANATOMY OF AN ARTICLE 16
ABSTRACT 16
INTRODUCTION/BACKGROUND 16
METHODS 17
STUDY DESIGN 17
STUDY POPULATION 17
TREATMENT ALLOCATION 17
INDEX OF ACCOMPLISHMENT 17
STATISTICAL ANALYSIS 18
RESULTS 18
DISCUSSION 18
SUMMARY 18
Numbers and the Nature of Research 20
INTRODUCTION 20
SAMPLE POPULATIONS 20
DATA 20
DISCRETE VERSES CONTINUOUS DATA 20
DISCRETE DATA 20
CONTINUOUS DATA 20
NOMINAL, ORDINAL, INTERVAL, AND RATIO DATA 21
NOMINAL DATA 21
ORDINAL DATA 21
INTERVAL DATA 21
RATIO DATA 21
DATA VALUES DISTRIBUTED AROUND MEANS 21
SUMMARY 22
Bias and Confounding 23

INTRODUCTION 23
BIAS 23
EVALUATION FOR BIAS 23
BIAS CLASSIFICATION AND NATURE 24
SPECIFIC EXAMPLES OF BIAS 24
BIAS ASSOCIATED WITH THE STUDY PARTICIPANTS 24
BIAS IN THE PROCEDURAL PROCESS 25
BIAS IN ANALYZING DATA 25
BIAS IN THE RESULTS AND CONCLUSIONS 26
CONTROLLING AND ELIMINATING BIAS 26
CONFOUNDING 26
EXAMPLE OF A CONFOUNDER 26
IDENTIFICATION AND CONTROLLING OF CONFOUNDERS 27
IDENTIFICATION OF CONFOUNDERS 27
CONTROLLING CONFOUNDERS 27
SUMMARY 28
Ethics in Research 29
INTRODUCTION 29
ACCELERATION OF BIOMEDICAL INFORMATION 29
LIFE AND DEATH 29
PRIVACY 29
THE THEORETICAL BASIS OF ETHICS 30
DEONTOLOGY 30
CONSEQUENTIALIST ETHICAL THEORY (UTILITARIANISM) 30
A THEORY OF JUSTICE 30
THE LIBERTY PRINCIPLE 30
THE DIFFERENCE PRINCIPLE 31
HISTORICAL NOTES ON RESEARCH ABUSE 31
NAZI ABUSES OF RESEARCH 32
JAPANESE RESEARCH ABUSES 32
AMERICAN RESEARCH ABUSES 32
THE SEVEN DEADLY SINS OF RESEARCH 34

REGULATION OF RESEARCH IN THE UNITED STATES 34
REGULATING AGENCIES 34
THE BASIS OF RESEARCH REGULATIONS 34
THE NUREMBURG CODE 34
THE DECLARATION OF HELSINKI 35
THE NATIONAL RESEARCH ACT OF 1974 35
THE BELMONT REPORT 35
THE INSTITUTIONAL REVIEW BOARD (IRB) 36
IRB FEDERAL POLICY REQUIREMENTS 36
INFORMED CONSENT 37
THE ELEMENTS OF FULL INFORMED CONSENT 37
ADEQUATE INFORMED CONSENT 38
THE REASONABLE PHYSICIAN STANDARD 38
THE REASONABLE PATIENT STANDARD 38
THE SUBJECTIVE STANDARD 38
EMERGENCY MEDICINE RESEARCH CONSENT 38
THE EMERGENCY EXEMPTION "WAVER" 39
ADDITIONAL IRB DUTIES UNDER EMERGENCY EXEMPTION WAVERS 39
EXEMPTION FROM CONSENT 39
EXPEDITED IRB REVIEW 39
FINANCIAL CONSIDERATIONS 40
SUMMARY 40
The Research Question 42
INTRODUCTION 42
THE FINER HYPOTHESIS CONCEPT 42
FEASIBLE 43
INTERESTING 43
NOVEL 43
ETHICAL 43
RELEVANT 43
BUILDING ON THE RESEARCH QUESTION 44
CAPTURING THE GOOD IDEAS 44

MAKING NOTES ON YOUR RESEARCH QUESTION 44
THE QUICK LITERATURE SEARCH 44
DEFINING THE STUDY POPULATION – INCLUSION AND EXCLUSION CRITERIA 44
THE CASE DEFINITION 45
THE OUTCOME MEASURE 45
OTHER DATA POINTS 45
FURTHER REFINEMENT 45
CREATING THE HYPOTHESIS 46
SUMMARY 46
Searching the Medical Literature 47
INTRODUCTION 47
GETTING STARTED 47
TEXTBOOKS AND THEIR SOURCES 47
MINING THE REFERENCES 48
THE FORMAL LITERATURE SEARCH 48
WHERE TO SEARCH? 48
PUBMED 49
THE IMPORTANT ELEMENTS OF PUBMED FOR RESIDENT RESEARCH PURPOSES 50
MECHANICS OF THE SEARCH 50
WHEN DO YOU HAVE ENOUGH ARTICLES? 51
THE MEDICAL LIBRARIAN 52
SUMMARY 52
Study Design 53
INTRODUCTION 53
THE SURVEY 53
HISTORICAL APPROACHES 53
THE FLOW OF INFORMATION PROBLEM IN HISTORICAL STUDIES 54
EXPERIMENTAL APPROACHES 54
THE SURVEY STUDY 54
TYPES OF SURVEY STUDIES 54
QUESTIONNAIRES 54
INTERVIEWS 55

WHAT SURVEY STUDIES CAN ACTUALLY DO 55
THE RESEARCH QUESTION IN SURVEYS 55
SURVEY SAMPLING METHODS 55
SURVEY LENGTH 56
REPEATING QUESTIONS 56
QUESTIONNAIRE DESIGN (THE SURVEY INSTRUMENT) 56
QUESTION STRUCTURE (OPEN OR CLOSED) 56
SURVEY SCALES 57
LIKERT (SUMMATIVE) 57
FORCED LIKERT 58
SEMANTIC DIFFERENTIAL (FORCED OR NOT) 58
GUTTMAN (CUMULATIVE) 58
VISUAL ANALOG (VAS) 59
NUMERICAL DESCRIPTOR SCALE (NDS) 59
SOURCES OF BIAS 59
CLARITY OF THE SURVEY INSTRUMENT 60
SURVEY LANGUAGE 60
ENGAGING THE RESPONDENT 60
Shared Interests 60
Strike While the Iron is Hot 60
Provide Inducement 60
FACE TO FACE DATA COLLECTION 61
ON LINE SURVEYS 61
PILOT TESTING 61
PLANNING FOR NON-RESPONDERS 61
RESPONSE RATES AND RESPONDERS 61
THE CHART REVIEW 62
ADVANTAGES OF CHART REVIEW STUDIES 62
DISADVANTAGES OF CHART REVIEW STUDIES 62
WHAT A CHART REVIEW CAN TELL THE RESEARCHER 62
ADDITIONAL REASONS TO AVOID DOING CHART REVIEWS 63
CREATING A CHART REVIEW 63

THE RESEARCH QUESTION 63
PERFORM A LITERATURE SEARCH AND WRITE A RESEARCH PROTOCOL 63
WRITE AN ABSTRACT MANUAL AND CREATE A DATABASE 63
OBTAIN INSTITUTIONAL REVIEW BOARD APPROVAL OR EXEMPTION 63
TRAIN ABSTRACTORS 64
ABSTRACTING THE DATA AND ENTERING DATA TO THE DATABASE 64
CLEAN UP THE DATABASE AND PERFORM THE ANALYSIS 64
COHORT STUDIES 65
HISTORY OF COHORT STUDIES 65
WHAT A COHORT STUDY CAN TELL THE RESEARCHER 66
CREATING A COHORT STUDY 66
A PROSPECTIVE COHORT STUDY 67
A RETROSPECTIVE COHORT STUDY 67
CASE-CONTROL STUDY 68
ADVANTAGES OF CASE-CONTROL STUDIES 69
DISADVANTAGES OF CASE-CONTROL STUDIES 69
CONFOUNDING IN CASE CONTROL STUDIES 69
BIAS IN CASE CONTROL STUDIES 69
DESIGNING A CASE-CONTROL STUDY 69
CASE-CONTROL STUDY EXAMPLE 70
META-ANALYSIS 70
WHEN TO USE A META-ANALYSIS 70
EXAMPLE OF A META-ANALYSIS 71
CREATING A META-ANALYSIS 71
ADVANTAGES AND DISADVANTAGES OF META-ANALYSIS 71
RANDOMIZED CLINICAL TRIALS (RCT) 71
WHAT A RANDOMIZED CONTROLLED TRIAL CAN TELL YOU 72
ADVANTAGES OF THE RANDOMIZED CONTROLLED TRIAL 72
DISADVANTAGES OF THE RANDOMIZED CONTROLLED TRIAL 72
METHODS OF RANDOMIZATION 73
RANDMOIZED CONTROLLED TRIALS IN THE CONTEXT OF CLINICAL TRIALS 73
PHASE 0 73

PHASE I 74
PHASE II 74
PHASE III 74
PHASE IV 74
CROSSOVER STUDY 74
SUMMARY 75
Writing the Research Protocol 77
INTRODUCTION 77
STEP ONE – GET INFORMATION IN ORDER 77
STEP TWO – CREATE A TITLE PAGE 77
STEP THREE – THE REASON FOR THE STUDY 78
INTRODUCTION 78
RESEARCH QUESTION 78
STUDY PURPOSE 78
STEP FOUR – A METHODS SECTION 78
CASE DEFINITION 79
INCLUSION AND EXCLUSION CRITERIA 79
RANDOMIZATION METHOD 79
MANAGEMENT OF STUDY RECORDS 79
STEP FIVE – THE DISCUSSION 80
STEP SIX - REFERENCES 80
SUMMARY 80
Hypothesis Testing – Part One 82
INTRODUCTION 82
HISTORIC PERSPECTIVE 82
RATIONALISM 82
EMPIRICISM 82
CRITICAL RATIONALISM 83
THE HYPOTHESIS IN HYPOTHESIS TESTING 83
THE NORMAL DISTRIBUTION OF DATA 84
CENTRAL LIMIT AND NORMAL DATA DISTRIBUTION 84
LINKING HYPOTHESIS TESTING TO STATISTICS 84

DEFINING A HYPOTHESIS TEST 86
GRAPHICAL PRESENTATION OF HYPOTHESIS TESTING 86
SUMMARY 86
Hypothesis Testing – Part Two 87
INTRODUCTION 87
PARAMETERS AND ERRORS 87
PARAMETERS 87
ERRORS IN STATISTICS 87
TYPE I AND TYPE II ERRORS 87
CENTRAL TENDENCIES IN DATA COLLECTION 88
THE MEAN ($\mathbf{X}$) OR (μ) 88
THE MEDIAN 88
THE MODE 89
MEASURING DISPERSION WITHIN DATA SETS 89
THE RANGE (R) 89
PERCENTILES (QUANTILES) 89
INTERQUARTILE RANGE 89
DEVIATIONS FROM THE MEAN ($\mathbf{x} - \boldsymbol{\mu}$) 89
VARIANCE ($\boldsymbol{\sigma 2}$) 89
STANDARD DEVIATION (σ) 90
SAMPLE SIZE IN RESEARCH 90
STEPS IN SAMPLE SIZE CALCULATION 90
ALPHA, BETA AND SAMPLE SIZE 91
THE REALITY OF ESTIMATING SAMPLE SIZE 91
SUMMARY 91
Statistics - Sampling and the “P” Value 92
INTRODUCTION 92
SAMPLING ISSUES 92
INDEPENDENT SAMPLES 92
THE “P” VALUE 93
P VALUE DEFINITION 93
CALCULATION OF THE P VALUE 93

INTERPRETING THE P VALUE 93
ERRORS IN P VALUE INTERPRETATION 94
HOW SIGNIFICANT IS SIGNIFICANT 94
ONE AND TWO TAIL P VALUE TESTS 94
SUMMARY 95
Select Statistical Concepts and their Usefulness in Medical Research 96
INTRODUCTION 96
PARAMETRIC VERSES NON-PARAMETRIC TESTS 96
PARAMETRIC TESTS 96
THE STUDENT'S T TEST (PARAMETRIC) 96
THE PAIRED T TEST (PARAMETRIC) 96
CHI-SQUARE TEST (PARAMETRIC) 97
FISHER'S EXACT TEST (PARAMETRIC) 97
ONE-WAY ANOVA (PARAMETRIC) 97
THE F TEST (PARAMETRIC) 97
NON-PARAMETRIC TESTS 97
MANN-WHITNEY U TEST OR WILCOXON RANK-SUM TEST (NON-PARAMETRIC) 97
KRUSKAL-WALLIS (NON-PARAMETRIC) 98
MCNEMAR'S TEST (NON-PARAMETRIC) 98
WILCOXON SIGNED-RANK TEST (NON-PARAMETRIC) 98
ANALYSIS OF VARIANCE ANOVA (NON-PARAMETRIC) 98
KRUSKAL-WALLIS TEST (NON-PARAMETRIC) 98
THE FRIEDMAN ONE-WAY ANOVA (NON-PARAMETRIC) 98
SENSITIVITY AND SPECIFICITY 99
POSITIVE AND NEGATIVE PREDICTIVE VALUES 100
SUMMARY 100
Confidence Intervals 101
INTRODUCTION 101
CONFIDENCE INTERVALS AND DATA DISTRUBUTION 101
THE 95% CONFIDENCE INTERVAL 101
EXAMPLE CALCULATION OF A 95% CONFIDENCE INTERVAL 102
INTERPRETATION OF 95% CONFIDENCE INTERVALS 102

SUMMARY 102
Statistical Associations between Variables 104
INTRODUCTION 104
EXAMPLE OF AN ANALYSIS OF ASSOCIATION 104
WHEN A STATISTICAL ASSOCIATION – IS NOT 105
EXAMPLE OF A FALSE RELATIONSHIP BETWEEN VARIABLES 105
SUMMARY 106

Introduction to Research

To see a world in a grain of sand and a heaven in a wildflower, to hold infinity in the palm of your hand and eternity in an hour. *William Blake*

WHY DO RESEARCH DURING RESIDENCY?

The primary advances in medicine over the last century have been by the application of scientific principles to the ancient art. Early healers and Shaman applied principles of nature as they understood them to those afflicted by injury or illness. These principles of nature were rarely based on scientific observation and often provided little more than placebo effect. Even today many of the patients we treat place stock in psychic forecasting, spiritualism, and junk pseudoscience. This can be understood both because it is human nature to develop belief patterns based on emotions, and because much of the "science" published on a regular basis is not necessarily good science. Like it or not, we practice in a world filled with superstitions as well as science. As physicians, it is our hope that good science and reason will guide our recommendations and treatments as we care for the patients entrusted to us. This means that as a practitioner of medicine you need to know how to tell the difference between good science, and bad. In addition, at some point in your career you may wish to add to the body of knowledge by conducting research yourself.

The creation, use, and interpretation of medical literature based on proper research and clinical testing is vital to the progression of medical science. Emergency physicians must be capable of interpreting the medical literature and determining if such literature will change the way they practice emergency medicine. The primary aim of this text is to introduce you to methods by which you can more readily understand and critique the medical literature. To do this, it is important that you have an understanding of how hypothesis generation, study methods, and analysis occur. Federal law prohibits some types of research, and stringently regulates all research pertaining to the use of human subjects. Consent, confidentiality, and ethical issues are vital in the conducting of research. Statistical analysis, rigorously applied to our medical observations lends confidence to our ability to accept or reject hypotheses.

The explosion of data within the medical literature has rendered it impossible for a single individual to assimilate all the information of even a single specialty area. Computers have become pervasive within the medical environment and are essential and retrieving information regarding new or standard diagnostic and therapeutic modalities. Computers have also made it possible for practitioners and researchers to consolidate worldwide research and opinions regarding discrete medical topic areas. Advances toward outcome based medicine make it necessary that physicians be capable of using computers for literature searches to acquire information regarding medical treatment of their patients.

By understanding the basic principals involved in the preparation of research protocols and the writing of research proposals, abstracts, and manuscripts, residents will acquire a deeper understanding of how the medical literature comes into existence. And they may indeed find themselves better prepared to perform their own observations in medical science helping us all to more fully understand the human condition.

In this text we will cover a considerable amount of material. Some terms and concepts may be confusing at first, but will become clear as one continues through the reading.

READING THE LITERATURE

One of the reasons to learn about research is to become proficient in reading the medical literature. Not everything published in medicine is worth reading. With the huge volume of writing being published, it is difficult

to determine what to read and what to discard. An understanding of the nature of research will assist one in making these decisions. We will start with considering why we want to be able to read the literature, how to select literature to read and finally a brief examination of the "*Anatomy and Physiology*" of scientific writing.

The physician may want to read clinical journals for the following reasons (1):

- To keep abreast of professional news
- To understand pathophysiology
- To discover how seasoned clinicians handle a particular problem
- To discover whether using a new or existing diagnostic test is worthwhile
- To learn clinical features and course of a particular disorder
- To determine etiology or causation
- To distinguish useful from useless or even harmful therapy through evidence based medicine
- To evaluate claims regarding indications and cost-effectiveness of health care processes
- To be stimulated by letters to the editor

This list provides multiple reasons to remain current with medical literature; however, the fact remains that biomedical literature is expanding at a rapid rate, doubling every 10-15 years. Emergency physicians provide care for patients with virtually every variety of disease and clinical presentation. To keep current with often-conflicting published research is a difficult task. Computerized databases have made accessing and sorting the medical literature somewhat simpler. However we are still faced with the task of determining which articles are deserving of our limited time in reading and contemplation.

SELECTING ARTICLES TO READ

Selecting what articles to read is a key skill in keeping up with the literature. There are several methods used by physicians to select articles. The following are a series of steps that may be used in deciding which article to spend time reading (2).

1. Analyze the title: Is this article potentially interesting or possibly useful in your practice? If not, reject it and move on to the next article.
2. Review the list of authors: One or more authors may be familiar. If the author's track record is good and of good reputation, read on. On the other hand, articles from authors with less than stellar reputations may be rejected. If the author is unfamiliar, consider the reputation of the journal, although this is not a foolproof guarantee of good work.
3. Read the summary/abstract or conclusion: The purpose of this step is to determine whether the conclusion, if valid, would be important to your clinical practice. At this stage, determining if the results are true is not important. If the results would be useful, read on.
4. Can the information be generalized: Finally, is the study site and population sufficiently similar to your practice so that the study's results, if valid, would apply to patients in your practice?

If an article has passed this basic 4 step filtering process, the full article should be read.

THE GENERAL STRUCTURE OF RESEARCH PROTOCOLS

Just as various methods can be utilized to perform research, various methods may be used to report it. There are, however, commonalities associated with the process of research that will influence the general way in which research results will be published. Understanding the underlying principles on which the research is conducted improves the ability of the Emergency Physician to read and understand the literature. Each research project is carried out utilizing a *Research Protocol*. The research protocol outlines the research question, the goals of the research, the methods to be utilized, and the techniques to analyze the information obtained. The research protocol acts as a "road map" for the performance of the research and a guide to insure that the research is

performed in a valid scientific manner. Following the performance of the research, the research protocol additionally provides the format for writing the scientific paper. As such, each published research report will have a general complement of similar elements based on the research protocol. These general elements of the research protocol include the following:

1. *Research question*. This discusses specifically why the investigator(s) are doing the research and what question(s) they hope to answer.
2. *Background information*. This element defines where the research fits into the overall scheme of scientific thought. It often contains a review of the pertinent literature pertaining to the project being proposed.
3. *Design*. This outlines the specific timeframe covered by the study and what design methodology the study will follow.
4. *Subjects*. This includes a definition of the subjects to be included in the research. It must include a definition of the inclusion and exclusion criteria for placing subjects into the research protocol, as well as what subjects cannot be placed in the study. One needs a clear definition of the subset of the population the subjects represent in order to determine how well any results may be generalized to the total population.
5. *Variables*. This element discusses specifically what is to be measured (variables) and how all the variables are to be measured and recorded. This should also discuss how bias and confounders will be dealt with.
6. *Outcome Measures.* In addition, the outcome measurements (final goals or endpoint) must be clearly defined.
7. *Statistics*. This element should describe the *Null Hypothesis* ($\mathbf{H_o}$) and the *Alternate Hypothesis* ($\mathbf{H_A}$). In addition, the calculated sample size required to achieve statistical significance and the analytical approach to be used in statistical analysis of the results should be defined.

Most research protocols will contain these 7 elements and they will be contained somewhere within the body of the research paper written - not exactly in this order, but the core information should be present. We'll discuss the development and writing of research protocols in detail later in this text. Suffice it to say for now, that the research protocol acts as a basis for not only conducting the research project, but for writing the results in a manuscript. As such, the research protocol forecasts the structure of the scientific manuscript eventually published.

ANATOMY OF AN ARTICLE

In addition to containing the elements of the research protocol, each article has specific divisions of organization. These divisions are relatively consistent although they may vary somewhat from journal to journal. Understanding what the purpose of each section is will aid in understanding scientific articles.

ABSTRACT

The abstract should provide an abbreviated summary of the entire study. Read the abstract to determine if the setting and patient population are useful. Do not depend on the abstract to tell you the "truth" about the article and thereby avoid reading the rest of the article. The abstract only represents the author(s) stated results and opinions about the article which may not be accurate (3, 4, 5, 6, 7).

INTRODUCTION/BACKGROUND

The introduction should acquaint the reader with the problem under study and explain the reasons for conducting the investigation. Background information may be presented, with a review of historical developments that led to initiation of the study. The most important portion of the introduction is identification of the study objective (study question and goals). The study objective identifies the specific question to be evaluated. All subsequent

sections should be read with the study objective in mind. The methods must be designed to answer the question and the investigator's conclusions should not extend beyond the stated objective.

METHODS

The methods section is the single most important section of any research publication in relation to being able to judge the validity of a study. Unfortunately in reality, it is the least read by most of us. A methods section should present information concerning experimental design, study sample, treatment allocation or observation period, index of accomplishment or outcome measure, and statistical test selection. It is only by reviewing the methods section that we may determine if the study is valid. The methods sextion should include at least the following information.

STUDY DESIGN

We will cover study design in detail during later sections of this text. However in the methods section the author(s) should describe the design of the study with regards to Experimental or Observational, Prospective or Retrospective, Cohort, Case-Control, Cross-Sectional, etc.

STUDY POPULATION

The author(s) must describe the study population well enough that the reader is able to precisely visualize the sample population under investigation. In addition, inclusion and exclusion criteria for admittance of individuals into the study must be clearly defined. These conditions enable the reader to determine whether the study sample sufficiently resembles their clinical practice to allow extrapolation. Exclusion criteria help to ensure that the study sample is as homogeneous as possible, identify patient subsets to which study results should not be extrapolated, and assure patient safety by excluding individuals for whom participation would be contraindicated or dangerous.

TREATMENT ALLOCATION

In an experimental study, the reader also must be able to determine in what way the study population was organized for any interventions (matched pair vs. independent samples) and how the population was assigned for the specific intervention (random vs. nonrandom). In addition was a control group used (placebo vs. standard therapy vs. non-controlled) and whether blinding was used to decrease biases.

INDEX OF ACCOMPLISHMENT

An index of accomplishment should be defined by the author(s). This is simply the measurement or observation that the author(s) have selected to assess outcome. The author(s) should specifically state the parameters measured when assessing response to treatment, after an observation period, or before reviewing the medical record. The index always should be as precise and reproducible as possible. This implies that it be free of potential patient recall bias, instrument measurement bias, and evaluator bias. A report of inter-observer reliability may be included for measures that are subject to subjective differences. Questions that may guide readers in determining whether the outcome measurement is appropriate include the following:

- Did the author use the right measurement to determine if study intervention was better than the existing standard?
- Was the selected outcome measurement one that previously has been used in other published research on the same topic?
- Was a measurement used that could determine outcome in the clinical setting?

STATISTICAL ANALYSIS

The methods section of any well written manuscript includes a summary description of the statistical tests used to evaluate data. The complexity of the field of biometrics and statistical analysis has increased substantially in recent years. Most physicians will not possess the knowledge of this field needed to determine if the statistical methods are fully valid. For this reason, all major scientific journals now employ specialists in analysis to help ensure valid methods in all publications. However, it is possible to understand if the "generally proper" approach has been taken by the author(s).

RESULTS

A well-written results section first describes patients involved in the study and their basic demographics. Determinations are then discussed as to whether the study groups were sufficiently similar for comparisons. The reader should be able to verify if any potential confounding variables were present that might affect a prognosis or treatment outcome and establish that baseline values of the outcome index were similar among treatment groups. All graphic summaries should be clearly labeled and appropriately scaled. Results of statistical analysis should be provided and all adverse effects should be reported. The process of analyzing any study's results should include the following 3 basic questions regarding results:

- What are the actual results?
- Are the results of the study valid?
- Will the results impact clinical care or project design?

DISCUSSION

The discussion section of a research publication provides an opportunity for the author to interpret the results and explain their clinical importance. The reader must remember that the discussion is the author's interpretation of clinical relevance, and then decide whether to agree. Statistical significance only decreases the possibility that the results could have occurred by chance alone. It implies nothing about actual clinical importance of the results. All studies potentially have design flaws or other threats to validity of the results that may affect the clinical relevance of the conclusions. Authors should always discuss a study's limitations, usually in one of the last few paragraphs, prior to the conclusion. Finally, conclusions must be justified by results and be consistent with the study objectives.

SUMMARY

Conducting scientific research is critical to the progression of medical science. One cannot practice modern medicine without a working understanding of how to interpret the medical literature and apply those principles to daily practice. The development of computerized access to databases and handheld devices are changing the face of medical practice towards an outcome oriented profile. All practicing physicians must have a practical set of skills regarding assessment of the medical literature and an underlying knowledge of how medical research is conducted and published literature is created.

REFERENCES

1. Buckwalter JA, Wright TM, Frank CB: Scientific credibility requires complete presentation of methods [editorial]. J Orthop Res 1997 Mar; 15(2): 161.
2. Canadian Medical Association: How to read clinical journals: I. why to read them and how to start reading them critically. Can Med Assoc J 1981 Mar 1; 124(5): 555-8.
3. Cuddy PG, Elenbaas RM, Elenbaas JK: Evaluating the medical literature part 1: abstract, introduction, methods. Annals of Emergency Medicine 1983; 12:9: 549-555.

4. Davis EA, Thompson C, Panacek EA: Basics of research (part 2): Reviewing the literature. Air Med J 1995; 14(2): 101-105.
5. Elenbaas JK, Cuddy PG, Elenbaas RM: Evaluating the medical literature part 3: results and discussion. Annals of Emergency Medicine 1983; 12:11: 679-686.
6. Niccoli JJ: Guide to critical evaluation of the medical literature [published erratum appears in J Am Podiatr Med Assoc 1994 Nov;84(11):585]. J Am Podiatr Med Assoc 1994 Oct; 84(10): 528-31.
7. Stewart DA: At the core of the medical literature [editorial]. Radiology 1995 Jun; 195(3): 597-9.

Numbers and the Nature of Research

"Knowledge of a matter may be obtained in two ways, by phenomenon observation, by inferential reasoning. …a brown spot on an apple can be seen with the eye; it is phenomenally given. But that there is a worm in the apple is something we may infer from the spot and from our general knowledge of apples and worms."

The Dali Lama (1)

INTRODUCTION

The goal of any scientific investigation is to evaluate a population for some characteristic of interest. Any numeric characteristic of such a population is known as a *parameter*. Examples would be the mean systolic blood pressure of a particular population or the proportion of a population the "responds favorably" to a dose of a particular antihypertensive. These parameters are then organized and analyzed by describing them and by the application of some *statistic*. A statistic is any mathematical process used on data that provides a parameter (2, 3).

SAMPLE POPULATIONS

If we are able to collect data (such as blood pressure) on every single individual (the whole population) then we could calculate an exact parameter of that population. This sort of measurement is usually not possible. To obtain the parameter of the mean systolic blood pressure of the population in Oklahoma alone would require over 3.5 million measurements. This clearly is impossible, so an alternate process needs to be employed. This alternate method is done by taking a sample of that population in such a way that any measurements done, and parameters defined, are generalizable to that entire population. Such a group selected for measurement is known as the *sample population*. Selection of a sample population is done by proper study design and randomization (if applicable). These topics will be covered in other chapters.

DATA

In research we deal with the collection of analysis of information. These bits of information are called *data* – and when analyzed properly, can allow us to make logical assumptions or inferences about the event or entity we are collecting the data on. Data come in various types – a concept that is important to understand as it limits what sort of inferences one can make with the data. In addition, the type of data dictates the type of statistical test that can be used.

DISCRETE VERSES CONTINUOUS DATA

DISCRETE DATA

Discrete data are those that have one value from a limited set of possible values. The data may be represented only by whole numbers (3). As example, the number of children born into a population during a specific time period is discrete – as it is exactly defined and represented by a whole umber – in this case there are no fractional children, only whole children.

CONTINUOUS DATA

Continuous data however may take on any value that occurs within a defined range. As example; suppose one is measuring the serum lactic acid levels from a group of long distance runners. These values can theoretically range from zero to infinity. The values may be expressed as whole numbers or fractional values – depending on the actual levels found as well as the sensitivity of the diagnostic test used. The lactic acid test value may assume any value continuously within its defined range.

NOMINAL, ORDINAL, INTERVAL, AND RATIO DATA

Along with being discrete or continuous, data also assume other characteristics that influence their nature.

NOMINAL DATA

As the term suggests, there are data consisting of "named categories." The data have no implied order among or between the categories. That is; the values do not imply that one value is "better" or "worse" than another, or that one value is "higher" or "lower" than another (3, 4). An example is the ethnicity of cases in a study population. There are a limited set of choices, but no implied order within the selections made.

A "special" type of nominal data are known as *dichotomous* data (they also go by the more nerdish name of *existential* data). As the name dichotomous implies these data take on one of two values (2, 3, 4). That is; the thing being considered either exists – or does not exist. In this case, suppose we are looking at mortality – (+ = dead and - = still alive). In this case, there are only two possible values – and one can make the case that it is preferable to be "-"and alive rather than "+" and dead. If this is accepted, then there is a difference in the ranking these values. On the other hand, suppose we are considering some test that has a color change and will either be red or blue. In this case, the value will be "R" or "B" but not ordered in any way as both R and B are equally "desirable" in this setting.

ORDINAL DATA

In the case of ordinal data, the data are organized into ordered categories – and the categories cannot be considered to be equal (2, 3, 4). An example would be the categories of severity of illness – none, mild, moderate, severe. In this setting the categories clearly are not equal in that it is better to be in the "none" category regarding disease rather than any of the other categories. In addition, it would be better to be in the "mild" disease rather than the "moderate" or "severe" categories. So clearly these categories are ordered in their nature, but there is really no direct relationship between the categories. That is, although it is clear that it is "better" to be in the "mild" disease category than the "severe" - we cannot determine "how much" better it is to be in the "mild" category.

INTERVAL DATA

In this setting, interval data values have equal "distances" or "strengths" between each other – but the zero point of the scale is arbitrary (2, 3, 4). That is; one may set the zero point wherever desired to adjust values, but the distances or relationship between each value remains constant. An example might be to add 400 points to the IQ scores of a group of individuals. Even with all the scores shifted up by 400, the relationships or distances between the scores remain the same. However, because the zero point has been arbitrarily shifted – the ratios of the values have no real meaning.

RATIO DATA

Ratio data are like interval data except in this case the zero point has meaning. Therefore the ratio of the data has meaning as well. The ratio data values have an equal and consistent interval between each other and they have a regular and consistent distance from the zero point. Examples here would be height and weight measurements.

DATA VALUES DISTRIBUTED AROUND MEANS

A property of data and the collection of it is that when the data are analyzed, the values will fall into some distribution around a mean value. This distribution around the mean is critical in the selection of the right statistical tests to analyze the data (2, 3, 4). We will consider the impacts of these distributions of data in later chapters.

SUMMARY

The selection of the sample population used in one's research in order to represent the overall population, the distribution of the data (information) around value means, and the nature of the data collected are critical in generating true observations and inferences from research.

REFERENCES

1. Dali Lama, *The Universe in a Single Atom: The Convergence of Science and Spirituality.* New York, Broadway 2006.

2. Motulsky H, Intuitive Biostatistics. Oxford University Press. 1995.

3. Norman GR, Streiner DL. Biostatistics: The Bare Essentials. 2nd Ed. PMPH USA. 2000.

4. Jekel JF, Katz DL, Elmore JG, et al. Epidemiology, Biostatistics, and Preventive Medicine. 3rd Ed. Saunders. 2007.

Bias and Confounding

We know that polls are just a collection of statistics that reflect what people are in thinking in 'reality.' And reality has a well known liberal bias.

Stephen Colbert, (to a White House correspondents' dinner, April 30, 2006).

INTRODUCTION

When one performs research, the initial products of these efforts are a set of data points. There are two major forces that can impact the nature of data collected in research studies and the nature of the subsequent analysis, these are *Bias* and *Confounding*. Although often confused, the impacts of bias and confounding on research are quite different and must be managed in different ways.

BIAS

Bias is defined as a "…systematic error in the design or conduct of a study that leads to an erroneous association between the exposure and the disease" (1). Bias will occur in any type of study, but is generally more common in retrospective studies than prospective studies, due to the design of these investigations.

EVALUATION FOR BIAS

When one evaluates for bias – three issues should be considered (2),

1. What is the source of the bias?
2. What is the magnitude of the bias?
3. What is the direction of the bias?

Determination of the source of the bias is critical in evaluation of the nature and impact of the bias on a data set. Knowing the potential sources of bias is paramount to understanding what the specific source of bias within a study might be, and in preventing or decreasing the impact of bias during the design phase of the research protocol.

One of the reasons to do research is to look for an association between variables. Estimation of the magnitude of bias is dependent on the magnitude of the association being measured, the type of bias involved, the ability of the study to measure an association, and the direction of the bias. In general, studies that indicate a large degree of association or non-association are unlikely to be impacted to a substantial degree by bias. On the contrary – studies that demonstrate a modest or weak degree of association (or non-association) are much more likely to be impacted by bias.

Bias direction can cause one to over-estimate or under-estimate the degree of association – depending on the bias direction away from or towards the null hypothesis respectively. This is demonstrated in figure 1 where the impact of bias on a frequency distribution is demonstrated. The null hypothesis (H_o) holds that there is no difference between two populations or associations between variables or outcomes, while the alternate hypothesis (H_A) – which is accepted only if there is limited or no support for the H_o - indicates that there is a difference between the two populations. Bias is generally unidirectional in that it shifts the mean value of a population in one direction or the other, but not in both directions.

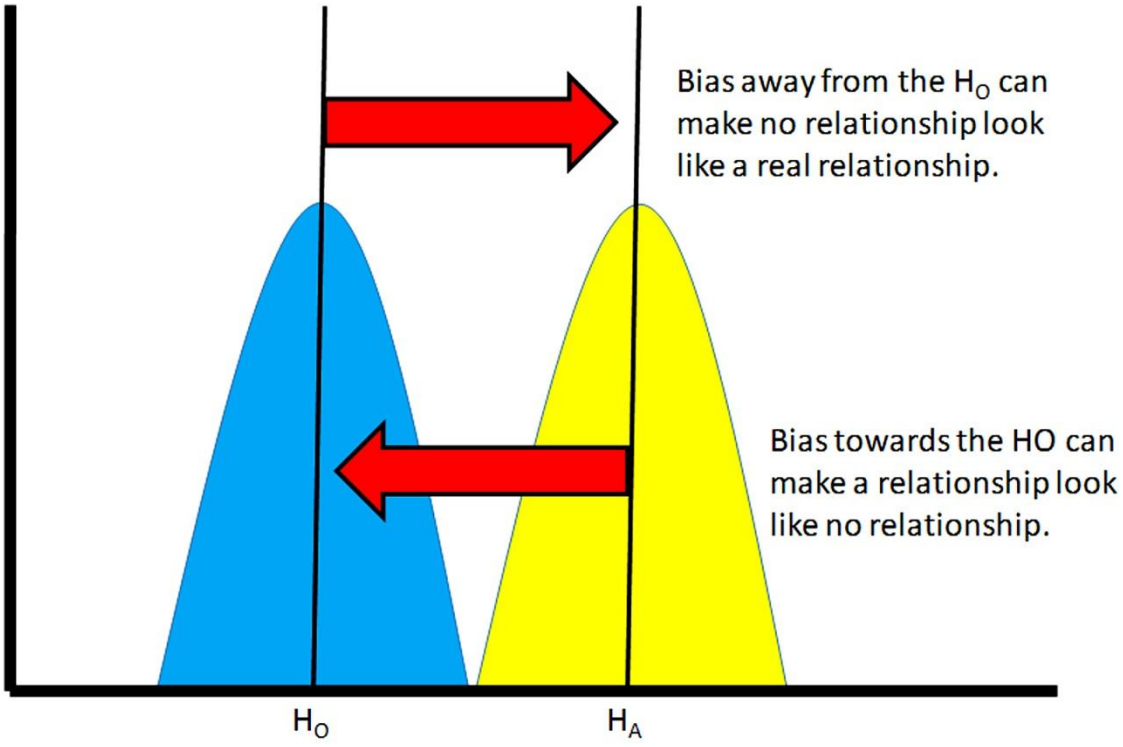


Figure 1. Bias towards and away from the null hypothesis (H_o).

BIAS CLASSIFICATION AND NATURE

Bias may find its way into research studies in a number of ways. In order to understand bias better, it may be classified in a number of ways. The primary classification of bias includes Selection Bias and Observational Bias (1). Bias may also be classified as to its association with [1] study participants, [2] the procedural process, [3] the analysis process, and [4] the development of results and conclusions (2, 3).

SPECIFIC EXAMPLES OF BIAS

The concept of bias may be more clearly understood by defining a few specific examples in the context of the study subjects, the procedural process, data analysis, results and conclusions and the experimenters (3).

BIAS ASSOCIATED WITH THE STUDY PARTICIPANTS

Healthy Worker Effect: Comparing health outcomes between healthy workers and the general population may be biased due to the fact that workers, who are actively holding down a job, often have better health to begin with and therefore often have better outcomes in similar situations when compared to the general population.

Non-response Bias: Non-respondents in a study may have exposures or outcomes that differ systematically from respondents. This is why it is important to evaluate and report on the basic characteristics if possible of both the respondents and non-respondents.

Admissions Bias (*Berkson's Bias or Paradox*): This effect can be found in hospital and clinic based studies. It occurs when hospitalization rates differ among different exposure or disease groups. The evaluation and care provided to those hospitalized and those not hospitalized can differ systematically thus introducing bias.

Membership Bias: The characteristics associated with a group membership (employed, runners, smokers, etc) may imply a degree of health that differs systematically from other groups and the general population. If not accounted for, this can introduce bias.

Volunteer Effect: Volunteers are motivated to do so for some reason. This may be associated with an underlying difference in exposures, health attitudes, or behaviors and eventual outcomes that will differ systematically from non-volunteers.

Prevalence-Incidence Bias (*Neyman Bias*): This bias can exist when, (1) a condition under investigation includes early fatalities that are missed because of a time lag between exposure and a study of its consequences (2) silent (undiscovered cases) exist or evidence of exposure disappears after the event or disease onset.

Procedure Selection Bias: The characteristics of individuals rather than random assignment are used to determine if a patient receives a particular clinical procedure (e.g. only those with a very poor prognosis are offered the experimental therapy).

BIAS IN THE PROCEDURAL PROCESS

Contamination Bias: Members of the control group inadvertently receive the experimental intervention.

Family Information Bias: Family history varies markedly depending on whether it is provided by an individual with disease or an individual without the disease.

Insensitive Measure Bias: Outcome measures (your method of measuring or detecting the outcome or endpoint) are not capable of determining disease or clinically significant changes.

Hawthorne Effect: Study participants change their behaviors simply because they are being observed in a study.

Detection Bias: Health outcomes appear more favorable because new, more sensitive diagnostic technologies detect disease sooner than older technologies. So treatment is started earlier in the stage of the disease than in previous studies. This can make a "new" method of treatment look better than the "old" method just because the disease is detected sooner and the treatment is started sooner, when in fact the "new" treatment is actually worse than the "old" one.

Diagnostic Suspicion Bias: Knowledge of the patient's exposure to a suspected risk factor influences the intensity and the outcome of the diagnostic process.

Compliance Bias: Compliance with treatments varies between groups because participants in one group find it easier or more pleasant to comply than participants in the other group.

Recall Bias: When asked to recall events, individuals in one group may be more likely than another group to remember events.

Therapeutic Personality Bias: Outcome and their measurement are influenced by the researcher convictions about the effectiveness of the intervention.

Exposure Suspicion Bias: Knowledge of a patient's disease status affects the intensity and outcome of the search for exposures.

BIAS IN ANALYZING DATA

Procrustean Data Torturing: Data that contradict a desired hypothesis are selectively suppressed.

Migration Bias: Participants in one group cross over into another.

Opportunistic Data Torturing: Data are analyzed in several different ways to detect statistical significance after which one may devise a biologically plausible hypothesis to fit the associations.

Withdrawal Bias: Participants who experience undesirable outcomes or who are lost to follow-up are "withdrawn" from the study and are omitted from the analysis.

Bogus Control Bias: Patients (in the treatment or the control group) who sicken and die before or during the administration of the experimental treatment are removed from the treatment group or reallocated to the control group.

BIAS IN THE RESULTS AND CONCLUSIONS

Significance Bias: Statistical significance is confused with biological or clinical significance.

Ecological Fallacy: The outcomes in groups are generalized inappropriately to individuals (e.g. There is a high correlation of lung cancer in a county with a cement plant operating in it. So the cement plant must be the cause of the cancer.).

Correlation Bias: Correlation (association) is equated erroneously with causation.

CONTROLLING AND ELIMINATING BIAS

The good news about bias is that it can be identified early in the study design process and minimized or eliminated almost entirely. Careful attention to study design and protocol will do much to prevent most forms of bias from entering the data. In addition, careful on-going monitoring of the research process and protocol compliance helps ensure that bias remains at a minimum during the data collection process. Little if anything can be done to reduce the impact of bias using statistical methods once it has entered the data set.

CONFOUNDING

Any time a study is performed, and especially when an association is noted between some variable(s) and an outcome, the investigator must evaluate the result for validity. This is due to the fact that there are several alternate explanations as to why one may have obtained evidence of an association. First, it may be due to random chance – which may be evaluated by statistical processes planned before data collection and executed after study completion. Second, it may be due to the effect of bias on the data – which is controlled by careful study design and adherence to good data collection procedures as discussed above. Finally, it may be due to confounding – which we will discuss now.

Confounding is defined as "the mixing of effects between exposure, and outcome, and a third extraneous variable known as a confounder" (1). As example, when one tries to make a comparison between an exposed group and a non-exposed group – one tries to balance all other risks and variables except the exposure in both groups. A confounder, in essence, changes the distribution of that risk such that it is different between the exposed and un-exposed groups. This then changes the rates of disease development or other outcomes between the two study cohorts in an often unpredictable or at least unaccounted for way. This effect changes the results such that the impact of the exposure or intervention is no longer "isolated" (or the only variable changing the disease or outcome rate).

EXAMPLE OF A CONFOUNDER

Suppose we are conducting a study to determine if there is an increase in the rate of caries in children who eat candy regularly verses those who do not. Suppose that the study is spread across a large number of communities.

Some of communities have fluoride in the water supply, some do not. The researchers are however not aware of this fact. The communities using fluoride will have a significantly lower fate of caries in both candy and non-candy consumers when compared to communities that do not.

By reducing the prevalence of cavities in candy consumers as well as non-candy consumers in some communities, fluoride acts as a confounder by reducing the overall rate of caries for the entire population more strongly in the candy consumers than the non-candy consumers. This pushes the data results towards the H_o (that there is no difference in caries between candy and non-candy consumers) and away from the H_A (that there is a difference). If the confounders are strong enough, they may entirely obscure any impact of the variable and hide a relationship – or indicate a relationship where one does not exist – depending on the direction of the confounding effect.

IDENTIFICATION AND CONTROLLING OF CONFOUNDERS

Confounding is controlled by identification of variables that can be confounders, then by controlling the confounders during the design and analysis phases of the study. In our example, this could be done by doing a separate analysis on the populations from fluoride and non-flouride using communities (a process called stratification) – assuming one knows the confounding problem exists.

IDENTIFICATION OF CONFOUNDERS

The first step is to determine if some variable associated with a study is likely to be a confounder before the study starts. This is done by carefully reviewing the previous knowledge and studies associated with the area of study one is about to undertake to identify known variables involved. Next these variables will need to meet several criteria to be able to function as a confounder within a study population. These are as follows.

There must be some association between the variable and the disease or outcome of interest. That is, the variable must be independently predictive (have an independent effect) of the disease or outcome in some way. An example might be advanced age and stroke. Although age doesn't cause stroke, the incidence of stroke in the older population is much higher than the younger population. If there is such an association, then the variable could be a confounder – if no such association exists – it cannot.

The variable must be within the population and associated with the individuals who are in the exposed cohort in some way. What this means is that everyone in the population both exposed and unexposed either has the confounding variable or could have the confounding variable in question, but the variable or its effect will be unequally distributed between the exposed and un-exposed groups. As example, suppose our exposed group has a higher proportion of older individuals than our un-exposed group – this is clearly going to alter the rates of some diseases or outcomes.

Finally, a potential confounding variable cannot be – itself – an intermediate step in the pathway of cause between the exposure and the disease. This means that the confounding variable must be outside of the direct chain of causation between exposure and disease (otherwise it would just be part of the causation event associated with the exposure and not result in much variation in the rate of the disease).

CONTROLLING CONFOUNDERS

Once a known variable has been identified as a potential confounder it may be controlled for during the design phase of the study by randomization, restriction of entry in the study with selection criteria, or case-control matching methods. Confounding may also be controlled for during the analysis phase of a study by standardization protocols, stratification of analysis based on specific characteristics, matched analysis in case-control studies, or use of multivariate analysis methods in some instances.

An additional problem with confounders is when science itself is unaware of the presence of a potential confounder. Unidentified confounders cannot be controlled for. However, their impact may be noted when results of a study do not go as expected. This event provides scientists with another question that requires investigation – and potentially results in expansion of human knowledge.

SUMMARY

Bias and confounding are important effects in research that can impact the collection and analysis of data substantially. Both bias and confounding however can be identified and minimized in the conduct of research.

The primary way to prevent the impact of bias is to prevent it entering the study through early recognition and proper design of the study. Once bias has impacted the data set collected in a study, little may be done to correct its impact.

Variables that can serve as confounders may also be identified and controlled for in research by careful study design and analysis. Confounders may be controlled for in the design phase of the research protocol by randomization, case selection criteria, or case-control methods. In the analysis phase confounding may be minimized by stratification, matched analysis, or use of multivariate analysis.

REFERENCES

1. Aschengrau A, Seage GR III (2008) Essentials of Epidemiology in Public Health. (2nd edition) Jones and Bartlett Publishers. Sudsbury MS.
2. Essex-Sorlie D (1995) Medical Biostatistics & Epidemiology: Examination & Board Review. (1st edition) Appleton & Lange East Norwalk, CT.
3. Jekel FJ, Katz DL, Elmore JG (2001) *Epidemiology, Biostatistics, and Preventive Medicine* (2nd edition) W.B. Saunders Co. Philadelphia, PA

Ethics in Research

"To the man who is truly ethical all life is sacred, including that which from the human point of view seems lower. Man makes distinctions...under the pressure of necessity, as for example, when it falls to him to decide which of two lives he must sacrifice in order to preserve the other. But all through this...he...knows that he bears the responsibility for the life that is sacrificed."

Albert Schweitzer MD (1)

INTRODUCTION

Human beings are ethical creatures. Given all the discord and trouble in the world this may seem like an irrational statement that deserves an explanation. The development of ethical behavior and the moral conduct in human beings has been a topic of controversy for centuries. At what point in time did an individual human begin to consider the impact of its own behavior and actions upon other humans? Why do we see in our society individuals willing to sacrifice their own lives for the benefit of others they may not even know? We have no clear answers to date as to when and how ethical behavior developed in human beings, but we have learned a great deal about how these ethical and moral concepts cause us to behave.

Because humans are not merely logical automatons, our species, as a rule, have certain limits we place on our behavior towards other humans. However history has shown us that we have a tendency to violate such limits – even for the best of intentions. These violations often cascade us down a dark path of human behavior. As such, ethics in research is a critical component of the overall process. In this chapter we will discuss the role of ethics in research as well as various regulatory processes employed to ensure ethical behaviors.

ACCELERATION OF BIOMEDICAL INFORMATION

In 1854 the modern age of health and epidemiology began with the investigation of a London cholera epidemic by John Snow (2). Using interviews and a mapping of the homes of cholera cases, the source of the cholera outbreak was traced to a specific water source in the city. The pump handle was removed and the epidemic abated.

Whereas the health improvements of the late 19th and early 20th centuries were usually matters of sanitation and education, the dynamic acceleration in biomedical and medical capabilities in the late 20th and early 21st centuries pose entirely new problems and ethical issues. Issues of genetic testing, genetic screening, and genetic manipulation are concerned with the very core of life itself and the potential creation of entirely new "genetic minorities" subject to discrimination. As intelligent sentient creatures humans have had little enough time to come to terms with their own nature, before being faced with the ability to alter that nature at will.

LIFE AND DEATH

The definition of "death" and the associated "rights" of the individual to make death related choices with or without the direct assistance of a physician are under intense consideration. Not just from the point of view of death as an end but also because of the application of technologies to prolong life, extend health, clone new organs, regenerate dying cells and the like. The complexity of the ethical-moral-theological issues associated with death and life-extending technology has grown astronomically in the past century.

PRIVACY

Medical informatics has expanded to the point that health information confidentiality can little be guaranteed. Coupling this with genetic screening, emerging diseases and life style choices the potential for various new "medical minorities" becomes clear – along with the associated potential for discrimination.

THE THEORETICAL BASIS OF ETHICS

The general topic of the theoretical basis of ethics is impossible to cover in any single discussion. What follows is a brief description of the major operational theories behind most ethical processes. For the most part, ethicists consider the theories of Deontology (duty based ethics) as outlined by Kant and that of Utilitarianism attributed to John Mill and Jeremy Bentham as the core of theoretical work upon which much ethical consideration is based. It is the dynamic interplay between these two theoretical approaches that is often relied on for illumination of ethical considerations.

DEONTOLOGY

Deontology by name is associated with a "duty" regarding a particular issue. This duty is not conditional on the probable outcome of any particular decision. The duty rather is dependent on the input elements associated with the decision. That is, Kant opined, because we cannot fully control the future – therefore the consequences of our decisions – we must judge the conditions under which the decision must be made at the start of the decision making process. The decision must be based only on the input conditions – not the potential or hypothetical outcomes. This is done by application of the *Categorical Imperative Principle* according to Kant. Very briefly, the categorical imperative means that you consider what the world would be like if everyone made the decision you are about to make.

In addition, Deontological Theory places emphasis on individual human rights (essentially as inviolate). The primary weakness of the Deontological approach are that it is a binary decision process not allowing for middle or alternate solutions and that it can actually generate moral dilemmas when ethical duties conflict (3).

CONSEQUENTIALIST ETHICAL THEORY (UTILITARIANISM)

Consequentialist Ethical Theory as embodied in the Utilitarian ideas usually attributed to John Mill and Jeremy Bentham takes a decidedly alternate approach to ethics. Whereas Deontology considers the input side of decisions to the exclusion of potential outcomes or consequences, consequentialism specifically considers the potential consequences of a decision. An abbreviated description of this theoretical approach is that one starts with a single goal in ethical decisions – that of causing the "greatest good for the greatest number." Decisions are then made with the objective of achieving that goal. As such, conseqentialism places its focus on the possible or hypothetical results of an ethical decision in a goal directed manner (4).

Utilitarianism (the primary form of consequentialism) allows for "gray zones" in decision making. It recognizes human individual rights – but places them in a different context in relation to the collective. That is, each individual is to count as one but no one as more than one if strict egalitarianism is observed. The major weaknesses with this theoretical approach is, as Kant stated that we cannot be sure of the future consequences of our decisions and actions. In addition, it is difficult to determine or predict what action will cause the greatest good. Finally, this theoretical approach can be somewhat caviler with individual human rights – actually allowing harm to some individuals if it is anticipated that the "greater good" will be served by such.

A THEORY OF JUSTICE

The work of contemporary philosopher John Rawls culminated in his "A Theory of Justice." Rawls used elements of both Deontology and Utilitarianism in his ethical moral theoretical constructs although the theory is not entirely dependent on either (5). Initially concerned with the ethical and moral evaluation of political and social institutions, the theory is increasingly being applied to specific ethical issues in health and medicine.

The construct of this theory is deceivingly simple, yet expands in complexity as complex ethical issues are considered. There are two primary principles associated with the Theory of Justice.

THE LIBERTY PRINCIPLE

Each person is to have an equal right to the most extensive basic liberty compatible with similar liberty available to others. This principle provides the basic tenant of respect for individuals as the baseline standard for ethical maxims.

THE DIFFERENCE PRINCIPLE

Social and economic inequalities are to be arranged so that they are both reasonably expected to be to everyone's advantage and attached to positions and offices open to everyone. Because people in the "real world" exhibit significant differences in social and economic status, maxims in ethical decision making should be (as much as possible) to benefit as many individuals as possible. In addition, the individuals involved in the ethical decision making process (as much as possible) should be open to everyone in the society. This is done in an effort to minimize the differences between individuals and groups.

HISTORICAL NOTES ON RESEARCH ABUSE

All humans are governed by a set of ethical behaviors from which they will individually and as a society derive moral philosophy. For the most part the definitions of ethical behavior and morality have revolved around theological and secular discussions of spiritual and logical reasons as to why people behave in a "good" or "bad" way. Recent research has removed some of this discussion from the realm of philosophy and into the hard science of biology, neurochemistry, and mathematics. Such scientific undertakings of late indicate that behaving in a "good' or altruistic manner is a genetic trait of the human species providing an evolutionary advantage. Indeed, it seems that human beings may actually be "wired" neurologically to be, for the most part, altruistic towards each other (6).

This innate altruistic behavior however, does not mean that all humans will behave in a compassionate or caring manner to all other people. All human behavior is modulated by ethical concepts as understood by the individual. Ethics itself is defined as study of human behavior involved in systematizing, defending, and recommending concepts of "right" and "wrong" behavior. A Moral Philosophy is the derivation of beliefs and behavior within an individual or society, towards the good or the bad based on the understanding of ethics. We have learned through history that human beings may behave in "unethical" (amoral) ways. These activities may even be contrary to individual and group human survival and against the general moral concepts of human society. Conversely ethical (moral) humans may also perpetrate acts of an "unethical" (amoral) nature based on their beliefs while following a highly developed moral philosophy.

Unfortunately although human beings often strive to follow ethical courses and some form of moral code in their lives, this does not ensure "humane" behavior. Many great philosophers and theologians have grappled with the problem of human duality in ethical behavior. In our modern era this apparent opposition of human nature has been restated by Stuart Stevens, "There's a dark side to each and every human soul. We wish we were Obi-Wan Kenobi, and for the most part we are, but there's a little Darth Vader in all of us" (7).

Substantial strides in the improvement of the human condition have been led by individuals with great ethical character. As human beings, we are capable of great compassion and beauty. The use of scientific methodology for the betterment of the human condition represents one of the most altruistic and pure behaviors of which human beings are capable. Surely those involved in scientific endeavors for the betterment of mankind are of good moral fiber and ethical strength? The truth of the matter is that human beings can convince themselves that almost any behavior is justified by their moral philosophy. Because of this, civilized societies have developed ethical guidelines outlining a moral philosophical approach to conducting human medical research.

There have been many abuses of research in the history of medicine. Some of the more dramatic examples are found in the Nazi physician experimentation on concentration camp prisoners. Doctors have been thought of as the protectors of mankind, the healers, and caretakers of our existence. Ancient civilizations revered the medicine men as having special power to protect life. The trust between a physician and patient is sacred. This is why the practice of medicine by some of the doctors of the Third Reich is so egregious, outrageous, and shocking when examined. The Nazi doctors violated the trust placed in them by humanity resulting in one of the most horrible

episodes in medical history to date. A most painful reality is that many of the doctors perpetrating these actions during WWII escaped their crimes against Humanity and lived a relatively normal life following the war (8).

NAZI ABUSES OF RESEARCH

During WWII a number of German physicians conducted experiments on thousands of concentration camp prisoners without their consent. Unethical medical experimentation carried out during the Third Reich has been divided into three categories. The first category consists of experiments aimed at facilitating the survival of Axis military personnel. In Dachau, physicians from the German air force and from the German Experimental Institution for Aviation conducted high-altitude experiments, using a low-pressure chamber, to determine the maximum altitude from which crews of damaged aircraft could parachute to safety (9, 10). Low temperature experiments were also carried out using prisoners to test the physiologic responses and find an effective treatment for hypothermia. In addition, prisoners were used to test methods of making seawater potable (11, 12).

The second category of experimentation was directed at creating and testing pharmaceuticals as well as developing trauma and medical protocols for injuries and illnesses encountered in the field by the German military and civilian populations. At the concentration camps of Sachsenhausen, Dachau, Natzweiler, Buchenwald, and Neuengamme, physicians tested immunizations and antisera for malaria, typhus, tuberculosis, typhoid fever, yellow fever, and infectious hepatitis which included exposing prisoners to these agents (13, 14). The Ravensbrueck camp was the site of bone-grafting experiments as well as tests of the efficacy of sulfanilamides. At Natzweiler and Sachsenhausen, prisoners were subjected to phosgene and mustard gas to test protective equipment, antidotes, and treatments (13, 14).

The third category of medical experimentation sought to advance the racial and ideological views of the Nazis (15). The most infamous were the experiments of Josef Mengele at Auschwitz. Mengele conducted medical experiments on twins. He also directed serological experiments on Roma (Gypsies), as did Werner Fischer at Sachsenhausen, in order to determine how different "races" withstood various contagious diseases (10, 16, 17).

JAPANESE RESEARCH ABUSES

Such medical atrocities were not limited to German physicians. Recent evidence over the last few years has brought to light the actions of Japanese physicians during the war. In the infamous Unit 731 unbelievable atrocities were committed in the name of medicine and science against prisoners and subjugated populations (18). This was only the first of such revelations to come out since the war. The anatomy department of Kyushu University was another site actually within the borders of Japan where downed American airmen were dissected while alive and awake for the purposes of “science” (19). A young physician at the time Dr. Toshio Toni recalled in an interview 50 years after the experiments, "*I could never again wear a white smock. It's because the prisoners thought that we were doctors, since they could see the white smocks, that they didn't struggle. They never dreamed they would be dissected*” (19). To date, only a few of the physicians known to have participated in these egregious acts against humanity have paid any penalty following the war (20).

AMERICAN RESEARCH ABUSES

Lest we feel too smug and superior in our moral ideologies there are plenty of examples of unethical behavior of agencies and physicians here in the United States. From 1932 to 1972, the *United States Public Health Service* (**USPHS**) conducted an experiment on 399 black men with tertiary syphilis. These men, illiterate sharecroppers from one of the poorest counties in Alabama, were never told what they were suffering from or of its seriousness. They were told only that they were being treated for “bad blood,” their doctors had no intention of curing them of syphilis. The data for the experiment was to be collected from autopsies, and they were thus deliberately left to degenerate and die from tertiary syphilis (21).

One of the most disturbing aspects of the experiment was how the USPHS kept the test subjects from receiving treatment. When a number of nationwide campaigns to eradicate venereal disease came to Macon County, the men were prevented from participating. Even when penicillin was discovered in the 1940s the Tuskegee men were denied the drug. During WWII, 250 of the men registered for the draft and were ordered to get treatment for syphilis, only to have the USPHS exempt them from active military service. The experiment continued in spite of passage of the *Henderson Act* in 1943, requiring the testing and treatment for venereal disease (21).

Other violations of trust in medical research have been discovered over the last few years here in the United States. All these studies seem to have been carried out under the best intentions, justified by "national security", or the "greater good". In preparing America for nuclear attack during the Cold War years following WWII, thousands of US citizens became the subjects of over 4,000 secret and classified radiation experiments conducted by the *United States Atomic Energy Commission* (**USAEC**) and numerous other government agencies. A secret USAEC document, dated 17 April 1947, reveals that physicians were aware of these radiation hazards but simply ignored them. Under the title "Medical Experiments in Humans," the memorandum read: "It is desired that no document be released which refers to experiments with humans that might have an adverse effect on public opinion or result in legal suits. Documents covering such field work should be classified 'Secret' (22, 23). Some of the classified government radiation experiments included:

- Exposing more than 100 Alaskan villagers to radioactive iodine during the 1960s.
- Feeding 49 retarded and institutionalized teenagers radioactive iron and calcium in their cereal during the years 1946-1954.
- Exposing about 800 pregnant women in the late 1940s to radioactive iron to determine the effect on the fetus.
- Injecting 7 newborns (six were Black) with radioactive iodine.
- Exposing the testicles of more than 100 prisoners to cancer-causing doses of radiation. This experimentation continued into the early 1970s.
- Exposing almost 200 cancer patients to high levels of radiation from cesium and cobalt. The AEC finally stopped this experiment in 1974.
- Administering radioactive material to psychiatric patients in San Francisco and to prisoners in San Quentin.
- Administering massive doses of full body radiation to cancer patients hospitalized at the General Hospital in Cincinnati, Baylor College in Houston, Memorial Sloan-Kettering in New York City, and the US Naval Hospital in Bethesda, during the 1950s and 1960s. The experiment provided data to the military concerning how a nuclear attack might affect its troops.
- Exposing 29 patients, some with rheumatoid arthritis, to total body irradiation (100-300 Rad dose) to obtain data for the military. This was conducted at the University Of California Hospital in San Francisco. (24)

A recent example of controversy is the United States sponsored AIDS drug research being performed in Kenya, Africa. This study is a double blind prospective analysis of the efficacy of AZT in reducing transmission by various means from mother to child (primarily by breast feeding). As such, the study requires a control group to receive a placebo instead of the actual AZT. This runs counter to the basic ethical principle that a placebo may only be used when there is no known effective treatment (which is part of the international agreements covered later in this chapter). It has been established since 1994 that AZT sharply reduces transmission of the HIV virus from mothers to babies in-utero. The argument is that it is unethical to withhold a drug known to be effective in other aspects of HIV progression and transmission to evaluate another aspect of the drug (25).

THE SEVEN DEADLY SINS OF RESEARCH

There are 7 common errors that most frequently cause physicians to fall into trouble with regulatory agencies regarding research. Avoidance of these will go a long way to improving the quality of your research, and perhaps make you a more "ethical human" in the process.

1. *Arrogance*: the patient wouldn't understand
2. *Paternalism*: doctors know what's best
3. *Immorality*: misrepresenting risks and benefits
4. *Vanity*: multiple redundant papers, foolish research, getting your name in print
5. *Irresponsibility*: careless research, sloppy adherence to methods
6. *Deception*: "fudging" data, making it fit the desired picture
7. *Greed*: for research dollars

REGULATION OF RESEARCH IN THE UNITED STATES

Regulation of research is clearly necessary considering the history of medical research abuse. Other factors play a role as well such as ignorance of the law and individual rights, which could jeopardize researchers and subjects alike. Regulation and guidelines had to be established to assist even the most well meaning researcher from going down the wrong path even if for the best possible reasons. Research regulation, although not perfect, hopes to assure that the research being performed is appropriate and well designed resulting in expeditious and useful results. Regulation also helps to assure the public that the research performed is necessary and fundamentally sound scientifically.

REGULATING AGENCIES

Currently there are two bodies which regulate medical research here in the United States. They are the *Federal Food and Drug Administration* (**FDA**) and the *Department of Health and Human Services* (**DHHS**). Each agency has been tasked with specific duties within the framework of federal law regarding the conducting of medical and human subject research in the United States. They function under a mandate provided by the authority of federal laws and in general monitor and regulate the activities of various research groups. The regulations for conducting medical and human research studies are clearly spelled out by these agencies.

THE BASIS OF RESEARCH REGULATIONS

Much of the existing laws and regulations regarding research have been derived by international activities pertaining to human research since the end of WWII. These activities include the Nuremberg Code, The Declaration of Helsinki, The National Research Act, and The Belmont Report. Each of these will be briefly discussed.

THE NUREMBURG CODE

Following the realization of serious medical misconduct in human research during WWII a series of world wide conferences have been organized to outline generally accepted guidelines for medical research. One of the first was during the *Nuremberg war trials* following the defeat of Nazi Germany. The set of recommendations coming out of the trials of physicians and scientists involved in unethical Nazi research became known as the *Nuremberg Code*. The Nuremberg Code thus established the ground rules for ethical research almost 50 years ago. Since that time researchers have been required to obtain informed consent from people participating in studies. Putting this requirement into practice however, stirs up a swarm of difficult questions. How much information does a person need to be informed? Do the study participants really understand the information, and thus, is their consent genuine? The basic tenants outlined by the Nuremberg Code are as follows.

- Voluntary consent is ESSENTIAL

- The researcher is responsible for the consent of the subject
- The research should be necessary
- The data should be unprocurable by other methods
- No research that purposely causes injury to the subjects is allowable
- The degree of risk must be less than the value of the research
- Researchers must be qualified to perform the research
- Subjects have the right to withdraw at any time

THE DECLARATION OF HELSINKI

Started by discussions within the *World Medical Association* and finally adopted as the *Declaration of Helsinki* in 1964 at the 18th World Medical Assembly in Helsinki, Finland the so called, Helsinki Declaration took further steps in safeguarding human participation in medical research. The declaration has undergone steady revisions over the years but has several essential elements that have remained consistent in addition to the elements of the Nuremberg Code (26). Although this is an extensive document, the core concepts are as follows.

1. In the purely scientific application of medical research carried out on a human being, it is the duty of the physician to remain the protector of the life and health of that person on whom biomedical research is being carried out.
2. The subjects should be volunteers - either healthy person or patients for whom the experimental design is not related to the patient's illness.
3. The investigator of the investigating team should discontinue the research if in his/her or their judgment it may, if continued, be harmful to the individual.
4. In research on man, the interest of science and society should never take precedence over considerations related to the well-being of the subject.

THE NATIONAL RESEARCH ACT OF 1974

Due to the adverse publicity from the Tuskegee Syphilis Study in the United States, *The National Research Act of 1974* was passed. The National Research Act created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. One of the charges to the Commission was to identify the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects, and to develop guidelines which should be followed to assure that such research is conducted in accordance with those principles (27).

The Commission was directed to consider a number of issues including;

- The boundaries between biomedical and behavioral research and the accepted and routine practice of medicine,
- The role of assessment of risk-benefit criteria in the determination of the appropriateness of research involving human subjects
- Appropriate guidelines for the selection of human subjects for participation in such research
- The nature and definition of informed consent in various research settings.
- The National Research Act codified the requirement that human subjects in research must be protected and set the stage for the issuance of the Belmont Report.

THE BELMONT REPORT

On September 30, 1978, *The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research* submitted its report entitled "*The Belmont Report: Ethical Principles and Guidelines for the*

Protection of Human Subjects of Research." The Report, named after the Belmont Conference Center at the Smithsonian Institution where the discussions which resulted in its formulation were begun, sets forth the basic ethical principles underlying the acceptable conduct of research involving human subjects. Those principles, respect for persons, beneficence, and justice, are now accepted as the three quintessential requirements for the ethical conduct of research involving human subjects (28).

Respect for persons involves the recognition of the personal dignity and autonomy of individuals, and special protection of those persons with diminished autonomy. *Beneficence* entails an obligation to protect persons from harm by maximizing anticipated benefits and minimizing possible risks of harm. *Justice* requires that the benefits and burdens of research be distributed fairly.

The Report also describes how these principles apply to the conduct of research. Specifically, the principle of respect for persons underlies the need to obtain informed consent; the principle of beneficence underlies the need to engage in a risk/benefit analysis and to minimize risks; and the principle of justice requires that subjects be fairly selected. Another important outcome of the commission was the definition and application of the principle of *Clinical Equipoise* in human medical research as defined by Dr Benjamin Freedman. Clinical equipoise essentially means that the risk to the patient in any arm of the study should be acceptable, treatment arms should be similar and developed based on earlier phase data, and both treatments should have a reasonable opportunity for benefit. The utilization of clinical equipoise in clinical research is defined as;

" *[a]t the start of the trial, there must be a state of clinical equipoise regarding the merits of the regimens to be tested, and the trial must be designed in such a way as to make it reasonable to expect that, if it is successfully conducted, clinical equipoise will be disturbed"* (29). In other words, the researchers should find that there is a difference between one research arm and another if it exists and the research is properly conducted. This is nothing more that assuring minimal risk to the patient and maximum scientific rigor in study design to detect differences in study populations.

THE INSTITUTIONAL REVIEW BOARD (IRB)

An *Institutional Review Board* (**IRB**) is the local board which authorizes research on human subjects. Each IRB is usually located at the hospital where the research is to be conducted. A local IRB is required by federal law at all facilities receiving federal funds for research. All human subject research requires IRB approval. Animal research is covered by the *Institutional Animal Care and Use Committee* (**IACUC**). The purpose of an IRB is to:

- safeguard subject welfare and privacy
- apply local standards to general ethics
- ensure federal compliance
- protect the institution

IRB FEDERAL POLICY REQUIREMENTS

The Federal Policy requires that IRBs have at least five members, with varying backgrounds to promote complete and adequate review of research activities commonly conducted by the institution. The IRB must be sufficiently qualified through the experience and expertise of its members and the diversity of their backgrounds, including considerations of their racial and cultural heritage and their sensitivity to issues such as community attitudes, to promote respect for its advice and counsel in safeguarding the rights and welfare of human subjects (30).

In addition to possessing the professional competence necessary to review specific research activities, the IRB must be able to determine the acceptability of proposed research in terms of institutional commitments and

regulations, applicable law, and standards of professional conduct and practice. The IRB must obviously include persons knowledgeable in these areas. No IRB, however, may consist entirely of members of one profession.

INFORMED CONSENT

The heart and soul of conducting clinical research revolves around Informed Consent of human participants to such research. There are very specific Federal Regulations in the United States that outline the requirements of informed consent (45 CFR 46.116, 45 CFR 46.117) (31). The process of obtaining and documenting informed consent must comply with these regulations. These regulations are intended to assure that such a fully informed patient can effectively participate in choices about their health care. Informed consent originates from the legal and ethical right the patient has to direct what happens to their body and from the ethical duty of the physician to involve the patient in their own health care.

THE ELEMENTS OF FULL INFORMED CONSENT

The most important goal of informed consent is that the patient has an opportunity to be an informed participant in their health care decisions. It is generally accepted that complete informed consent includes a discussion of the following elements:

1. *Disclosure*: The potential participant must be informed as fully as possible of the nature and purpose of the research, the procedures to be used, the expected benefits to the participant and/or society, the potential of reasonably foreseeable risks, stresses, and discomforts, and alternatives to participating in the research. There should also be a statement that describes procedures in place to ensure the confidentiality or anonymity of the participant. The informed consent document must also disclose what compensation and medical treatment are available in the case of a research-related injury. The document should make it clear whom to contact with questions about the research study, research subjects' rights, and in case of injury.
2. *Understanding*: The participant must understand what has been explained and must be given the opportunity to ask questions and have them answered by one of the investigators. The informed consent document must be written in lay language, avoiding any technical jargon.
3. *Voluntarism*: The participant's consent to participate in the research must be voluntary, free of any coercion or promises of benefits unlikely to result from participation.
4. *Competence*: The participant must be competent to give consent. If the participant is not competent due to mental status, disease, or emergency, a designated surrogate may provide consent if it is in the participant's best interest to participate. In certain emergency cases, consent may be waived due to the lack of a competent participant and a surrogate.
5. *Consent*: The potential human subject must authorize their participation in the research study, preferably in writing, although at times an oral consent or assent may be more appropriate.

In order for the patient's consent to be valid, they must be considered competent to make the decision and their consent must be voluntary. It is easy for coercive situations to arise in medicine as patients often feel powerless and vulnerable, particularly in emergency situations. To encourage voluntarism, the physician can make clear to the patient that they are participating in a decision, not just signing a form. With this understanding, the informed consent process should be seen as an invitation to participate in a health care decision. The physician is also obligated to provide a recommendation and share their reasoning process with the patient comprehension on the part of the patient is as important as the information provided. The discussion (invitation to participate) should be carried on in layperson's terms and the patient's understanding should be assessed along the way.

Basic consent entails letting the patient know what you would like to do and asking them if that will be all right. The level of intensity to this process will vary based on the type of research being carried out. For example, basic consent (covering just the core issues as noted above) is appropriate, when drawing blood during treatment,

particularly if this was to be done for usual diagnostic purposes anyway. Decisions that merit this sort of basic informed consent process require a low-level of patient involvement because there is a high-level of community consensus.

ADEQUATE INFORMED CONSENT

When does one know that they have adequately informed the patient? There are three measures that are generally suggested by the literature (30).

THE REASONABLE PHYSICIAN STANDARD

What would a typical physician say about this intervention? This standard allows the physician to determine what information is appropriate to disclose. However, it is probably not enough. Most research in this area of consent indicates that the typical physician tells the patient very little. This standard is also generally considered inconsistent with the goals of informed consent as the focus is on the physician rather than on what the patient needs to know.

THE REASONABLE PATIENT STANDARD

What would the average patient need to know in order to be an informed participant in the decision? This standard focuses on considering what a patient would need to know in order to understand the decision at hand. However, this standard alone may be difficult as it puts the physician in the place of the layperson but being unable to divest themselves of the knowledge and insight of a physician.

THE SUBJECTIVE STANDARD

What would this patient need to know and understand in order to make an informed decision? This standard is the most challenging to incorporate into practice, since it requires tailoring information to each patient. But it probably provides the best overall informed consent for complex research protocols.

Most states have legislation or legal cases that determine the required standard for informed consent. The State of Oklahoma has adopted the Subjective Standard for obtaining informed consent in medical research (32). According to 21 CFR 50.20, "*no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative*" (31). The potential participant must be given the opportunity to give full consideration regarding the decision whether or not to participate in the research study without undue influence from his or her physician, family, or the scientific investigator. No informed consent may contain any exculpatory language by which the participant waives any legal rights or releases the investigator or sponsor from liability for negligence. The best approach to the question of how much information is enough is one that meets both your professional obligation to provide the best care as well as respecting the dignity of the patient while providing them with the right to have a voice in their health care decisions.

EMERGENCY MEDICINE RESEARCH CONSENT

Emergency Medicine is faced with a unique set of challenges when it comes to some types of research. As emergency physicians we often treat patients both in the hospital and via paramedic medical control who are in extremis and require invasive resuscitative measures. As such, we are ethically bound to determine the efficacy of the resuscitative methods currently used as well as develop improved life saving modalities (33). Unfortunately due to the nature of the circumstances in which such severe life threatening conditions occur, patients are often in coma or unable to consent to research and families are usually not readily available. Recognizing this as a serious impediment to scientific advancement in all areas of resuscitation research, new regulations were passed in 1996 (21 CFR 20 and 21 CFR 814) permitting some instances where consent could be wavered for such research (34).

THE EMERGENCY EXEMPTION "WAVER"

IRBs must follow strict guidelines to approve a project for an emergency "waiver." The situation must meet "and the IRB must document" the following criteria (35);

- The human subject faces a situation that is life-threatening, and the research is necessary to determine "the safety and effectiveness of particular interventions."
- Informed consent is not feasible; for example, the patient is in a coma. An intervention is needed before consent from an authorized representative is possible.
- Participation is in the subject's best interest.
- The research "could not practicably be carried out without the waiver."
- The researcher must have IRB-approved procedures and guidance to use when providing a family member the opportunity to object to a subject's participation in the emergency research.
- Community consultation requirement has been met. This requirement involves advising the community from which the subjects are drawn about the plans for the study and the risks and benefits of the research trial. After the study is completed, researchers must report the findings to the community. The demographics of the human subjects and the study results are provided at this time. Again, the requirements stipulate that the public disclosure be completed both before the project starts and after it has been completed.

ADDITIONAL IRB DUTIES UNDER EMERGENCY EXEMPTION WAVERS

The IRB is required to ensure that the subject is informed as soon as possible of their inclusion in the research. The communication of research details, as well as the usual requirements for informed consent that would have taken place under normal circumstances, are then performed. If the subject cannot be informed, the legally authorized guardian must be informed, and then he or she may withdraw the subject from the investigation (25,30). The institutional review board must consider community feedback from community consultation and the adequacy of public disclosure in deciding whether to approve the proposed study. In addition, all studies that use the emergency exception must be overseen by an independent data and safety monitoring board (30).

If the legally authorized guardian is not available, a family member has the same rights. The new regulations have changed the definition of "family member" to include "any individual related by blood or affinity whose close association with the subject is the equivalent of a family relationship." An unmarried partner or other significant person has rights to "speak" for the person unable to give consent within the "therapeutic window." The Federal regulations fully explain the "*therapeutic window*" as the short period of time in emergency situations where the medical intervention or research could save a person's life (30). Unfortunately even with the new regulations very few institutions have been able to conduct serious research in the prehospital arena or in resuscitation medicine. Further discussions with federal agencies and communities are hoped to improve the process of research in this area.

EXEMPTION FROM CONSENT

Some research with humans is eligible for "exempt" status from the IRB. If your research is part of a routine educational experience, an observational non-interventional study and if your participants will remain completely anonymous (with no identifying code to link them to their identity), you may apply to the IRB for a certificate of exemption. Your study proposal will still be reviewed by a member of the IRB, but the application process is considerably shorter, even though, the IRB process will not necessarily be "faster (32)."

EXPEDITED IRB REVIEW

A study may also qualify for *expedited review* if an IRB reviewer determines that it meets assessment criteria for minimal risk, and involves only procedures that are commonly done in clinical settings, such as taking hair, saliva,

excreta or small amounts of blood. A study that qualifies for expedited review is still held to the same standards used in full board review, but the approval process may (but not always) take less time (33).

FINANCIAL CONSIDERATIONS

Financial considerations can pose a substantial ethical dilemma to medical research and should always be approached cautiously. For investigators, the primary ethical consideration is avoiding conflict of interest. Obviously no amount of money should override the three principles of the Belmont Report outlined earlier (beneficence, justice, respect). Examples of compensation which can pose ethical problems are finder's fees, research funds, trips, and academic promotion.

For subjects, the primary concern is to provide incentives, not coercion. Financial compensation shouldn't be so large as to compromise voluntary choice. Conversely it should be commensurate with the discomfort suffered by the subject. It should not be contingent on completion of the study but can be given incrementally over the course of the study.

SUMMARY

Altruism, moral philosophy, and ethical behaviors are all part of the human condition. It is also part of the human condition to alter these behaviors to suit our own wants and needs. Medical research represents one of the highest activities practiced by humans, but also one that can potentially result in violations of basic human rights. Because of this, regulation of research by federal statutes and monitoring of research activities by various agencies are needed.

As we seek to discover the truth of the universe in which we exist, we cannot fail but be struck by the limitations of our intellect and technology. At the same time we must be inspired by the power and depth of a simple quality of our species - the human spirit.

REFERENCES

1. Schweitzer A: Out *of My Life and Thought: An Autobiography*. The John Hopkins University Press, Baltimore, MD. 2009.
2. Aschengrau A, Seage GR III: Essentials of Epidemiology in Public Health (2nd ED) Jones and Bartlett Publishers Inc. Sudbury, MA. 2008.
3. Kant I. *Critique of Practical Reason*. CreateSpace 2010 Amazon.Com
4. Mill JS. *Utilitarianism*. Dover 2007 Seattle WA.
5. Rawls JS. *A Theory of Justice*. Belknap Press – Harvard University Press 2005 Princeton MA
6. Ridley M: *The Origins of Virtue: Human Instincts and the Evolution of Cooperation*. Penguin Books LTD. London,1998
7. Stuart Stevens, Northern Exposure, Jules et Joel, 1991
8. Aziz, P: *Doctors of Death. Geneva*: Ferni Publishers, 1976. (R 853 .H8 A9513 1976)
9. Annas, G J., Michael GA, eds.: *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation.* New York: Oxford Univeristy Press, 1992.
10. Lerner, BH, Rothman DJ: Medicine and the Holocaust: Learning More of the Lessons. Annals of Internal Medicine, 122(10), 15 May 1995, 793-794.
11. Angell, Marcia. "The Nazi Hypothermia Experiments and Unethical Research Today," in New England Journal of Medicine, 322(20), May 17, 1990, 1462-1464.

12. Berger RL: Nazi Science: The Dachau Hypothermia Experiments. New England Journal of Medicine 322(20), May 17, 1990, 1435-1440.
13. Boozer JS: Children of Hippocrates: Doctors in Nazi Germany, in I. Shur, F. Littell, and M. Wolfgang, eds. Reflections on the Holocaust. Special issue of The Annals of the American Academy of Political and Social Science, 450 (July 1980), 83-97.
14. Bernadac C: *Devil's Doctors: Medical Experiments on Human Subjects in the Concentration Camps.* Geneva: Ferni Publishing House, 1978.
15. Müller-Hill B: *Murderous Science: Elimination by Scientific Selection of Jews, Gypsies, and Others in Germany 1933-1945.* Plainview, N.Y.: Cold Spring Harbor Laboratory Press, 1998.
16. Lagnado LM, Dekel SC: *Children of the Flames: Dr. Josef Mengele and the Untold Story of the Twins of Auschwitz.* New York: William Morrow, 1991.
17. Seidelman WE: "Mengele Medicus: Medicine's Nazi Heritage," The Milbank Quarterly, 66(2), 1988, 221-239.
18. Harris SH: *Factories of Death: Japanese Biological Warfare 1932-45 and the American Cover-up.* London: Routledge, 1994.
19. Getz PC, Pross C: *Cleansing the Fatherland: Nazi Medicine and Racial Hygiene.* Baltimore: Johns Hopkins University Press, 1994.
20. Harris SH: "Japanese Biological Warfare Research on Humans: A Case Study of Microbiolgy and Ethics," in Zelinskas RA ed: The Microbiologist and Biological Defense Research: Ethics, Politics and International Security, New York: New York Academy of Sciences, 1992.
21. Jones JH: *Bad Blood: The Tuskegee Syphilis Experiment,* expanded edition (New York: Free Press, 1993).
22. Faden R; "The Advisory Committee on human radiation experiments: Reflections on a presidential committee." Hastings Center Report 26 (no.5): 5-10, 1996
23. Faden RR, Lederer SE, Moreno JD: "U.S. medical researchers, the Nuremberg Doctors Trial, and the Nuremberg Code: A review of findings of the Advisory Committee on human radiation experiments." JAMA 276:1667-1671, 1996.
24. Gallagher C: *American Ground Zero: The Secret Nuclear War.* The Free Press, New York, 1993.
25. http://www.tuskegee.edu/global/Story.asp?s=1209892 (Accessed 8/05/2004)
26. http://ohsr.od.nih.gov/guidelines/nuremberg.html (Accessed 8/05/2004)
27. http://www.cr.nps.gov/linklaws.htm (Accessed 8/05/2004)
28. http://www.unh.edu/osr/compliance/belmont.html (Accessed 8/05/2004)
29. Freedman B. Equipoise and the ethics of clinical research. New England Journal of Medicine 1987: 317: 141-145.
30. IRB Guidebook: http://www.hhs.gov/ohrp/irb/irb_guidebook.htm c (Accessed 8/05/2004)
31. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm (Accessed 8/05/2004)
32. http://www.medicalcrossfire.com/debate_archive/2003/Oct03/1003_OneOnOne.pdf (Accessed 8/05/2004)
33. Lewis RJ, Berry DA, Cryer H, et al. Monitoring a clinical trial conducted under the Food and Drug Administration regulations allowing a waiver of prospective informed consent: the Diaspirin Cross-Linked Hemoglobin Traumatic Hemorrhagic Shock Efficacy Trial. Ann Emerg Med. 2001;38:397-404.
34. Biros MH, Fish SS, Lewis RJ. Implementing the Food and Drug Administration's final rule for waiver of informed consent in certain emergency research circumstances. Acad Emerg Med. 1999;6:1272-1282
35. Department of Health and Human Services. Protection of human subjects: informed consent and waiver of consent requirements in certain emergency research: final rules, 61 Federal Register. 51500-51533 (1996) (codified at 21 CFR §50.24).

The Research Question

You can tell whether a man is clever by his answers. You can tell whether a man is wise by his questions. – Naguib Mahfouz

INTRODUCTION

The Research Question is the single most important aspect of the research project. Poorly developed questions do not produce "do-able" projects or useful answers. The research question is like the Samsara of Buddhism, the Atman of Hinduism, the Alpha and Omega of Christianity, it is the eternal entity that seeks all knowledge and eventually finds itself. In a universe of quantum uncertainty, the question is the starting point for the investigation, and the endpoint for the answer. Questions represent the method by which the human species have developed our understanding of the universe and how we have repeatedly tested that level of understanding against reason.

Good ideas are easy to find. The hard part is transforming good ideas into good questions that are interesting, important and feasible. In Emergency Medicine we come up with ideas about research all the time. These ideas may come from good health skepticism about a practice principle or dogmatic practice directives. The current literature may present us with alternative ways of approaching a clinical problem that we wish to investigate further. Our patients may challenge us with unusual presentations or questions that bring a new perspective to our practice and our knowledge base. Conversations with colleagues or others bring out varying opinions and ideas about clinical problems. Formal "brainstorming" sessions with other physicians or researchers are frequently fruitful in opening up new ideas that can be translated into research questions. Journal clubs have evolved over the last century from a way to share the expense and the load of compiling information from medical and scientific journals, to a method of learning to read and write scientific literature and generate new ideas about research. At times long held clinical and popular belief may be challenged at the level of "folklore" with new ideas about whether or not any scientific principles underlie these practices. Ideas may be found in the course of teaching sessions or during conferences. In reality research ideas may be found in the course of completely unrelated activities as the "muse" may strike.

If research ideas are so easy to come up with, why is it that there is so little good research done? The major problem lies with the fact that most of these ideas are too nebulous, broad, and grandiose to be of any real use in solving a problem. It is inspiring to contemplate the meaning of the universe as a whole and our relationship as individual humans to it, but it is another thing to actually start to answer some of the real questions associated with that idea. It is also a good idea to wonder "What is the best therapy for a patient with a hospital acquired pneumonia?" But to be of any use, this idea needs to be more specific. About whom is this question directed? There are considerable differences in the pediatric and elderly populations who acquire pneumonia in the hospital. Are you interested in antibiotic therapy, ventilator settings, physio-therapy or other modalities? To be sure, the treatment options and population with hospital acquired pneumonia represents a complex mixture of options and persons. To be of any use to us as researchers, we have to be able to move from the realm of ideas to the realm of questions. The secret of starting a good research protocol, is being able to hone down great ideas, to answerable questions. We have to seek concrete, measurable, palpable, and feasible questions from our ideas. How do we achieve this?

THE FINER HYPOTHESIS CONCEPT

One method that has been used to move from ideas to real questions is called "Developing the *FINER hypothesis*." The neumonic **F.I.N.E.R**. stands for Feasible, Interesting, Novel, Ethical, and Relevant. Creating a research question

that fulfills the requirements for these categories will greatly improve the quality of the research protocol and make the effort worthwhile for the researcher (1).

FEASIBLE

Are there adequate numbers of patients to answer the question? Is there adequate technical experience available? Is there enough time and money? Don't waist your time on questions that you don't have the resources to answer. Many studies miss the boat because of under-powering the results. Therefore, a good sample size estimate is your best friend. Always anticipate losing some patients to follow-up and consider them in your sample size calculation. Make sure at the outset that you will have the appropriate experts, equipment, tools, money etc. If the scope of the question is too broad the question or questions may not be answerable. Don't bite off more than you can chew!

INTERESTING

Is the question interesting to the investigator? What is the motivation on the part of the researcher (Money, Grants, Career progression, Scientific or clinical interest, passing your residency)? What ever the motivation, it needs to be strong. Research can be difficult and (dare I say) boring at times. So it is vital that the researcher be motivated and interested in the project enough to see it through. It is inadvisable to enter into a research protocol without a substantial interest in the topic area and the outcome of the research. To do so will either result in an unfinished project with the attendant waste of time, money, and human resource, or poorly conducted and reported research of which there is a plethora in the literature. Not all research can change the world, but good research can bring a world of change.

NOVEL

Will the answer contribute good clinical information to the body of knowledge? Not every question has to be completely new. In fact a principle of the scientific method is that research must be reproducible to be valid. At times reproducing someone else's protocol and project on your own population is just what the doctor ordered. Particularly if there is a variation in the population tested from the one in the first protocol or if the results of the first protocol seem too good (or bad) to be true. In addition, at times reproducing a study (especially observational non-interventional studies) may be very useful to confirm observations done in one location with another. In addition, reproduction of some studies may be done to learn how to conduct research. If the answer to the research question will make at least some small difference in your clinical practice, it is likely worth doing.

ETHICAL

Your question must be ethical. There is a concept called "fruit of the poison vine" that applies to scientific research. It is the reason that none of the research done in the concentration camps of WWII Germany or Japan is considered valid. Research done outside of the realm of human moral and ethical concepts is not acceptable for use. This is the reason for the existence of the institutional review board (IRB). Some studies cannot be performed in the most proper scientific method due to ethical constraints. There may be ways to address ethical problems such as performing retrospective reviews instead of interventional studies or using animal instead of human models. Ethical matters were covered in chapter 4 of this text.

RELEVANT

Finally, would your results be of value in changing current ideas or validating clinical practice? Although we do not expect most research to change the world, we do expect it to be of some use in the practice of medicine. Research that is done "just to be done" (to complete a research requirement etc,) that will add nothing to the body of human knowledge, should not be done.

BUILDING ON THE RESEARCH QUESTION

The research question drives the entire research protocol from development to publication. Most research questions start out in the range of broad to grandiose, and will need to be narrowed from its original scope to an answerable concrete statement. It should be expressed in simple straightforward terms that an intelligent lay person can understand. It must be as unambiguous and as specific as possible. For example the question; "*Do prophylactic antibiotics reduce the incidence of wound infection?*" is a good first idea but not a good question. It is too broad, poorly defined, and unanswerable. In contrast; "*Does the use of ciprofloxacin reduce the incidence of infection associated with puncture wounds to the foot in healthy non-diabetic patients*?" is more precise and more likely to yield useful information if properly researched.

CAPTURING THE GOOD IDEAS

Once we have created a reasonable research question, how do we go about refining it and developing it into a question worthy of actually creating a research protocol? An initial step is to be certain you actually capture the research question before it gets away. The productive conversations that typically occur at 03:00 hrs in the Emergency Department often produce great ideas and occasionally great research questions. However, by the time you get home, drop off to sleep, and return for the next night shift, they are forgotten. Keep a journal in your jump bag and use it. When one has the occasional discussion session, or the stray brilliant thought, write it down in the journal with the time and date. You can also use the journal to organize and clarify your thoughts on research ideas and questions. Small hard cover journals are available at most book stores and survive the stress of being carried around in the cluttered Emergency Medicine Resident's jump bag better than the soft cover type.

MAKING NOTES ON YOUR RESEARCH QUESTION

Having written down the research idea, devote a paragraph about it to each of the FINER method categories. This will point out where problems may lie in getting from the idea to the question. Once you have a question that seems to fit the FINER model, go to the basic Emergency Medicine Textbooks (Tintinalli, Rosens, etc.) and see what is written and referenced in those sources. Make a short series of notes in the journal regarding what you have found in these texts and start to organize these notes in a logical sequence. It is occasionally helpful to recopy and reorganize things you have initially written as your thoughts progress about the topic area.

THE QUICK LITERATURE SEARCH

The next step is a literature search. Chapter 6 of this text is devoted to how to search the medical literature for your research protocol. However, at this time a short initial search may be done to scratch the surface and obtain a couple of relevant articles (2). Again, this is not the formal literature search you will do before writing the research protocol. Using key words from your topic area perform a basic search with an eye to recent articles, review articles, and any major landmark articles used by most of the other sources you find about your topic area. You should come up with 1 to no more than 3 basic articles in you first search (3). Download the abstracts if available. Some publications actually let you get the full article on line for free. If your are an ACEP or SAEM member you can access the Annals of Emergency Medicine and Journal of the Academic Society of Emergency Medicine on line and down load articles. Once you obtain the articles and abstracts read them, make notes to summarize them, and put them in a logical order in a file or 3-ring notebook.

DEFINING THE STUDY POPULATION – INCLUSION AND EXCLUSION CRITERIA

After reading these articles and abstracts it is time to define your study population. Carefully decide who should be included in the protocol and who should not. You want a study population that will be as representative of the type of cases you will see in your practice but will also be reasonably available for entry into the protocol. An example of the thought process might be as follows.

- You want to study patients presenting with headache.
- After defining your study question and reviewing the literature you decide that patients with a known history of migraine cephalgia are the cases you want to include in the study.
- Next you need to decide specifically how to include these cases. How must the cases have been diagnosed with migraine cephalgia? Do they need to have had a CT of the head to rule out other pathology? Do they need to have been diagnosed by a neurologist? After consideration of all the possible reasons to include a patient in the study, you have arrived at the *Inclusion Criteria* for cases.

After you have defined who can be included in the study, you need to define who cannot be in the study. Federal law limits the inclusion of some "protected" populations such as the mentally handicapped, prisoners, pregnant patients, and some pediatric cases, except under special conditions. These populations are therefore often excluded. You also want to exclude cases that will bring potential and real sources of bias into your study. Does it matter if patients are on preventative medications for migranes? What if they have already taken medications to try to stop the headache before coming to the ED? After considering all these potential problems, the reasons to exclude patients from your study are listed, these are the *Exclusion Criteria*.

THE CASE DEFINITION

The case definition stands as the final description of the nature of the subject who may participate in the study. These individuals are essentially defined by the exclusion and inclusion criteria.

THE OUTCOME MEASURE

Now that the case definition has been clearly delineated, one needs to determine exactly what it is that will be measured (4). What is the endpoint of the study? As example; in the previous discussion of a headache study, what criteria or measurements are to be documented and at what point will the measurement period be over? Perhaps the patient in the headache study will rate their pain on a visual analog scale every 30 minutes till the pain is resolved or for a period of 2 hours. The point is, that you need to carefully define exactly what is going to be measured, how it is going to be measured, and at what point you will stop doing measurements. This then defines the *Outcome Measure* to be used for the study.

OTHER DATA POINTS

In addition to the outcome measures, other information regarding the research subject will probably need to be obtained. These data almost always should include age, sex and ethnicity. These demographics allow for stratification of data in later analysis. Depending on the nature of the study additional data such as social habits, medical conditions, medications and other data points may be obtained as well. It is important not to overwhelm the process by attempting to obtain too much information, yet it is always better to have too much data rather than too little.

FURTHER REFINEMENT

Other measures can be taken to help refine a research protocol. Often it is difficult to find sufficient numbers of patients to enroll into a study. In this setting, several measures may be taken to increase the number allowed into the protocol. These measures include but are not limited to; decreasing the number of variables that are in the process to be measured, narrowing the question further, expanding the inclusion criteria or decreasing the exclusion criteria to get more subjects into the study. You can also increase the duration of the study or the number of institutions involved enrolling more subjects.

CREATING THE HYPOTHESIS

The hypothesis is not the question, it is the anticipated answer (3, 4). In research, there are two types of hypotheses, the *Null Hypothesis* and the *Alternate Hypothesis*. The Null Hypothesis states that in a properly constructed scientific study, there will be NO difference between the control group and the study group when the outocmes are analyzed. The Alternate Hypothesis states that in a properly constructed scientific study, there WILL be a difference between the control group and the study group when the outcomes are analyzed. Why do we bother to state a research hypothesis in the null and alternate manner? In classical hypothesis testing the alternate hypothesis cannot be tested directly primarily because it must have the size of the population defined. We live in a universe of uncertainty where the very fabric of space-time is based on limits of our ability to observe and define. Because of these limits we cannot statistically accept an alternate hypothesis and generalize the results to the entire population without enrolling the entire universe of the population. Obviously we cannot do this in most settings. However, statistical methodology gives us an "out" of this quandary. We cannot fully accept an alternate hypothesis, but we can reject a null hypothesis within a specified degree of statistical probability. If we can reject the null hypothesis at a level of statistical probability (that we define) we can accept the alternate hypothesis at the same level of probability. *The key to a successful research project is that the hypothesis must be testable*. This will be discussed in more detail in later chapters.

SUMMARY

Research ideas must be refined down into answerable research questions. The FINER method is one way of getting from a nebulous research idea to an answerable research question. Keeping a research journal handy is an excellent way to keep ideas from getting away, and to organize thoughts into research protocols. Following a few simple steps in refining a research question take you to the next level and get you set to write a research protocol. Defining the hypothesis sets the research question in motion and places it in a scientific format.

REFERENCES

1. Seehusen D. The FINER Criteria. HTTP://www.fmdrl.org/group/index. (accessed 10/14/2010).
2. Elenbaas JK, Cuddy PG, Elenbaas RM: Evaluating the medical literature part 3: results and discussion. Annals of Emergency Medicine 1983; 12:11: 679-686.
3. Panacek EA, Thompson C: Basics of Research (Part I): Why conduct clinical research and how to get started. Air Med J. 1995;14:33-36.
4. Kwiatkowski T, Silverman R: Research fundamentals: II. Choosing and defining a research question. AEM 1995'5:1114-1117.

Searching the Medical Literature

"It is in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curious of inquiry."

Albert Einstein (1987 – 1955)

INTRODUCTION

Research is never carried out in a vacuum. The investigator must answer the research question in the light of current knowledge base in the scientific literature. In addition, the research question must be viewed in the context of present clinical practice. As such, having a detailed knowledge of pertinent scientific evidence and practice guidelines regarding your area of research is imperative. The Emergency Physician must be able to use information resources available to review and retrieve scientific publications pertinent to the research question of interest. The steps discussed in this section will guide the resident researcher through obtaining sufficient information and theoretical background on which to base a research protocol.

This chapter relies heavily on the National Library of Medicines' (NLM) PubMed link for the research literature search. This is due to the fact that this portal is simple to use, readily available, accurate and detailed, and taxpayer funded. Although this chapter provides the resident researcher with enough information to get started, the NLM provides several excellent on-line tutorials to learn the use of the PubMed portal in detail. The resident is encouraged to review these tutorials as time allows to gain a deeper insight into the use of this powerful resource.

GETTING STARTED

Prior to going to the medical literature in detail, the resident will need to have completed the following steps.

- Identified the area of interest, or topic area in which you will be working.
- Defined at least your initial research question by the FINER method.

At this time, the resident is now ready to approach the medical literature with an open mind, and take a detailed look into what is known about their area of interest. The resident will already have some knowledge about the research topic, but will need to expand their knowledge base and familiarity with the topic before moving on to protocol writing. The level to which you educate yourself in your area of interest will be reflected in the quality of the research you are able to perform. However, it is not necessary to become the World's expert, just be sure you have a good working knowledge of the topics involved in answering your research question. Not much is worse that finding out you omitted learning basic information about your topic area that would have changed your experimental approach or result, after your have conducted your protocol.

TEXTBOOKS AND THEIR SOURCES

One of the first steps is to obtain information on your topic from available textbooks. Recall the old maxim that "*textbooks are 90% opinion with 10% research*." Although the basic principles covered in textbooks are usually quite solid, much of the information presented is more "artful" practice of medicine than hard fact. Discussions of topic areas within textbooks are often colored deeply by the author's opinion. However, recall that usually that author will have a reasonably valid opinion or they would not have been asked to write the book. After reading the sections in the textbooks pertaining to your topic, look at the references used in the textbook to support the ideas discussed. If the cited articles look worthwhile, obtain them and read them as well.

Practice Point

One must devise a method of organizing the information obtained during a literature review. Although there are many ways to do this, the following method has proven effective for the author. Make a hard-copy of the chapters and articles you have decided to use. Place these documents in a 3 ring binder in an organized manner. The individual documents may be tagged with "sticky-tabs" for easy reference. Make notes on the articles (these are your working copies) that outline what is important about the chapter or article. This keeps your references organized and at hand for rapid reference and reading.

The articles you have obtained from the textbook references will also have articles that they reference. Review the citations in these papers and if any seem to be worthwhile, obtain them as well. You should review, make notes, and file them as you did the original articles.

MINING THE REFERENCES

There are two things you want to look for at this point. First, most good research papers will include in their citations a *review article* that covers the topic area in which the research was done. Look for one that is the most up to date (although you will look for this on the online portion of your literature search as well) and obtain it. Second, often in an area of research there is a *"classic"* article that historically initiated research or medical practice in a particular area. This may be an original article that proved a relationship or casual connection. If such an initial classic article exists, you will often see that almost all subsequent research and discussions of related topic areas will cite that article. Obtain such articles and review them.

Practice Point

In some of the original tornado injury research done by the author, there were 6 major research papers that were consistently cited by everyone else who had later published about tornado injuries. Review of these articles demonstrated the "CORE" principles and findings associated with tornado research.

THE FORMAL LITERATURE SEARCH

The next step is to perform the formal literature search for information regarding your research topic. Once entirely in the prevue of the medical informatics specialist, performance of a detailed literature search has become a much less formidable activity. This is primarily due to the development of computerized database systems with human friendly interfaces. These database systems available have replaced most of the older methods of obtaining medical literature. We will discuss these presently.

WHERE TO SEARCH?

Scientific Journals are the primary source of scientific medical information. They are the primary source of scientific communication within the medical community. In addition, scientific journals in general contain the most up to date information regarding a specific medical topic. Scientific Journals provide the process of peer review in which each scientific paper is reviewed in a blinded manner by others who are recognized experts in the field before papers are accepted for publication. This allows a full and detailed evaluation of the content as well as the opportunity for revision and comment by other experts. Such a review by peers allows the best (although not perfect) process we have available to date of ensuring accurate and acceptable scientific content. Many scientific journals have recently initiated a policy of posting scientific articles in their final form, on their web site as soon as it has been accepted for publication (often months before the paper edition is released). As such, no research protocol is complete without a search of the pertinent scientific journals. Currently such early publications will be

entered into the NLM as "epub ahead of print" articles and will therefore be detectged with the use of the PubMed search – discussed shortly.

Internet postings of medical information have exploded along with the other aspects of electronic communications. A number of sources such as The Centers for Disease Control, National Institutes of Medicine, Accredited Medical Organizations or Colleges, etc, may be of high quality and helpful in finding other valid information sources. However the reliability of much of this rapid fire information is inconsistent and usually not peer reviewed. Many individuals today have a tendency to post their PhD or Masters Thesis on-line. There are even postings of local research projects by individuals and organizations that seem authoritative – but may not be. At this point much of the information posted by proprietary or unaccredited organizations on the internet cannot be verified to a satisfactory degree. In addition, sources quoted by such internet sites cannot usually be confirmed. Only reputable internet sites should be used as part of your information search.

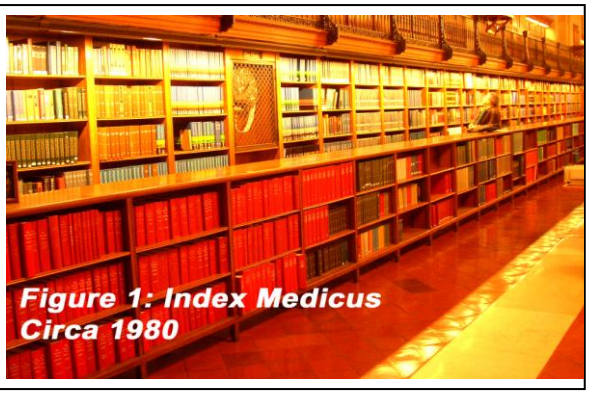


Figure 1: Index Medicus Circa 1980

Performing the Literature Search

Considerable changes have occurred in the mechanism or searching the literature. Prior to the 1980s one was required to physically dig through a series of thick books constituting the Index Medicus (Figure 1) to scan a particular topic area of research across the series of scientific journals (1). This was time consuming and tedious to say the least, requiring a considerable level of training to perform properly. The *Index Medicus*, at the time, consisted of abstracts from all member scientific journals of nursing and allied health care literature, indexed under *Medical Subject Headings* (**MeSH**) representing specific medical topic areas (1, 2). Using the Index Medicus, one would decide on the topic area, check the index for listings in the volumes of scientific journals, obtain the volumes of interest, locate the abstract listings within the volumes, and ask that any articles of interest be retrieved. The process, although laborious sounding (and in reality laborious) could actually be completed with some efficiency once one became familiar with the process. The primary problem with this manual system was that if one did not have the correct set of MeSH headings, one would be unable to locate scientific articles on the topic.

Starting in the late 1980s, with the proliferation of computer technology, substantial changes began to take place in the methods of access to scientific literature. Under the *National Institutes of Health* (**NIH**) congress authorized the *National Library of Medicine* (**NLM**) to create the *National Center for Biotechnology Information* (**NCBI**) in 1988. The NCBI was created with a mandate to develop public databases to cover and link a wide variety of biotechnological information sources. The NCBI was additionally tasked to create the software needed to maintain and use such databases for the public as well as to conduct computational research in the area of database creation and access. The database system created by the NCBI stands as the most complex and successful example of computerized management of scientific information (3, 4).

PUBMED

PubMed is the system available via the NCBI for public access of the NCBI database systems. PubMed is the initial search portal found on the World Wide Web located at http://wwww.ncbi.nlm.nih.gov/pubmed that provides access to full-text articles at journal Web sites and other related Web resources (5). Publishers participating in PubMed electronically submit their citations to NCBI prior to or at the time of publication. If the publisher has a

web site that offers full-text of its journals, PubMed provides links to that site as well as biological resources, consumer health information, research tools, and more.

THE IMPORTANT ELEMENTS OF PUBMED FOR RESIDENT RESEARCH PURPOSES

MEDLINE is the NLM's bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. This database is what PubMed searches for the most part when looking for medical literature. MEDLINE contains bibliographic citations and author abstracts from more than 5,000 biomedical journals published in the United States and 70 other countries. The database contains over 19 million citations dating back to the mid-1960's. Coverage is worldwide, but most records are from English-language sources or have English abstracts.

MECHANICS OF THE SEARCH

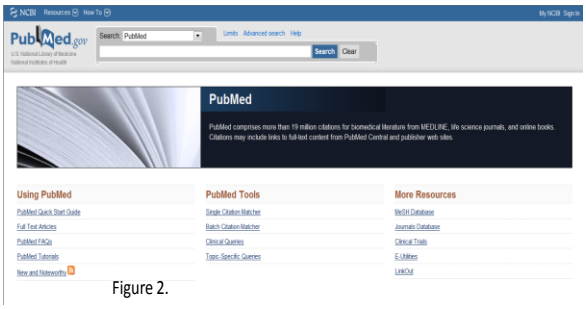


Figure 2.

On arriving at the PubMed site one will be presented with a screen similar to the one demonstrated in Figure 2. The most useful aspect of the PubMed site is that you need only enter Standard English terms into the search box. You no longer need to know the exact MeSH term to be able to find information related to your research topic.

As an example, in Figure 3, the researcher is interested in the need for antibiotic treatment for patients suffering dog bite wounds. Rather than searching these topics under various MeSH headings individually, they simply enter the associated terms in plain English and click the "Search" button to obtain results.

Figure 4 represents the first result page you will get back after having entered your English term and hitting the "search" button. Each numbered line represents a specific citation from the literature. The first line lists the title of the article, followed by the author(s) and finally the journal name with date and journal/page numbers.

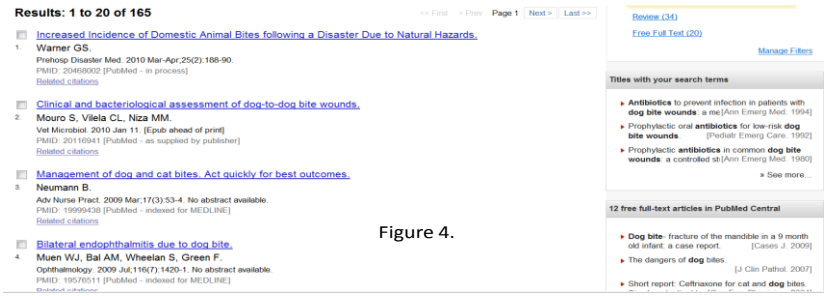


Figure 4.

One can scroll through the citations, usually covering multiple pages. They are usually arranged from the most recent to the more distant. In this case the researcher notes an article from 1994 in figure 5 on the next page, that may not be of much value scientifically, but should at least provide a good laugh. To bring up the abstract of the article the researcher clicks on the highlighted article title as in figure 5.

Clicking on the article title brings up the abstract as demonstrated in figure 6 on the next page. This should provide the researcher with enough information regarding the article to decide if it is worth obtaining the entire text for review. It should be noted that the abstract page also includes side boxes with links to related citations

Figure 5.

that reference the article as well. These are often excellent links to obtain additional articles and information related to ones current literature search.

As a minimum you will be able to obtain an abstract of the citation. That abstract will generally allow you to determine if the article will be of use to you in your research. If the publication is available free online (about 20% of the time it seems to be) you can click the icon, and save a copy to your hard drive or print out a copy for your notebook. Otherwise, you can get the citation information and have the librarian pull it for you.

The abstract page (demonstrated in Figure 6) provides you with additional information regarding the type of article you are looking at, what type of journal it is in, and if

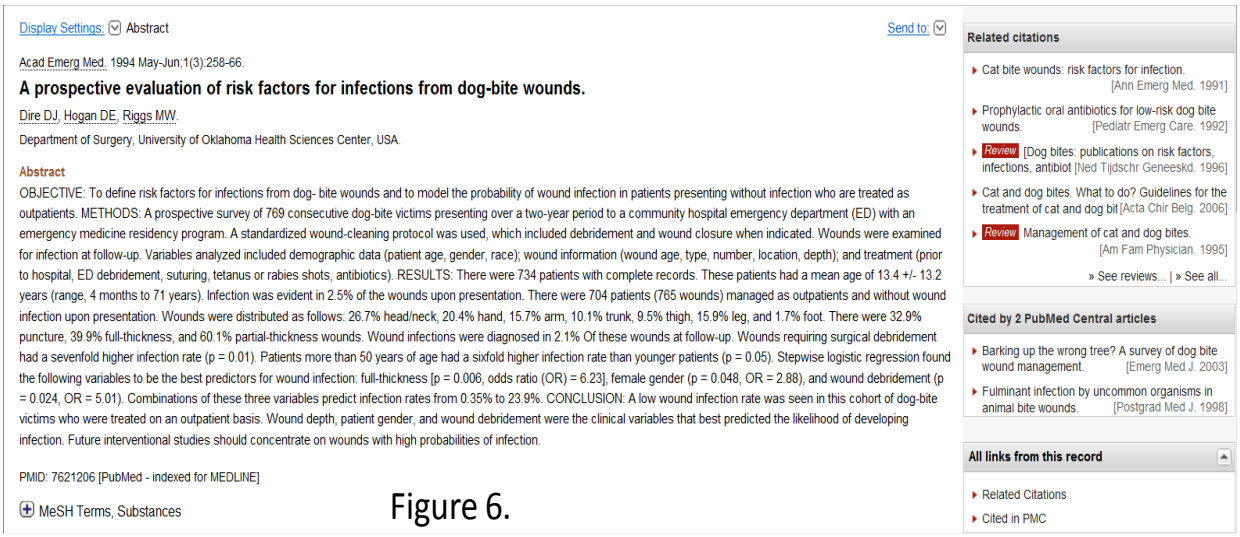


Figure 6.

there are any associated comments, rebuttal, letters to the editor, revisions etc. There are also usually notations as to where the article has been cited in other literature with links to those sources as well.

Another important point is that the information from the abstract may be copied directly from the website page to your word processing program. This can allow you to make an abstract document by placing the abstracts sequentially (or by some other organizational strategy) into a word processing program, printing and placing it in your notebook.

As noted before, there are on-line tutorials available from this same entrance page and the author highly recommends that novice researchers avail themselves of these tutorials. Although you can obtain excellent search results with no more expertise than putting the English terms in as shown above, one can streamline the process and increase accuracy considerably using the methods available in the tutorials.

WHEN DO YOU HAVE ENOUGH ARTICLES?

How do you know when you have gone far enough? When the references and ideas start to converge you have probably obtained the majority of the needed citations. That is, you find that the articles are not citing any new

articles and no new ideas or points of view about the topic are coming to light when you review them. Another indication is when the cited articles are extending so far back into the past as to be useful only as historical footnotes. When you have obtained the original landmark article (or set of articles) that initiated the whole area of research. And finally, you are probably done when you have simply had "enough". There comes a point when you have a good feeling that you have covered the topic area well enough and you seem to be finding nothing new in further research reading. In most resident research protocols this is at the level of 20 to 25 articles. The author recommends in general to residents that you limit your acquisition of articles at this point of the project to about 20 total, not counting a couple of textbook chapters.

THE MEDICAL LIBRARIAN

The description of search processes in this section are not designed in any way to denigrate the role and usefulness of a professional medical librarian. No formal study intended for publication in a peer reviewed medical journal should be submitted without the services of such a professional. However, if the resident will simply follow the suggestions in this section, they will obtain more than sufficient literature to support the creation of a research protocol for the selected research question of interest.

SUMMARY

A simple process may be followed by residents to initiate the creation of a research protocol. Critical to this process is a general review of the pertinent scientific literature available regarding the research topic area and research question. Using the available relational database systems such as PubMed is simple and effective even for the novice researcher. Organization of the information obtained (in 3 ring binders, indexed and annotated) will prove highly beneficial in subsequent steps of protocol creation and paper writing.

REFERENCES

1. Lowe HJ, Barnett GO: Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA 1994;271:1103-1108.
2. Corrall CJ, Wyler PC, Zick LS, et al.: How to find evidence when you need it, Part I: Databases, search programs, and strategies. Ann Emerg Med 2002;39:303-306.
3. Gllagher PE, Allen TY, Wyler PC, et al.: How to find evidence when you need it, Part 2: A clinician's guide to MEDLINE: The basics. Ann Emerg Med 2002;39:436-440.
4. Gllagher PE, Allen TY, Wyler PC, et al.: How to find evidence when you need it, Part 3: A clinician's guide to MEDLINE: Tricks and special skills. Ann Emerg Med 2002;39:547-551.
5. Pubmed HTTP://www.ncbi.nlm.nih.gov/pubmed/ (accessed 10/14/2010)

Study Design

"Basic research is what I am doing when I don't know what I am doing."

Wernher Von Braun (1912 – 1977)

INTRODUCTION

Prior to finalizing your study design, you will need to have completed the following steps in your research process.

- Defined the research question and refined it through the **FINER** method.
- Considered and defined a hypothesis based on the research question.
- Considered the ethical implications of the research.
- Reviewed the medical literature in detail.

After focusing your research question, there is nothing more important in your protocol process than to properly design the study methods such that you can actually answer your research question. Obviously you will have had some general idea as to how you intend to conduct your research protocol during the previous steps in preparing your research. However, you will need to review and discuss your methods ideas with a mentor who has some experience in conducting research. This is also the time to involve a biostatistician in your research protocol process. The objectives of this chapter are to define the most common research designs that are applicable to most resident research and discuss the advantages and disadvantages of each. There are 3 broad categories of research study design; *The Survey, Historical Approaches, and Experimental Approaches (1)*.

THE SURVEY

The *Survey Approach* is an effort to determine an answer to a question with the assumption that the answer lies in current practice or opinion. Variables cannot be controlled in survey methodology. This approach, therefore, is non-experimental. Surveys provide researchers with information on current opinions, practice methods, etc.. They may also provide the prevalence of such ideas and methods within the population of those surveyed. Surveys have a low level of scientific validity but may still provide highly useful information if designed, conducted, and interpreted properly.

HISTORICAL APPROACHES

The *Historical Approach* is also non-experimental. Relevant data to the research question (which have already been generated – usually in medical records or population databases) are collected and reviewed. This design attempts to determine the influence that those data have on current practice. Researchers have only to make some sense out of information to which they have access in relation to their research question. Historical studies suffer from numerous sources of bias. The primary problem is that data have already been collected and recorded. This means that the data may not have been collected in an unbiased manner with attempts to control sources of bias as it would have been if the data collection had been performed prospectively under a strict research protocol. In addition, often specific data points are not available to the researchers due to the incomplete nature of most medical records.

Historical studies may be improved in quality greatly by the creation of a specific research protocol for data collection prior to recording data from historical sources. This means that before the researcher goes to the data sources (aside from gathering an idea as to what data points may be available in those sources) one creates a protocol clearly defining what data are to be collected, along with a careful definition of what each data point is. This helps prevent the practice of *data dredging* (a process of going through a data source and creating a "research

paper" out of whatever one can find). An example of an historical study would be a review of medical records to answer a specific research question.

THE FLOW OF INFORMATION PROBLEM IN HISTORICAL STUDIES

The *Flow of Information Problem* merely states that the investigator, by the nature of the study design, is removed from the actual data collection. Each step between the investigator and the original data increases the potential for error. For example, an LPN measures a blood pressure of 164/ 95. When asked for the vital signs by the RN, the LPN states 160/90. When asked for the respiratory rate the LPN estimates a rate of 18. The intern writing in the progress note later misreads the numbers and transcribes a BP of 150/90 and a respiratory rate of 18. The study assistant who is abstracting the data cannot read the handwriting of the intern (especially since the note was written 3 years ago and is obscured by a coffee stain). She misreads the numbers as 150/70 and 16. The blood pressure which is used for data analysis is obviously much different than that recorded by the LPN and the respiratory rate is complete fiction. While this may be an extreme example, this type of error could clearly alter the results of a study.

EXPERIMENTAL APPROACHES

Only the *Experimental Approach* allows for control of variables. In fact, the entire point of experimental approaches are to isolate and control potential sources of bias and to focus on the points of change or intervention of interest (*study variables*) so as to see the impact of study variables on the measured outcome. Randomized, blinded studies seek to control interventions, time, environment, and other confounding variables in a rigorous manner. *Randomization* takes on major significance in the setting of experimental studies and will be discusses later in other sections.

THE SURVEY STUDY

If the answer to the research question lies in current practice and/or opinion, the survey method may be an effective research study design. Surveys are one of the simplest methods available to the researcher as well as one of the least temporally intensive. It can be of moderate expense as it involves contacting the survey population directly in person, on the phone, or via the postal services. Surveys are also the least scientific of all research methods and are highly limited in the ability to obtain objective information and produce information that can easily be generalized to larger populations.

While surveys may seem to require less work than other types of studies, this is not always the case. Much work must go into the survey design and validation of the survey study tool. If the survey is thoughtful and well researched, surveys can provide vital data that would otherwise be unobtainable. Surveys, however, have a bad reputation in the scientific community. This is partially due to bias in the minds of most researchers toward more *objective* data, but it is also due to the fact that there are so many bad survey papers published and referred to every year. If the survey tool is thrown together quickly, without proper review and pre-tests, there is little chance of producing accurate data. Survey research requires a copious amount of preparation. Well done survey studies can add substantially to the body of medical knowledge. Poorly done survey studies are not only useless, but can be harmful by diverting both public and scientific attention and resource away from real issues.

TYPES OF SURVEY STUDIES

If general surveys can be performed by two methods, Questionnaires and Interviews.

QUESTIONNAIRES

Questionnaires are usually self-administered forms which require the subjects to enter data themselves, either at home or at the study site. This allows greater subject privacy and better standardization of test questions.

Questionnaires are less time consuming on the part of the researcher as they put the bulk of the work on the respondent. However the response rate can be very low.

INTERVIEWS

Interviews are usually conducted one-on-one with a researcher. They usually require a standardized set of questions and provide better control over the data collection and a more flexible approach to qualitative information. They are more time consuming and resource intensive than questionnaires but the response rate of data collection is quite high.

WHAT SURVEY STUDIES CAN ACTUALLY DO

Surveys have proved to be a fairly good tool for modeling peoples' opinions and behaviors. Cultural anthropology, sociology and marketing research are all carried out primarily by survey research and statisticians have developed a number of techniques to analyze this type of data. Topics such as patient satisfaction, preferences, behaviors, attitudes, can only be approached through a survey design. This approach can be particularly useful when studying pain control, customer service quality, disaster preparedness, practice preferences, health care systems and other topics that require a flexible approach to the outcome data. Small but well prepared exit polls have predicted the results of presidential elections even in very close races - some of these exit polls have even been correct.

THE RESEARCH QUESTION IN SURVEYS

Like any study it is important to have a clear goal when attempting to develop a survey study. You must have a well-reasoned definitive research question. Without a clear question, you will never be able to adequately interpret your data. This is the problem with many surveys. A question such as "How do parents feel about pediatric emergency care?" may seem interesting but it must be narrowed down to be approached. A more appropriate question would be "How do parents feel about pediatric emergency care in an urban trauma center as measured by patient waiting time, pain control, impression of the facility, impression of nursing care, impression of physician care and treatment expectation as compared to parental pre-visit expectation of emergency care?" While this question may seem overly cumbersome, it clearly delineates the approach to the general question. In short it is more answerable.

Such a question also provides guidance as to how the survey instrument needs to be developed and formatted. In the above example about pediatric emergency care, the real question is "did the experience in the pediatric ED meet the expectations of the parents?" In general it is better to stick to one topic (in this case the comparison of the pediatric ED experience and the parent's expectations) but it is acceptable to approach the question via a number of measurable criteria. In the above example the research question actually lists a number of elements of the ED encounter that are to be assessed in the survey. It is less acceptable to piggy-back other studies onto the primary question, such as attempting to determine parental attitudes on anti-smoking education in the middle of the survey on the ED encounter. It is tempting to try to answer a number of questions with one study (and therefore get a number of papers out of the project) but the addition of more questions will only decrease the response rate and the quality of the data. Stay focused on only the elements of experience and opinion that will help answer the research question.

SURVEY SAMPLING METHODS

Sampling methods are just as important in surveys as they are for any other research method. To continue the example of the pediatric ED survey, one will get very skewed data if they only sample parents in the department between 7 and 9 in the morning when there are fewer patients and generally quicker visits. One would probably get the opposite result if the population is sampled only between 7pm and 9pm, despite the fact that the researchers are sampling the same general population. A truly random sample from many different parts of the

day will provide less biased results which will be more likely to reflect the true population statistics. Obtaining large samples through continuous data collection or other methods such as mass mailings will always provide more accurate data.

SURVEY LENGTH

If you have ever personally been solicited to perform a survey, you are well aware that the longer the survey, the less likely that you will complete it. The urge to include more questions and criteria must be tempered by an understanding of human nature. People do not like to be asked the same questions over and over again. The longer they have to concentrate on the survey, the more likely they will give the quick answer instead of the thoughtful answer. Obviously this will erode the validity of the study. Remember that the respondent to the survey is rarely as interested in the topic as the researcher. To get at the data quickly and effectively, one needs a well thought out question with clear goals. This data should be approached with the minimum number of questions necessary.

REPEATING QUESTIONS

Repeating questions – usually by asking the same question in a different way – is commonly used as a method of assessing internal consistency in a survey study. One would hope the respondent would answer the questions in a similar way as an indication of their true opinion on an issue. However one should avoid unnecessary repetition of questions. Do not ask every question twice. Select only a few important ones and scatter them throughout the questionnaire randomly.

QUESTIONNAIRE DESIGN (THE SURVEY INSTRUMENT)

The actual questionnaire or interview script should be as simple as possible. Begin with a description of the study and very clear and explicit instructions. Consider providing 1 or 2 examples of questions with responses to demonstrate how the form should be marked. Begin with basic questions, such as name or sex. Save sensitive questions for the end when respondents have already put some effort into the form and are close to finishing. Each question should contain only one idea. Attempting to combine ideas into a single question will detract from the validity of your results. Keep the form spacious and align answers in a vertical fashion. If you are planning on quantifying data, consider pre-coding the data collection forms with the corresponding numbers to be entered into the database.

QUESTION STRUCTURE (OPEN OR CLOSED)

An open question is one which invites the respondent to answer in whatever manner they deem appropriate, e.g.

What do you think contributes to the development of heart disease? _____

This allows the respondent to fill in the blank with their response. Open questions are more suitable if one desires actual quotes, if looking for typical answers to develop a more specific questionnaire or when investigating topics which are not easily broken down into discrete answers. This is useful in that there is a great degree of freedom in how the respondent can answer the question. It is bad in that the respondent can actually answer in ways that are of no use in the data collection process. In this setting – depending on how the question is asked – the researcher may end up putting responses into "categories" rather than using the actual data provided. This can introduce more bias into the process by adding the interpretation of the researcher into the mix. As such, when using open ended questions, one must be very specific in the wording.

A closed question is one which forces the respondent to answer in a specific way, e.g.

Which one of the following do you believe contributes most to heart disease?

(check one box)

[] 1. Smoking

[] 2. Stress

[] 3. High Blood Pressure

[] 4. Diabetes

Notice that this example is pre-coded for easier data entry. That is, the numbers 1, 2, 3, 4, are easy to enter and analyze in a database and correspond to the specific answers of smoking, Stress, High blood pressure, and Diabetes. In general closed questions are easier to quantify and analyze. Closed questions do however limit the responses of the respondent and "force" an answer that may not be what the respondent actually wants to answer. This obviously introduces bias into the process. This bias can be limited to some degree by allowing an "Other" response with or without allowing the respondent to write in an open ended answer. Most medical survey research will be done with closed questioning.

SURVEY SCALES

Quantification of survey data requires some form of measurement scale. A number of these have been proposed. The scales on the following list are used commonly and are generally acceptable. An example of each is provided in the following pages.

- Likert (summative)
- Forced Likert
- Semantic Differential
- Guttman (cumulative)
- Visual Analog (VAS)
- Numerical Descriptor (NDS)

If you use a new or novel scale, you will be required to validate that scale by pre-testing it and measuring statistical significance and internal validity. The resident should consult a statistician for more information on this process.

LIKERT (SUMMATIVE)

The Likert Scale is somewhat quantified, provides easy data entry characteristics into a database – which then allows for at least some descriptive and comparative data analysis. An example of a question using a Likert Scale is;

How comfortable is this bed?

1	2	3	4	5
uncomfortable		neutral		Comfortable

FORCED LIKERT

The Forced Likert scale makes the respondent choose, there is no neutral choice. As example;

Is this bed comfortable?
1 2
uncomfortable comfortable

This obviously severely limits the responses possible and forces the data into the form the researcher wants. Such forcing may be allowable in some settings (as above when really all the researcher wants to know is did the respondent consider the bed comfortable or not and not "how comfortable" the bed was). Although adding some bias, it also limits some bias in that the decision process for the respondent is simplified in terms of what the researcher actually wants to know.

SEMANTIC DIFFERENTIAL (FORCED OR NOT)

This scale is very similar to the Likert Scale with the exception that the respondent is placing a response in boxes or some represented scale between two extremes. It is useful in that one may evaluate several specific issues about a specific element with the same question. In the example below;

This bed is….(place X in box)
comfortable |_|_|_|_| uncomfortable
too long [] [] [] [] [] too short

one is able to examine a number of specific issues about a single element of importance to the researcher (ie the characteristics of the bed in the ED). It is an efficient way to get the respondent focused on a single topic or element and then obtain a rapid-fire assessment of their opinion about the topic. Obviously the Semantic Differential is subject to the same sort of advantages and disadvantages as the Likert Scale.

GUTTMAN (CUMULATIVE)

This form of survey scale requires the subject to demonstrate a logical process in answering the question. Such questioning can be used to help determine the validity of a study. This is due to the fact that the process of answering actually requires that the respondent to consider the issues behind a particular response and arrive at a logical conclusion. In this setting, the response is actually the question itself and one is asked to either agree with the response or not. Although the Guttman Scale can sample reasonably complex issues and the respondents understanding of those issues, it is also more difficult for the respondent to complete. If the survey is long, the nonresponse rate will be high. An example of Guttman Scale questions is as follows;

Circle the letter for every statement for which you agree
a] Drinking can cause injury
b] Drinking is an important cause of injury
c] Drinking is an important cause of injury and death
d] Drinking is the most important cause of injury and death in the United States

VISUAL ANALOG (VAS)

Visual Analog Scales are used in surveys in an effort to make the responses more analytical and thus more amenable to statistical analysis. This is performed by asking a question and having the respondent mark their answer on a line (usually 10 cm long) between two extremes of response. The question is scored by actually measuring the distance from the negative response on the left in centimeters and millimeters. The measurement is equal to the score. Obviously the higher the score, the more the respondent agrees with the positive response. An example might be;

Mark on the line where your pain falls.

_____**X**_____

None Severe

In this setting one must have good photocopies or the original data sheet in order to do accurate measurements. The VAS is commonly used in the collection of highly subjective data such as in pain studies where the information provided by the respondent will vary greatly from participant to participant. One may have the patient evaluate their pain before and then after some specific intervention in order to compare the differences between the score on the VAS. This makes it easy to create means and medians for various groups and subgroups of individuals and to quantitize highly subjective information. The tendency however is to over analyze VAS scores by "pretending" that they are actually objective data. The measurements are nothing more than subjective responses forced into a number that can be treated as objective data points. There is no good way to judge the internal validity of the respondents when they go from one response to the next. There is a large level of bias in the data, and it is impossible to determine which direction the bias is pushing the results. Despite this, the VAS can provide very useful information in the analysis of difficult research areas (such as pain research) as long as these limitations are kept in mind.

NUMERICAL DESCRIPTOR SCALE (NDS)

The Numerical Descriptor Scale (NDS) is essentially the same as VAS but quantified rather than physically measured on a line. This still allows quantification of the response but forces a simplification of the actual data into a number from 1 to 10. When the amount of bias in these sort of scales are considered, there is probably little difference between the NDS and the VAS statistically. The NDS is subject to all the limitations of the VAS, but is faster to enter into a database as it requires no direct measurement with a ruler. An example of an NDS is noted below.

Circle the number where your pain falls.

1 2 3 4 5 6 7 8 9 10

no pain moderate pain worst pain of life

SOURCES OF BIAS

There are many causes of inaccurate or incomplete data. Some causes such as a lack of knowledge or inability to recall the information cannot be helped. The types of bias and their causes in research have been discussed in chapter 3 of this text. Some causes, such as vague and poorly worded questions, are completely under the researcher's control and can be minimized with simple attention to detail on the part of the researchers. There are some procedures that can help to reduce the bias in surveys.

CLARITY OF THE SURVEY INSTRUMENT

As noted before, the questionnaire should be clear, concise and easy to read. A clean and easy to follow process will encourage and properly engage the respondent into the process. This also reduces confusion when the process of answering the question is easy and clear (some readers may recall the Florida Presidential Election of 2000 where unclear ballots raised a substantial amount of controversy).

SURVEY LANGUAGE

The language of the questions should be unbiased. Accusatory questions are not likely to be answered (eg. When did you stop beating your wife?). In addition, the language must be understandable to the potential respondents. Most physicians use language skills well beyond the education level of the patients they treat. Even if the physician has learned to effectively communicate verbally, they tend to write at a much higher level than the general public. As such, it is easy for a written survey instrument to be beyond the usual reading comprehension of the average patient. There are "readability" tests that may be used. An example is the Flesch-Kincaide Readability Test that is available as part of many word processing programs. Using a mathematical protocol the test analyzes written documents and determines the likely age or grade level at which the written material will be understandable. Simple versions of this test are available on the internet for free. The researcher should use one of these readability tests to assist targeting the survey to the test population. In general, a 6th grade reading level is considered appropriate by most survey takers.

ENGAGING THE RESPONDENT

The more the survey process engages the respondent, the more likely it is that the subject will complete the survey and provide accurate data. There are a number of ways to increasing the level of engagement – most of which rely on human nature and behavior.

Shared Interests

If the survey is collecting data on some issue that is of high interest to the respondent, the response rate will be higher. This is particularly true if the respondent belives that the response they give will be noted by the researcher and particularly if it is possible that their response may have some positive impact (from their point of view) on the issue of interest. One way to increase response is to offer to send the participant a copy of the scientific report created (or refer them to a web-site where the analysis will be posted).

Strike While the Iron is Hot

Some issues are of temporary interest to respondents. As example, if one is seeking information on activities associated with a disaster (such as an earthquake or flood) the longer one waits before getting a survey into the hands of the subjects, the less likely they will be to participate. In general, most acute issues are only of significant interest to people for 30 to 60 days following an event.

Provide Inducement

If people are rewarded for participation, the participation level is increased. The sort of reward offered for participation in a survey will vary with the type of survey and the budget of the researcher. At times, simply offering some form of "credit" for the information is sufficient. At other times providing gift certificates or even money will improve participation (although the author admits that even when a fresh $1.00 bill is provided in the mail for participation in an unsolicited survey, he still forgets to participate most of the time). The reward must be commensurate with the amount of work required but must not be so large as to induce bias into the results (highly paid respondents will likely want to "please" the researchers in their responses).

FACE TO FACE DATA COLLECTION

When the opportunity presents itself data may be collected by a face to face meeting with the respondent. This may be done by interviewing subjects while they are in the ED or clinic, by setting up interviews with subjects from a list, or any number of other arrangements – including telephone interviews. Such active data collections are more expensive and resource intensive, but have a much higher response rate. They also allow the researcher to explain the question if the respondent has questions, and to clarify any answers given (although this can also introduce bias). It is however not usually possible to obtain as many responses in total due to the resource intensity with face to face data collection as with postal based processes.

ON LINE SURVEYS

Creating on-line survey processes has been simplified reciently with not only the proliferation of simple web-site construction and software, but the development of specific web applications such a "Survey Monkey." Survey Monkey (available at http://www.surveymonkey,com) is a site that allows one to place simple surveys (developed by the methods previously discussed in the chapter) on line for free. Respondents can go to the website and participate based on an email with a link or being given the web-site address. The results of the survey are only collected by the on-line process and are then available to the researcher for more detailed analysis. Additionally researchers may choose to use their own web-site and create a data collection process based on other available programs.

PILOT TESTING

Pilot testing is the best way to improve and sharpen the survey tool. Poor wording, annoying fonts, confusing formatting and many other problems can easily be identified by third party reviewers who are given the survey in a pilot test format. Just about anyone can provide input into the survey: research directors, staff, residents, medical students, technicians, nurses, patients, friends, family etc. The more people who review the form, the more likely you are to discover problems, and consequently the better the data you will get. The pilot study should include a number of individuals who are representative of the population the researcher intends to study.

PLANNING FOR NON-RESPONDERS

Obviously you will not get a response from every subject you solicit. It is not uncommon for surveys to have a 50% non-responder rate or higher. Your goal should be 80% or greater response. One should plan on doing multiple mailings or telephone follow-ups to increase the number of responses. Your estimated sample size should be based on your total responders, see chapter 10 on sample size calculations. Consider extending your sampling if you continue to have inadequate numbers.

RESPONSE RATES AND RESPONDERS

Do not be discouraged by the initial response rate, 50% or less is not uncommon. Increasing the response rate is desirable but it is not the only method to deal with this problem. The other approach involves testing whether or not the non-responders would have substantially changed your data. People who respond promptly to a survey are usually different than those who respond much later in the process. It can be useful to characterize these different groups to discern if non-responders are substantially different. Early responders usually have strong opinions on the questions being asked and therefore a motivation to reply quickly. Later responders usually have less strong opinions and may not have taken as much time with the questions as you would like. In general, non-responders are usually less interested in the subject. They may have less resources and education than the responder group. However, they may also be well educated, but just busier than the responders.

In general, non-responders tend to be similar to late responders. Analysis of these groups may provide insight into whether the non-responders would have changed the results of the study. The first and last 25% of responders

should be compared to discover differences between the groups. Subsequently, the researcher might personally call a random sample of the non-responders. If that survey demonstrates that the non-responders are similar to the late responders, then categorize them that way and publish the results. If, however, that sample is substantially different than the late responders, then the researcher must contact and survey enough of the non-responders to reach the target rate (sample size) of the study.

THE CHART REVIEW

Chart review studies are a form of archival (secondary) data research and are one of the most commonly used *Historical Approaches* to research. This is a popular research method for resident and pre-hospital studies. The popularity of this form of research is due primarily to the perceived ease with which it is performed. There is no informed consent required and the *Institutional Review Board* (IRB) process is easier than that required of prospective studies. While aspects of chart review research are certainly easier, there are a number of drawbacks. Frequently many hundreds of pages of charts must be examined. The data required may not be available in each record. And, these studies may not receive the respect afforded to prospective trials. This final criticism is often due to the fact that chart reviews frequently lack clear published methods or may be the result of data dredging - which is defined as determining the research question after examining the set of data. While data dredging can produce important research questions, it is imperative that the study to investigate any such questions be performed on a different data set than the one which generated the question.

ADVANTAGES OF CHART REVIEW STUDIES

Obviously there are many advantages to doing chart reviews. As stated earlier, the IRB process is usually less cumbersome and there are usually no informed consent issues. In addition, these studies are much less expensive and can usually be performed on little or no budget. They are relatively quick to perform and can be done at a convenient time. These studies also lend themselves to the examination of multiple questions and the generation of further questions (which should of course be examined on a different data set). In short chart reviews may provide very useful information in a short period of time at little expense – as long as certain rules of the process are carefully adhered to.

DISADVANTAGES OF CHART REVIEW STUDIES

It is impossible to control sources of bias in the data collection as the data are already collected and documented in the chart. It is also difficult to control bias on the part of the individuals reading the chart and recording the data. The data is always incomplete. In addition, the chart review, as an example of an historical approach to research, is subject to the *Flow of Information Problem* as discussed previously.

WHAT A CHART REVIEW CAN TELL THE RESEARCHER

Chart reviews are commonly used in general epidemiologic studies for all the advantages noted above. However the limitations of this mode of research must be clearly understood. At times, researchers may have a tendency to extrapolate and generalize more from a chart review than is actually possible. In analysis, a chart review can determine the prevalence of some data element in the population (of patients who have charts) studied. These data elements may be stratified by age, sex, or other components of the population that were collected in the database. As such, rates, proportions, and ratios can be expressed.

In addition, if there are two groups or more, means can be compared. An example might be if data on death or survival of casualties from car crashes were to be obtained from a review of charts from a trauma center. Suppose such data were collected on whether or not the casualties were wearing seat belts at the time of the crash. The researcher could then compare the mean fatality rates between those wearing seatbelts and those not to see if there is a significant difference. While a chart study can suggest an association between some data element found

in the chart with some other data element (as example a higher proportion of fatalities in car crashes where the occupants were not wearing seat belts when compared to occupants wearing seat belts) it cannot link CAUSE between these data elements. In other words, one cannot make the statement that not wearing a seat belt during a car crash CAUSES a higher fatality rate than wearing a seat belt. It can however suggest a cause that can be evaluated with experimental forms of research studies.

ADDITIONAL REASONS TO AVOID DOING CHART REVIEWS

Many additional arguments can be made against doing chart reviews. Medical records are a secondary source of data, which are often in error. Observations in the charts can vary widely based on the skill of the clinician making the chart. The data is frequently incomplete or inconsistent. The process suffers from the *Flow of Information Problem*. The abstraction of the data is often be biased on the part of the investigator. This does not mean that chart reviews are never appropriate. There are certain questions that can only be performed by chart reviews. In addition, small chart reviews have contributed substantially to the advancement of medicine. For example the association between aspirin use and Reye's syndrome in influenza A cases was initially suspected from a small chart review. Obviously that chart review had a substantial impact on clinical practice. Just be certain that the limitations and capabilities of the chart review process are clearly understood, and appropriate for the research question being asked.

CREATING A CHART REVIEW

Chart reviews should be created and managed much like any other sort of study. That is, they should be based on a refined research question and research protocol. Doing so increases the quality of the data obtained, reduces bias, and strengthens the overall study results.

THE RESEARCH QUESTION

Using the FINER method discussed in chapter 5, one should come up with a research question that can be answered by a chart review. As example if one wants to know, "does exposure to methyl chloride cause liver cancer?" a chart review would not be appropriate. However if one asks "what is the initial clinical presentation of cases found to have spinal epidural abscesses seen in the Emergency Department?" – a chart review would potentially work.

PERFORM A LITERATURE SEARCH AND WRITE A RESEARCH PROTOCOL

The resident, now armed with a proper research question, performs a literature search and writes a research protocol as described in chapter 8.

WRITE AN ABSTRACT MANUAL AND CREATE A DATABASE

This step forces one to select the individual data points that will be looked for during the abstraction of the medical records. It consists of a paragraph or two that describe the intent of the chart review study (which can be cut and pasted from the research protocol) followed by an individual listing of the data points to be collected along with the definition of those data points. Because one is listing the specific data points to be collected, a database can easily be constructed at the same time using Microsoft Excel or any other user friendly relational database program.

OBTAIN INSTITUTIONAL REVIEW BOARD APPROVAL OR EXEMPTION

There are few journals today that will accept and few institutions that will allow research that has not been reviewed by an IRB. It has become the standard of research nationally. As such, even a chart review must be submitted and reviewed through the local IRB process before it can be conducted. Most chart reviews will meet criteria for exemption from consent as there will be no actual contact with the patient and there will be no identification of individual patients in any of the research documents or database files. This IRB review however is

still required in order to access the patients *Private Health Information* (PHI) which are in the custody of the medical institution. In addition, an IRB review can help ensure that the study is being done on a sound scientific and medical-legal format.

TRAIN ABSTRACTORS

Often the researcher will need more than just themselves to review the charts and transfer the information therein onto the data forms (called abstraction of the charts). As such, the researcher may need to employ the assistance of others, who will need to be trained in the protocol and abstraction methods before being put to the task. Doing this greatly increases the quality of the study and reduces bias due to the abstraction process. One uses a short presentation giving an overview of the study, and then provides a copy and performs a review of the abstraction manual to the abstractors. This usually takes about an hour for most studies. If funds are available to reimburse the abstractors for their time it is beneficial. However, resident studies of this sort are usually done without funding so other means of reward may be offered. This can range from a heart-felt "Thank You" to food and or gift certificates funded out of the researcher's pocket. Enlisting fellow residents to the task may be reciprocated at a later date as they work on their own projects.

ABSTRACTING THE DATA AND ENTERING DATA TO THE DATABASE

Medical informatics must be notified ahead of any chart abstraction process and provided with a protocol and a letter of approval from the local IRB. If physical charts are needed, this will require that the medical informatics personnel actually pull the paper charts needed and make them available, a procedure that may take a period of time – and represents additional work for the informatics staff. Being very nice to the informatics staff (and perhaps providing coffee, snacks, cookies, pizza etc.) can go a long way towards getting the records one needs to complete a project.

In most cases, the charts will be sequestered in medical records in a specific location. At selected times the abstractors may present themselves and identify themselves as abstractors working on the research protocol during working hours of medical informatics. An alternate is when the record may be accessed electronically from a location in the facility, which can provide a wider range of times available for the abstraction process. It is important in this setting that the protocol be followed carefully in this case in that no hard-copies of the record with patient identifiers should be printed out for convenience unless this is part of the official protocol approved by the IRB.

Following the collection of data onto the abstract form, the data must be entered into the database. If one has designed the abstract form and the database simultaneously, it is usually an easy task to transfer information from the paper to the database – line for line. An alternate to using the data sheet created in the abstraction manual process is to enter data directly into the computer database using a laptop computer or hand held device. Some researchers still like the paper abstract form as it provides a backup should the electronic format become damaged.

CLEAN UP THE DATABASE AND PERFORM THE ANALYSIS

One now reviews the database to make sure all the data points available are entered in the proper cells. Errors in entry must be found and corrected. Following this the data are ready for analysis. All studies need a general descriptive analysis of the study population entered into the database. This includes reporting the mean and median ages, the gender distribution, and ethnicity (if collected). Such information allows one to determine how much like other populations the study population is likely to be. If the these demographic elements are similar to other populations to which the authors wish to generalize the results, it strengthens the evidence.

Next one looks at the specific data elements collected in relation to the research question. As example, a study done by one resident looked at the clinical presentation of patients diagnosed with spinal epidural abscesses and found the following;

- 90% abused intravenous drugs
- 100% were smokers
- 90% abused alcohol
- 90% were febrile with a mean temperature of 38.7°C
- 80% had an elevated erythrocyte sedimentation rate.

Multiple other findings were noted but the above gives the reader an idea of the sort of information that can be obtained. With enough data, a profile of the clinical presentation of cases with spinal epidural abscesses may be developed – in this case to indicate to the clinician what presentations may be "high risk" for the infection if properly generalized.

Further analysis may be carried out depending on how the study was constructed. There may be various groups of patients that can be compared to look for differences etc. These procedures are covered in detail in the chapters on hypothesis testing and statistics.

COHORT STUDIES

Cohort studies are observational studies that can either generate or test a hypothesis depending how they are conducted. Case-Control Studies (CCS) are a related method that will be discussed in the next section of this chapter. These studies by nature do not involve an intervention or randomization process. However, it is important to understand that when properly designed and conducted, cohort studies can have almost the same power as randomized controlled trials (RCT) to identify certain issues in nature. They are somewhat easier to conduct than RCTs. Cohort studies are also termed "natural" studies as they generally let nature take its course without intervention.

A cohort is defined as an identified group of the population to be included in the study. The word comes from the Latin word denoting a particular group of soldiers in the Roman Army. There are two general types of cohort studies 1) Prospective cohort studies, and 2) Retrospective cohort studies. The nature of the cohort study and its possible impact on clinical medicine may be demonstrated in the following example.

HISTORY OF COHORT STUDIES

During the 1920s physicians in Great Britain and elsewhere in the world began to suspect an association between cigarette smoking and lung cancer. This suspicion was based on the observed fact that most of the patients being seen with lung cancer also had a strong history of cigarette smoking. From the early 1930s through the 1960s a number of epidemiologic studies were conducted to attempt to determine if there was an association between cigarettes and lung cancer. The two most important were performed by Doll in 1947 and Hill in 1951.

Sir Richard Doll used a case-control method where he identified patients in the hospital with lung cancer and compared them to a similar group of hospital patients who did not have lung cancer. Then both groups were assessed for a number of risk factors including cigarette smoking. The results were striking. Almost all the lung cancer patients had a history of heavy cigarette use, while hospital patients of the same sex and age without lung cancer did not have history of smoking for the most part. Because there were such a large number of patients included in the study, the association between cigarettes and lung cancer was quite clear.

Based on this, A. B. Hill used a cohort method to further evaluate the apparent association. In this setting Hill created a study group of young physicians (the cohort) some who smoked and some who did not smoke. They were followed forward in time and monitored till death. The causes of death were then evaluated for death due to lung cancer, or other causes. This was then matched with the history of cigarette smoking. What he determined was that there was a very strong association with death from lung cancer and a history of cigarette smoking.

WHAT A COHORT STUDY CAN TELL THE RESEARCHER

Based on these two studies (and a myriad of additional similar studies that confirm these results) we generally accept cigarettes as a "cause" of lung cancer. The reader should note at this point that neither the cohort or case-control study method can firmly establish anything as a "cause" (or assign causality). This typically requires a stronger experimental process such as a randomized controlled trial (RCT). However, we accept cigarettes as a cause of lung cancer based on two issues. First, because we already know that cigarettes are harmful based on the cohort and case-control studies, it would be unethical to set up an experiment where individuals were assigned to a group where something known to be harmful is being given to them. Second, the association detected in the cohort and case-control studies is so strong as to leave almost no doubt that cigarettes are the cause of the lung cancer.

Despite this example, most cohort studies do not provide such a strong level of association as the Hill study. They do however allow one to detect potential impacts of various factors on outcomes. As example, a recent study at the author's facility looked at cutaneous abscesses evaluated in the ED. The cohort consisted of all patients presenting with abscesses that were treated. All physicians were allowed to treat each patient as they saw fit, without interference. Along with a number of other data points collected were the time it took the wound to heal and whether or not the patient was placed on antibiotics. There were some patients in the group who were given antibiotics and some who were not. This allowed the researchers to divide the overall cohort into two sub-cohorts based on use of antibiotics and to compare the time to heal for both groups. There was no difference in the time to heal for abscesses between those patients given antibiotics and those not. If the finding is strong enough and there are enough patients in the study (all elements of the statistical analysis) then it supports the idea that antibiotics are not generally needed for cutaneous abscesses.

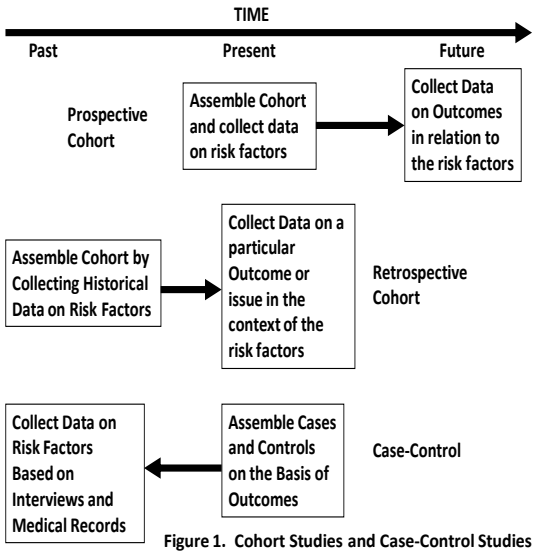


Figure 1. Cohort Studies and Case-Control Studies

CREATING A COHORT STUDY

In creating a cohort study, the resident should consider the fact that there is a limited time in which to complete the study during residency. As such, one will not be likely to be able to follow patients over a long period of time. Figure 1 demonstrates the general differences between prospective and retrospective cohort studies as well as Case-Control studies (described in the next segment of this chapter). Using the processes covered in this book to this point (developing the research question, performing a literature search, etc.) one settles on the area of interest and develops the specific research question(s). One then decides on a prospective or retrospective cohort approach.

A PROSPECTIVE COHORT STUDY

As the example noted before, suppose we wanted to know if irrigating a simple cutaneous abscess verses not irrigating it has any impact on the time to heal for that lesion. First one assembles the cohort – all patients presenting to the ED for treatment of cutaneous abscesses. This means one must carefully identify and define the specific criteria used to determine if the patient has an abscess so that the "cases" being selected are rigidly defined. This is done by a careful set of inclusion and exclusion criteria.

Once the cohort has been identified, one must decide what data points must be collected in order to answer the research question and minimize confounding and bias. In general this includes demographics such as age, sex and ethnicity. Then such data as vital signs, whether or not the abscess was irrigated, the size of the abscess and other information is added in order to answer the research question. Each of these data points must be carefully defined so that individuals collecting and recording them for the study are consistent.

After the initial data is collected from the first visit, each case must be followed up through whatever method has been decided upon. This may be done through a repeat clinic visit or by letter or telephone. This represents the outcome data and may be such information as how long it took the lesion to completely heal or if there were complications. The time frame of follow up here obviously needs to be reasonably small for a resident project, on the order of weeks to a month. These data are then all placed into a relational database file for organization and analysis.

Although specific statistical processes are intentionally left out of this text – for the most part – examples will be used to support the illustration of data analysis in both cohort and case-control studies. All the data collected in the study are put into a relational database. In this case, one of the questions asked is whether irrigation or the abscess cavity had any effect on the time it took the abscess to heal. Each database is different, but one would arrange the data with the information about the time it took the lesions to heal into groups of those irrigated and those not irrigated. Then using the specific directions for each database, one performs a Student t-test on the data. This provides the researcher with a comparison of the mean time to heal between each group in such a way as to determine if any difference is real, or simply due to random chance. If a difference is found (suppose those abscesses irrigated took less time to heal) and the student t-test indicated that the result was "significant" – that result must now be interpreted in the context of the clinical problem being investigated.

From this, one can see how the cohort study design is able to evaluate multiple findings, risk factors or exposures. As example, suppose we also recorded the presence or absence of surrounding erythema, necrosis, lymphangitis or lymphadenopathy. We could then do the same student t-test by dividing these groups up based on having or not having the finding above and again see if there is a difference in the time to heal between groups. The number of data point observations like this that can be included in a study are essentially unlimited – but should be moderated by the physical ability to collect the data, and it's probable relation to the outcome (time to heal in this case) being studied.

A RETROSPECTIVE COHORT STUDY

Because retrospective cohort studies start with the problem or diagnosis, and then look back into the past to try to determine if there are any risk factors or exposures that may be associated – this study method is more commonly used in epidemiologic and outbreak evaluations. However, it can be used in topic areas directly related to acute care processes as demonstrated in the following example.

As discussed previously, suppose we want to evaluate the effect of wearing a seatbelt on the degree of injury found in *Motor Vehicle Crash* (**MVC**) trauma casualties. A retrospective cohort case method can serve our purposes here. After developing the research question we could obtain information on sequential trauma

casualties arriving at a trauma center. In most trauma centers, a myriad of data points are collected regarding the trauma event and recorded in the trauma registry. Looking at the data, we find the information about whether or not the trauma casualty was known to be wearing a seat belt or not at the time of the crash. In addition, we match up the level of injury (as defined by some trauma scale we decide to use) into none, minor, major or death. We then set up a contingency table with 2 X K columns (with K being the number of "disease states" or categories desired). This is demonstrated as follows;

Seat Belt	Injury None	Injury Minor	Injury Major	Injury Death	Totals
Yes +	75	160	100	15	350
No -	65	175	135	25	400
Totals	140	335	235	40	750

Table 1. A 2 X 4 Contingency Table of Seatbelt use and Level of Injury

Although there is a fair amount of information we can obtain from this table, let us restrict our thoughts at this time to what we can learn about the data above with simple proportions and ratios. As example, calculate the proportion of seatbelt users in each level of injury.

Level of Injury	Calculation	Proportion Wearing Seatbelt
None	75/(75+65)	58%
Minor	160/(160+175)	48%
Major	100/(100+135)	43%
Death	15/(15+25)	38%

Table 2. Calculation of the Proportion of Trauma Casualties Wearing Seatbelts and Level of Injury

Even with this simple use of proportions from the contingency table we can see a trend in the proportion of seatbelt use dropping as the severity of injury increases from no injury to death. Therefore the more severe the injury of the trauma casualty, the more likely it is that the casualty was not wearing a seatbelt. This method certainly cannot show a causal relationship between seatbelt use and injury, but it rather suggests that there is one. We could further investigate this apparent association with odds ratios and other statistics, but the point has been made about how a retrospective cohort study can provide valuable information.

CASE-CONTROL STUDY

In the standard *randomized clinical trial* (**RCT**), an outcome is determined after exposing or not exposing randomly assigned patients to an intervention. Logistical, ethical, and financial considerations may make this difficult. The retrospective *case-control study* (**CCS**) has been used as an effective alternative, especially in the study of disease etiology. The term "*case-control study*" was coined by Sartwell to offset the implication that the retrospective nature of this design is its essential feature (2). In studies of disease etiology, for example, researchers begin by selecting a group of patients (*the cases*) with a particular disease. They then determine the rate of previous exposure to the presumed etiologic agent. The exposure rate also is determined in a group of patients who do not have the disease in question (*the controls*).

Case-Control Studies are conducted to evaluate patients who are kown to have a particular disease or cutcome and compare them to patients who do not have that particular disease or outcome. As such, it can evaluate for exposures and risk factors that may have been involved in the development of a particular disease or outcome. Case-Control Studies are observational in nature as no intervention is carried out on the part of the researchers, and retrospective in that one starts with the known disease or outcome and looks back into the past to determine if there have been specific exposures or risk factors. The CCS differs from the Cohort Study in that one also collects data on a group of people (the controls) as similar to the patients with the known disease (the cases) as possible, but who have not developed the disease or outcome of interest. In this way one can calculate the odds ratio of developing the condition of interest or not developing it based on any number of exposures or risk factors.

ADVANTAGES OF CASE-CONTROL STUDIES

The major advantages of the CCS design include savings in time and actual cost. It can be used to study reasonably rare diseases or conditions in clinical practice. The CCS can also evaluate multiple risk factors or exposures at the same time and can establish associations between these risk factors or exposures to particular outcomes. The CCS can also be deployed rapidly to solve real-time problems such as disease outbreak issues. In settings in which diseases take years to develop or occur infrequently, the cohort or RCT becomes expensive and logistically difficult. In the CCS, the disease has already occurred in the case population so one only needs to evaluate the historic exposures (see figure 1). The CCS format is also useful in the study of exposures (such as alcohol in pregnancy) that cannot be randomized for ethical or logistical reasons.

DISADVANTAGES OF CASE-CONTROL STUDIES

Case-Control studies are retrospective by nature and are therefore subject to all the confounders and bias inherent to retrospective studies. These problems may be decreased by the creation of a good study protocol and adherence to that protocol during the performance of the study. In addition, CCS researchers face the same methodological problems as their counterparts who plan prospective studies. These problems include case and control group selection, definition and detection of the disease under study, definition and determination of the exposure, baseline susceptibility of the 2 groups to the disease, and statistical considerations. Since the purpose of most CCS is to determine the association of specific risk factors or exposures to a disease, selection bias becomes a potential problem in both case and control selection. Arbitrarily including or excluding certain groups seriously skews (and may invalidate) the results. For example, excluding patients who have a particularly high or low rate of exposure falsely raises or lowers the rate of exposure among those who remain as cases or controls.

CONFOUNDING IN CASE CONTROL STUDIES

A particular problem (although certainly not unique) to CCS is the problem of confounding. This is when there seems to be an association between some risk factor or exposure and a particular disease or outcome, when in fact, the exposure or risk factor and the outcome or disease are not associated, but both are associated with a third (usually unknown) confounding variable.

BIAS IN CASE CONTROL STUDIES

Multiple forms of bias can easily enter in to the case-control design. For convenience, samples are frequently drawn from hospitalized patients; however, these patients are among those most likely to have been exposed and become diseased when compared to non-hospitalized individuals in the general population. Referred to as *Berkson's bias*, this is a very difficult impediment to overcome. Another source of selection bias involves studying a group that is not representative of all patients with a particular disease, as would occur if some patients suffer transient illness or experience early death. Regardless of what method is used to select cases and controls, it should be established before the data are obtained and analyzed. Manipulation of study groups afterwards to achieve desired results is not scientifically permissible.

DESIGNING A CASE-CONTROL STUDY

The disease, condition or outcome to be studied must be defined in clear, specific, and unambiguous terms by using operational criteria. Cases consist of individuals who meet the diagnostic criteria for the medical problem of interest. Controls are selected from age and sex matched individuals that are as similar in life style as possible to the cases. Diagnostic tests and procedures must be performed and interpreted equally in both groups. In so doing, the researcher is assured that members of the control group do not have the disease.

Exposure must also be defined in a precise, unambiguous fashion. Other investigators should be able to use the same definition in an attempt to reproduce the findings. If possible the person collecting the data should not be

aware of the hypothesis being tested and should not know whether the subject is a case or control (termed blinding). Issues of timing also can be important in the determination of exposure. The members of the case and control groups must have had an equal likelihood of developing the disease under study in their baseline pre-exposure state. This is the methodological issue of *baseline susceptibility*. Variables that are linked to exposure and predisposition to disease are referred to as *confounding variables*. The impact of those factors must be assessed and reflected in the controls. This goal can usually be accomplished by matching members of the treatment and control groups on the pertinent variables or, at least, by showing comparability among groups with respect to those variables. Another method for minimizing the impact of confounding variables is to employ multivariate methods in data analysis.

CASE-CONTROL STUDY EXAMPLE

An example of a CCS can be noted in the following. A group of body-builders began presenting to a local ED with severe muscle pain and swelling. Biopsy of the cases muscles determined an unusual eosinophillic infiltrate was present. A CCS was assembled consisting of the known cases of "Eosinophillic myalgia syndrome" (EMS) and controls consisting of fellow body-builders from the same gyms who were not afflicted. Data were collected on risk factors and exposures in which the data as shown in figure 2 were assembled. It was noted that although both cases and controls often consumed the amino acid L-tryptophan as a supplement, 45 of 46 cases obtained their L-tryptophan from the Shows-Denka Company. When analyzed for an odds ratio, it can be seen that body builders who consumed L-tryptophan from the Shows-Denka company had a 57.5% chance of developing EMS. It was later determined that this supplement was contaminated with a toxin that caused the disorder.

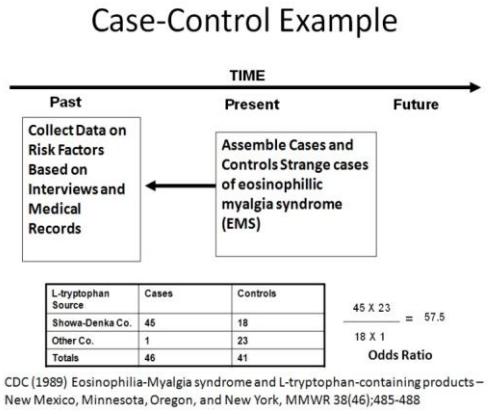


Figure 2: Case Control Study Example

META-ANALYSIS

The application of statistical procedures to examine tests of a common hypothesis from more than one previously published study is known as a Meta-Analysis. It is in reality an "analysis of analysis." A meta-analysis attempts to perform statistical analysis using published data from a collection of studies of the same methods in the body of available literature (3). The meta-analysis provides a statistical process for testing a conceptual hypothesis using these previously published papers. The origins of the meta-analysis can actually be traced back to the statistician Ronald Aylmer Fisher – who is generally considered to be the creator of modern statistical science (4). As such the meta-analysis is not a new invention, although it has only come into common use in the last decade or so.

WHEN TO USE A META-ANALYSIS

One problem commonly seen in clinical research is that individual studies published often have numbers too low to provide a high level of confidence in the results. Although such studies may meet the statistical requirements associated with study number calculations, they are actually too low to provide adequate "resolution" (the ability to clearly see the difference between groups beyond a statistical difference). This is particularly true in the setting where the disease or issue being studied is relatively rare. One then ends up with a series of published papers about a particular topic with low numbers at the edge of the statistical level of resolution. The meta-analysis seeks to increase the resolution regarding the issue (based on a new conceptual hypothesis) by combining the numbers of cases from the multiple studies. The meta-analysis can also help clarify an issue just by analysis of the methods used in each study. In this setting the meta-analysis can look at the quality of the studies done and stratify the results based on high quality, or low quality studies.

EXAMPLE OF A META-ANALYSIS

One meta-analysis was performed to investigate confusing results in the literature regarding the effectiveness of the Bacillis Calmette-Guerin (BCG) vaccine used in many locations of the world in preventing new cases of tuberculosis. The researchers located 8 controlled trials in the literature evaluating this question. Three studies demonstrated benefit from the vaccine and 5 studies demonstrated no benefit or actual harm from the BCG vaccine. On the face of it, it appears that the "majority" of the studies demonstrate that the BCG vaccine is not beneficial. However, when the selection criteria that were created for determining the quality of the study to be included in the meta-analysis is applied the researchers found that the three studies with the best quality all found a benefit from the BCG while the 5 studies that demonstrated no benefit all had poor methods and were unacceptable for inclusion into the meta-analysis due to method problems and wide confidence intervals. This alone suggests that one should put more stock in the studies supporting the use of BCG. When the numbers from the three good studies are combined and statistics applied, the researchers found a significant association with the use of BCG and a decrease of new TB cases, with a reasonably narrow confidence interval.

CREATING A META-ANALYSIS

It is obvious from the above example that the methods used in the meta-analysis must be quite strict and followed carefully to decrease the amount of bias allowed to enter the process. According to Thacker, the steps in creating a meta-analysis are as follows (5):

- Define the problem to be evaluated and the criteria for which studies can be included
- Locate research studies in the literature
- Classify and code study characteristics of each study
- Place the data from the studies on a common scale for analysis
- Analyze, interpret, and report the results

ADVANTAGES AND DISADVANTAGES OF META-ANALYSIS

There are a number of benefits to meta-analysis. It forces systematic thought about methods, outcomes, categorizations, populations, and interventions. It allows one to increase the resolution of the data sets by combining them and thereby increasing the available numbers of cases. The combination of data from numerous studies increases the ability to generalize from results and potentially increases statistical power. It also should decrease investigator bias if the bias of the various studies pull the data in different directions.

The primary disadvantage is that one does not actually perform the data collection themselves and are therefore dependent on the original researchers for the quality of the data. In addition, because of the variation of research methods, it can be difficult to find papers of sufficient quality and study design to be included into the meta-analysis. There are even those who espouse that meta-analysis is not actually "research" and that the results are not valid. However, it is clear that if the study is well designed, rigorous and the protocol strictly followed, that valuable information may be gained by this process.

RANDOMIZED CLINICAL TRIALS (RCT)

A *Randomized Clinical Trial* (**RCT**) has become the accepted standard for evaluating therapeutic efficacy. Randomized trials are done in a prospective fashion to allow as much control of variables as possible. The outstanding feature of such trials is the use of randomization to help prevent bias in the assignment of subjects to specific intervention groups. If randomization is carried out properly, differences in outcome that are observed among treatment groups will tend to result from treatment effects and not from inherent differences among the groups (6, 7).

WHAT A RANDOMIZED CONTROLLED TRIAL CAN TELL YOU

Randomized Controlled Trials (RCTs) are a true experimental method. This means that when designed and conducted properly a RCT can not only demonstrate an association between some experimental variable and an outcome but can also demonstrate causality between some variable and an outcome or disease (8).

The RCT allows the researcher to isolate and control variables and confounders to a greater degree than the other research methods discussed previously. As such, the RCT method is best used to study single agents such as pharmaceuticals, or single exposures in an effort to link their clinical effects and outcomes. In addition, the RCT is able to isolate the effects of such agents well enough in most settings that is can often resolve even small differences between groups (on the level of 20% or less). This makes the RCT valuable when the actual change in a study group expected from some intervention might be small – but the potential clinical significance is high. An example might be if a particular anti-tumor pharmaceutical was able to reduce the tumor burden in treated patients by only 20% - but that 20% reduction in tumor might translate into several extra years of quality life survival.

ADVANTAGES OF THE RANDOMIZED CONTROLLED TRIAL

As noted, the RCT is the primary experimental method used when strong association or actual causality between some agent and an outcome is sought. In addition, it is most commonly used to compare one treatment to another (usually some "gold standard") to determine if the new agent is equally effective or superior to the gold standard. RCTs may also be used to determine effectiveness of a particular agent in comparison with a placebo or "sham" therapy to determine if an agent has clinical efficacy beyond the placebo effect or random chance. RCTs are also commonly "blinded" meaning that the subject, researcher or both do not know what intervention or treatment is being given. Blinding substantially reduces potential treatment bias.

DISADVANTAGES OF THE RANDOMIZED CONTROLLED TRIAL

Few residents will be able to conduct their own RCT for a resident project during their tenure at a residency program. This is due to the complexity, cost and length required for most studies. RCTs are difficult or unfeasible to conduct in many settings. Many clinicians and patients are reluctant to accept randomization, especially if one of the proposed interventions is particularly desirable or undesirable. Many potential research questions are simply not ethical to be resolved by an RCT – particularly if this involves the withholding of helpful interventions or the administration of harmful interventions. Furthermore, huge numbers of patients are required if the disease prevalence or clinical impact of an intervention is low. This may necessitate the involvement of multiple institutions with all the attendant increases in complexity and cost. All RCTs require the most stringent of informed consent processes, thorough monitoring by the IRB and strict adherence to Federal regulations. Substantial financial and criminal penalties may be levied for violation of such statutes. At the least researchers found to have violated the monitoring requirements for consent, ethics or confidentiality may be banned from further research in the United States. Finally RCTs are usually of such complexity that they require at least a minimal research staff to carry out in addition to financial backing in the form of research grants or loans.

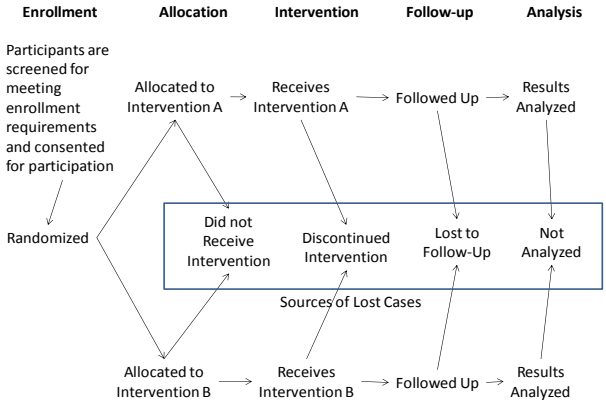


Figure 3: Randomized Controlled Trial

METHODS OF RANDOMIZATION

Randomization is at the heart of the RCT. Two critical assumptions required in an RCT in formulating the probability distribution are that; *(a) the sample observations are independent* and *(b) the samples are identically distributed*. Therefore, patient eligibility (inclusion) criteria must be well defined and strictly followed (9). It follows that the clinician investigator and the patient must agree to accept any random treatment assignments before the patient is formally entered into the study. If that does not happen, selection bias is introduced. Figure 3 depicts the general structure of an RCT through the Enrollment, Allocation, Intervention, Follow-up, and Analysis phases.

Enrollment is performed by carefully designating what subjects are eligible for entry into the study. This is again done by careful inclusion and exclusion criteria depending on what issue is being studied and what the research question is. Individuals meeting the inclusion criteria may then be invited to participate in the study and undergo proper informed consent.

Randomization is next performed to determine what intervention is to be allocated to what individual. There are a number of classic ways that randomization is accomplished. The simplest (and crudest) is the equivalent of a flip of the coin. A stack of papers, each labeled with a 0 (for control) or 1 (for case), is shuffled, a paper is drawn for each subject, and the subject is assigned accordingly. This method has been criticized (and largely abandoned) for lack of blinding and the possibility of inadequate shuffling. In many clinical trials, a randomization list is constructed from random number tables or from random numbers generated by a computer. In 1955, the Rand Corporation generated one million random digits by designing an electronic roulette wheel specifically for that purpose. Realistically, computer generated lists probably will be used.

The interventions are then carried out with at least one group getting one intervention and the other receiving another (or a placebo). Additional arms may be created with more than one intervention or drug being tested with or without a placebo arm of the study.

In the follow-up phase, those subjects completing the intervention are followed-up carefully and the results of the particular intervention are documented.

Finally the results of the interventions are statistically analyzed and comparisons are made with conclusions being drawn.

The RCT has the advantage of being highly flexible with the ability to add multiple arms and subgroups for analysis when the study is designed. This can however lead to highly complex and cumbersome processes than are costly and take considerable time to complete.

RANDMOIZED CONTROLLED TRIALS IN THE CONTEXT OF CLINICAL TRIALS

Rather than providing a specific example of an RCT, this method of study may be better understood in the context of Clinical Trials (10). In the United States and much of the western world, the study of new agents has been partitioned into a set of clinical trial categories. These allow for both safety and efficacy data to be obtained on the agent in a sequential process. The point is to maximize the protection of human test subjects while still allowing medical science to progress (11).

PHASE 0

This is a reasonably new class of trial also known as a human mico-dosing trial. The phase 0 trial had allowed a substantial reduction in the use of animal trials which have historically been the initial step in the study of agents prior to entering phase I trials. This is not only beneficial to the animals, but avoids the often complex and

conflicting results obtained in animal trials due to the differing physiologies between humans and the test animals. This has become increasingly more important as more pharmaceutical agents are developed based on human genomic data that are highly specific for human physiology. In phase 0 trials a new agent is given to human test subjects (usually fewer than 100 subjects who are paid for the experience) in doses much lower than those expected to be required for treatment of a disease. Data on the pharmacokinetics and pharmacodynamics of the agent are obtained. This process does not provide safety or efficacy information, but does produce basic information on which the researchers may determine the likely dose and monitoring methods needed to further study the agent.

PHASE I

This had traditionally been the first stage of testing done on human subjects (and still is if animal trials are the first step rather than a Phase 0 trial). Usually done in an inpatient setting, a set of subjects (usually around 100) are given what are thought to be therapeutic doses of the agent to establish tolerability, pharmaokinetyics and pharmacodynamics. Healthy subjects are usually used but on occasion patients with a particular disease (such as cancer) who lack other options may be used. Subjects are paid for their participation. Multiple types of Phase I trials exist but will not be covered here for brevity sake.

PHASE II

When the initial safety of the agent has been determined in Phase I trials, Phase II trials are done to determine dose and efficacy. Often phase II trials are divided into Phase IIA (designed to determine the dose) and Phase IIB (designed to determine the efficacy). Some phase II trials are of a cohort design as well as RCT design. Phase II trials use several hundred participants.

PHASE III

These are large multicenter studies of several thousands of subjects in a RCT design. The study agent is usually compared with an established agent (termed the “gold standard”). These are highly complex and expensive processes. On occasion, other uses for a particular agent will be suggested during this phase of testing and researchers may wish to expand the uses of an agent under the FDA regulations (termed label expansion). Additional studies conducted along these lines are termed Phase IIIB trials. In general Phase III trials confirm the efficacy and safety of an agent on a large segment of the population. The results of these multicenter trials are complied and analyzed to become the information found in the package insert as approved by the FDA.

PHASE IV

A less known trial is the Phase IV trial or “Post Marketing Surveillance Trial” conducted after the release of a particular agent to the open market. These trials are conducted to detect any unanticipated effects such as drug interactions or toxicities that were not detected in the Phase III trials. This involves monitoring adverse drug reports by clinicians as well as active surveillance of particular sectors of the health care market for evidence of problems. Detection of problems may result in the withdrawal of an agent from the market, or modification of the package insert with warnings or changes in indications.

CROSSOVER STUDY

Although frequently used in the past, *Crossover Studies* have fallen out of favor in the past few years. In a multi-treatment crossover study, each patient receives 2 or more treatments in sequence, and the outcomes in the same patient are contrasted (11). For example, in a 2-treatment protocol, a patient's response to treatment A is compared with the same patient's response to treatment B at a later time. To a certain extent, this design allows individual patient characteristics that influence response to be disregarded. In a variant of this design, matched pairs of organs (eg, limbs, teeth) are measured at the same time instead of measuring the same patient at different

times. The term "crossover" refers to the fact that at some predetermined point in time, the experimental population and the control population are switched without the knowledge of the subjects or the persons administering the treatment.

The decision to use a crossover design usually is based on the potential saving in sample size. However, recent opinion among research professionals has increasingly discouraged the use of crossover studies due to a number of associated problems with the design that are likely to invalidate the results of such studies. Although crossover studies have provided for a number of good studies in the past, they are rarely used in current research methodology, particularly in Emergency Medicine.

SUMMARY

Researchers must decide which study design they will use. Choice of design is based on a number of factors, including time, finances, type of data required, and available resources. The study design must be consistent with answering the research question. Proper study design insures the best opportunity for the conduct of a valid research protocol. This chapter has covered only a few of the standard research design methods and their variations. As always consultation with a research consultant on study design is advised before proceeding with a major project.

REFERENCES

1. McCabe JB: Research: general principles and emergency medicine application. In: Van de Leuv JH, ed. Management of Emergency Services. Rockville, Md: Aspen Publishers; 1987:355-361.
2. Hayden GF, Kramer MS, Horwitz RI: The case-control study. A practical review for the clinician. JAMA 1982 Jan 15; 247(3): 326-31.
3. Chalmers TC, Levin H, Sacks HS, et al: Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. Stat Med 1987 Apr-May; 6(3): 315-28..
4. Glass GV: Primary, secondary and meta-analysis research. Educ Researcher 1976; 5: 3-8.
5. Thacker SB: Meta-analysis. A quantitative approach to research integration. JAMA 1988 Mar 18; 259(11): 1685-9.
6. Guyatt G, Sackett D, Taylor DW, et al: Determining optimal therapy--randomized trials in individual patients. N Engl J Med 1986 Apr 3; 314(14): 889-92.
7. Rimm AA, Bortin M: Clinical trials as a religion. Biomedicine 1978 Sep; 28 Spec No: 60-3.
8. Ellenberg SS: Studies to compare treatment regimens: the randomized clinical trial and alternative strategies. JAMA 1981 Nov 27; 246(21): 2481-2.
9. Moeschberger ML, Williams JA, Brown CG: Methods of randomization in controlled clinical trials. Am J Emerg Med 1985 Sep; 3(5): 467-73.
10. Zelen M: The randomization and stratification of patients to clinical trials. J Chronic Dis 1974 Sep; 27(7-8): 365-75. Connelly LM. Retrospective chart reviews. MedSurg Nurs 2008;17(5):322-323.
11. Aschengrau A, Seage GR III. Essentials of Epidemiology in Public Health (2nd Edition) Jones & Bartlett Sudbury, MA 2005.
12. Louis TA, Lavori PW, Bailar JC 3d, Polansky M: Crossover and self-controlled designs in clinical research. N Engl J Med 1984 Jan 5; 310(1): 24-31.

Design Type	Definition	Use	Limitations
Experimental	An experiment designed to investigate the role of some agent in the causation, prevention, or treatment of a specific disease.	High fidelity determination of causality and associations under conditions maximizing control of confounders and variables. Best in the study of single agents or exposures. Useful when the difference between groups is expected to be < 20%.	Expensive, time, and resource consuming, potential serious ethical issues. The strictly controlled nature may make it difficult to generalize results.
Observational	A general class of design that observes individuals in a population for specific outcomes or exposures without performing interventions.	Able to study a wider range of agents and exposures than experimental. Naturalistic in format and therefore usually more generalizable.	Difficult to limit the impact of confounders and variables.
Cohort	A study that observes cohorts (groups of people with a similar characteristic) over some temporal span monitoring for an outcome or exposure.	Useful in following differing cohorts (usually exposed and unexposed) forward (prospective), backward (retrospective) or both (ambidirectional) through time. Used to compare the occurrence of symptoms, disease, and death between the two cohorts.	Sensitive to the selection of a proper comparison group and to follow-up. Complete sources of information may be difficult to find. May be time and resource consuming.
Case-Control	A study using two groups that are selected (cases with a disease) and (controls without a disease) who have their exposure histories obtained and compared.	Used to evaluate relationships between exposures and disease.	Usually considered less rigorous than Cohort studies. Exposure information may be difficult to find.
Cross-Sectional	A study that obtains information from a select population at a specific period in time with regard to disease prevalence, conditions, and exposures.	Usually carried out for public health planning and disease etiologic research. Provides a "snapshot" of disease prevalence and conditions.	Sensitive to temporal and conditional forms of bias.
Ecological	These studies evaluate rates and relationships of disease on a population rather than an individual level in which the exposure is a property of the population.	Mainly used to evaluate population level disease rates under large scale levels of environmental exposure over time and geographic location.	The lack of individual information results in sensitivity to the "ecological fallacy."

Table 3: General Study Types with Definitions, Uses, and Limitations.

Writing the Research Protocol

Writing is not necessarily something to be ashamed of, but do it in private and wash your hands afterwards.

Robert Heinlein (1907 - 1988)

INTRODUCTION

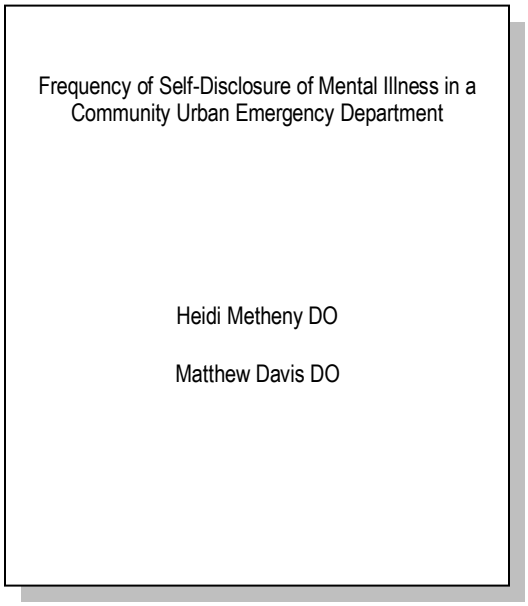
Writing a research protocol is necessary for the successful completion of any research project. This section will provide you with the needed instructions to start and complete a basic research protocol for use in the research program at almost any residency. This is not the "only" way to outline and write up a research protocol, but it is a simple method that covers all the basic requirements. After following these steps, it should be a simple matter to complete an IRB application. In addition, when the protocol is completed it is usually a matter of cutting and pasting sections of the protocol into the draft of the paper you are preparing for publication.

STEP ONE – GET INFORMATION IN ORDER

Get your information in order. Before you can write your research protocol you will need to have completed the following steps that have been covered in earlier parts of this manual.

- You will have already decided on your research topic, and refined your research question by the FINER method.
- You have considered various research methods to answer your question and have, in general, decided what method to use.
- You have conducted at least a preliminary literature review and have considered your reasons for conducting your research and your research question in light of the existing published material on the topic.

Using this information you are now ready to start getting the protocol down on paper were it will do some good.



STEP TWO – CREATE A TITLE PAGE

Create a title page. This page should include the title of the protocol (don't worry about it too much as the title will change before any publications come of the project anyway). The title should be as specific as possible as to what the project is all about (this will sometimes make the title lengthy). The title page should also contain the names of all persons involved in the project with the *primary investigator* (**PI**) listed first. The PI is the person who is ultimately responsible for the conduct of the project. Finally there should be contact information listed at the bottom of the page for the PI. This page is not numbered and should have no running header. An example of a title page is demonstrated in figure 1. It is amazing how important such a seemingly silly thing as this page is. The title serves to help the investigators focus on the task at hand, as well as provide needed contact information. In addition, there is something about seeing one's name on a page of paper that can stimulate further literary activity.

STEP THREE – THE REASON FOR THE STUDY

The Reason for the study: The next page is page "one" of the protocol. It should contain three sections; an *Introduction, Research Question, and Study Purpose.* This page will define your reasons for doing the study.

INTRODUCTION

As the reader, the *Introduction* should make one interested in your topic and provide the reason you think it is important to learn something about this area. You should mention any "classic" or "landmark" papers (if any) that opened up this area of study or practice that you obtained from your literature search. Briefly cover basic background information to set the reader up for your research question and study purpose to follow. You think this is important stuff or you would not be conducting a study, make the reader as excited and interested in it as you are.

RESEARCH QUESTION

Next, create a paragraph on the *Research Question*. Under this you should phrase the full research question as you have developed it. Remember this will need to have been refined by the FINER method before being placed here.

Introduction

Patients with mental illness represent an often misunderstood and overlooked segment of the patient population in the Emergency Department (ED). This is especially true in the harried environment in which most emergency physicians practice. The problem oriented patient approach to patient care often used in the ED may leave little room for exploring pertinent historical elements of the medical history that the patient does not readily divulge. Prior psychiatric diagnoses fall into this category. Such psychiatric problems are often noted only as an afterthought once the ED physician has initiated diagnostics and/or therapy for a non-psychiatric complaint. At times the presence of a preexisting psychiatric disorder or even the use of psychoactive pharmaceuticals remains undisclosed by the patient. Existing psychiatric disorders and the spectrum of action of psychoactive pharmaceuticals can have a substantial impact on the application of ED therapy. Once this information is elucidated, the clinician must consider unique medication interactions, comorbidities, coping skills, and other special needs in treating these patients.

In order to appreciate how mental illness impacts the decision-making process in an emergency department, a conscious effort must be made initially to determine its prevalence in the population of the community served. Commonly, physicians have remarked on the reluctance of patients to divulge comorbid mental illness in the initial triage and interview period of their emergency department visit. Many times patients will only relay a psychiatric diagnosis if directly asked by the interviewer. This implies that the majority of patients with such disorders may never disclose their mental illness, and the treating physician may remain unaware of a vital part of the health history.

Research Question

What is the frequency of undisclosed psychiatric illness in a community based urban Emergency Department and can a simple screening method with specific questioning techniques improve identification of such disorders in this patient population?

Study Purpose

Our goal is to evaluate a simple screening method that could be integrated into the initial interview process. Using triage notes, patient history, review of current medications, and finally direct questions as to psychiatric illness, we hope to identify the presence of psychiatric illness and the use of psychoactive pharmaceutical among a cohort of patients not initially reporting such information.

Figure 2: Introduction, Research Question, Study Purpose

STUDY PURPOSE

Next create a paragraph titled *Study Purpose*. It is probably best to start out with the phrase "Our Goal" or something similar to get across WHY you are doing the study. This will essentially be what you intend to accomplish with your study by answering your research question. An example of this step is provided in figure 2.

Much of this page you will be used in the introduction of the manuscript you put together for publication. Don't worry too much about the style at this point as you will have plenty of time to change it before a paper is sent off. Just get your point across and try explain why it is important to do this study what you are trying to answer and why you became excited (or at least interested) about looking into this area of medicine. As noted before the *Research Question* is critically important as everything else in the study springs from this. It should be the item you have spent the most time thinking about up to this point.

STEP FOUR – A METHODS SECTION

Create a Methods Section: Now it gets hard, or at least slightly difficult for some. You need to create a *Methods section*. This will be two or three paragraphs relating exactly how you intend to collect the data, record the data, and analyze the data. In addition if your protocol includes using humans or information about humans by collecting their *Private Health Information* (**PHI**) you must discuss how you intend to provide *Consent* for the study, and how you intend to manage any PHI that could identify individuals involved in the study.

This section is an important part of the research protocol as it defines how you intend to answer the research question. Remember that an IRB exists to protect humans involved in human use research. Although an IRB will

usually review the "science" behind your methods, it is not their job to carefully review your statistical methods. Do not expect that review and approval of your research protocol by an IRB equals a "scientifically sound" protocol. You should discuss your methods with a mentor, research director, and / or a biostatistical consultant before writing. It is heartbreaking to have gone through the process of creating a protocol, getting IRB approval and collecting data only to find that there is a fatal statistical flaw in the study methods that invalidate everything. However, it is also true that one need not know a great deal about statistical analysis to design a great research protocol. That is, as long as you understand the research methodology and have selected the right research method for the question you want to answer. An example of this page is seen in figure 3.

Methods

This study is designed as a blinded prospective historical interventional study. All cases will be triaged and processed into the Emergency Department in the standard manner with no alteration in the usual clinical and administrative processes. The intervention occurs with the initial physician interview in the ED where the physician (prior to looking at the past medical history or pharmaceutical list) will ask the patient a short series of questions regarding psychiatric disorders and psychoactive pharmaceutical use. These answers will be documented on the patient encounter chart as usual. The remainder of the ED encounter will continue as usual with no alteration in care standards or procedures. At the end of the ED encounter patients meeting the inclusion criteria will be asked to read and sign a Health Insurance Portability and Accountability Act (HIPPA) form for use of their information in this study.

Case data will be abstracted onto the data collection form associated with this study (see attached). This form will be placed in a secure lock box in the Emergency Department and collected daily by the PI or co-investigators.

Data will be entered anonymously into a custom relational data base in Microsoft Access®. Descriptive analysis will be carried out and rates calculated. Comparative statistical analysis and analysis of associations will be carried out where appropriate. The results will be compiled, discussed, and presented in a scientific report.

Case Definition

All patients age 18 years and above are considered for inclusion in the study.

Inclusion Criteria

- Presenting complaint not identified as related to psychiatric illness.
- Patient age greater than or equal to 18 years of age.

Exclusion Criteria

- Age less than 18 years.
- Presenting problem is identified as associated with a psychiatric illness by the patient or ED health care providers.

Management of Study Records

All documents related to the research protocol will be maintained by the primary investigator in a locked limited access file in the Department of Emergency Medicine office, Integris Southwest Medical Center for the duration of time required by the IRB.

Figure 3: Methods Section

CASE DEFINITION

Next under the heading of *Case Definition,* describe exactly who will be considered for entry into your study. Be very specific. This is the study population or "sample" of the population you will be conducting your protocol with. They should be as much as possible, identical to the general population in your clinical practice; that are affected by the elements of the study question you are investigating. This is to allow you to *Generalize* your results from this study group to the entire population of patients in your practice.

INCLUSION AND EXCLUSION CRITERIA

In the next two headings, *Inclusion Criteria* and *Exclusion Criteria*, define carefully what is exactly needed to include a case in your study and what will disqualify a person from getting in to the study (some federal and local regulations may limit your selection criteria also). Remember that these criteria are vital in being able to enter enough cases in your protocol to get a valid sample of the population (so on one hand you want to be as inclusive as is possible). Conversely, there will be limits on the nature of people who may be entered in the protocol due to the presence of confounding variables (coexisting medical conditions, competing medications, etc.) or ethical considerations and legal restrictions (children, or other protected classes under Federal Regulation *45 CFR 46 subparts B-E*).

RANDOMIZATION METHOD

If your protocol will use any *Randomization Method*, this method needs to be clearly outlined under this step (although not shown in figure 3). The reader is referred to the section on randomization methods in this text in chapter 7 for a more detailed description if needed.

MANAGEMENT OF STUDY RECORDS

End this step with a description of the *Management of the study records*. This is important as the primary concern of an IRB is the protection of people involved in research. If patient identification is in any way possible from the study information (which it usually is) every possible protection must be afforded to prevent the inadvertent release of PHI from study records. Should always be safe and secured and under the management of the PI or a designee. In addition they should be kept on file for a period of time that will be defined by federal regulations and local IRB protocols. Following that time, after the study has been published the records should be destroyed in a manner that prevents their recovery. An example of how this might be written is as follows;

- **"**All cases diagnosed as having cutaneous abscesses will be asked to participate in the study and will be given a standardized *Health Insurance Portability and Accountability Act* (**HIPPA**) form for their signature. This form and all documents related to the research protocol will be maintained by the primary investigator in a locked limited access file in the Department of Emergency Medicine office for the duration of time required by the IRB. Following the completion of the study, publication of scientific reports and safe storage of study materials as required by regulations, the PI or associates will notify the IRB and all study materials with patient identifiers will be shredded or removed from digital storage devices. An anonymous database of study information will be maintained for archival purposes."

STEP FIVE – THE DISCUSSION

This section starts the Discussion. It may be as long or short as needed but should cover what is known about this topic area based on the literature search and articles you have already collected. What you are doing in this section is providing a more detailed description of your research question in the light of current research. Most of this will be used later (cut and paste) into the scientific article you will create with your study for publication. This section is referenced with the first study cited as number 1 and each other study numbered in order. An example of a discussion section is provided in figure 4.

Discussion

It has been postulated that a large number of ED visits occur in the mental health population, and that a majority of those patients are either reluctant to divulge their history or find it irrelevant. [1] It is our opinion, however, that this information is not only significant in regard to diagnosis but can be vital in terms of treatment options, drug interactions, drug reactions and compliance. Previous researchers have found that a large number of emergency department cases have underlying psychiatric illness. [2] Marchesi et al. found that a significant number of repeat ED visits were made by those with the diagnosis of depressive or anxiety disorder especially regarding somatic complaints. [3] Depressive symptoms and other disorders of mental health were very common among patients presenting to the ED with somatic complaints. [4] Others also have gone so far as to state that not only are ED visits increased but the utilization of laboratory and other diagnostic studies are greatly elevated in this same population often with no clear diagnosis found. [5, 6, 7, 8, 9]

The problem faced by busy Emergency Physicians is how, in a brief period, to successfully seek out and address issues associated with mental illness. Some researchers have used a (DSM)-IV screening questionnaire with good results. [10, 11] They were able to determine that repeat 911 utilization and emergency visits in a one year period were dramatically higher in those that screened positive for panic disorder. [12, 13, 14, 15] Although useful, (DSM)-IV screening can be time consuming and cumbersome.

Once the existence of psychiatric illness has been identified, the Emergency Physician may more efficiently address the patient's complaint with proper follow-up, social service or psychiatric consultation, as well as diagnostic and therapeutic modalities. If such a simple screening process proves useful in identification of undisclosed psychiatric disorders, further research may be carried out to see if the number of non-emergent ED visits is altered by identification and more efficient management of these underlying psychiatric disorders. Whatever the final outcome is, the simple fact is that the more we know our patients and know their disease the better we will be in both diagnosis and treatment therefore better physicians overall.

Figure 4: The Discussion

Most of this discussion will be intertwined with the discussion of the results you obtain from your own study later when you are writing your paper for publication. If well written to start with, it provides a good framework for comparison and contrasting your findings with the literature.

STEP SIX - REFERENCES

Finally attach a reference page citing the full references you have used in numerical order of their use. Do not use the endnote or footnote function of word processing software to generate a list of references. Number references consecutively in the order of their appearance in the protocol. Type the list of references in their order of mention in the text, not alphabetically. Abbreviate journal names according to index medicus. Indicate abstracts by "abstract" in parentheses. List the first 3 authors, followed by "et al" if there are more than 3. Figure 5 demonstrates a reference page.

References

1. Marchesi C, Giannini A, et al. The use of an emergency ward by patients with depressive or anxiety disorders: a one year follow-up study. International Journal of Psychiatry in Medicine. 2001;31:265-275.
2. Jones GN, Ames SC, et al. Utilization of medical services and quality of life among low-income patients with generalized anxiety disorder attending primary care clinics. International Journal of Psychiatry in Medicine. 2001;31:183-198
3. Fleet RP, Marchand A, et al. Comparing emergency department and psychiatric setting patients with panic disorder. Psychosomatics. 1998;39:512-518.
4. Katon W. Panic Disorder: Relationship to high medical utilization, unexplained physical symptoms, and medical cost. Journal of Clinical Psychiatry. 1996;57 (Suppl 10):11-18.
5. Klein E, Linn S, et al. Anxiety disorders among patients in a general emergency service in Israel. Psychiatric Services 1995;46:488-492.
6. Katerndahl DA, Realini JP. Where do panic attack sufferers seek care? Journal of Family Practice. 1995;40:237-243.
7. Yingling KW, Wulsin LR, et al. Estimated prevalences of panic disorder and depression among consecutive patients seen in an emergency department with acute chest pain. Journal of General Internal Medicine. 1993;8:231-235.
8. Johnson J, Weissman MM, et al. Service utilization and social morbidity associated with depressive symptoms in the community. JAMA. 1992;268:199-200.
9. Katerndahl DA. Factors associated with persons with panic attacks seeking medical care. Family Medicine. 1990;22:462-466.
10. Callahan CM, Hui SL, et al. Longitudinal study of depression and health services use among elderly primary care patients. Journal of American Geriatrics Society. 1994;42:833-838.
11. Katon WJ, Von Korff, et al. Panic disorder: relationship to high medical utilization. American Journal of Medicine. 1992;92:7S-11S.
12. Padgett DK, Struening EL. Influence of substance abuse and mental disorder on emergency room use by homeless adults. Hospital and Community Psychiatry. 1991;42:834-838.
13. Kelly TM, Donovan JE, et al. Psychiatric disorders among older adolescents treated in emergency departments on weekends: a comparison with a matched community sample. Journal of Studies on Alcohol. 2003;64:616-622.
14. Zane RD, McAfee AT, et al. Panic disorder and emergency services utilization. Academic Emergency Medicine. 2003;10:1065-1069.
15. Torres S, Nazario S, et al. Increasing the identification of depression in an inner city population. Academic Emergency Medicine. 2003;10:549-550.

Figure 5: The References

SUMMARY

The protocol you have just put together in this manner will provide you with essentially all the information you need to fill out any IRB application. In addition, this protocol will be used (in many cases verbatim) in a cut and paste style when you write your scientific paper. In essence, you have everything already written in a "first draft form" for the final paper except the data results, data analysis, discussion of the data in relation to current scientific knowledge, and your final conclusions. Organization of your scientific protocol in this manner promotes good reasoning, good science, and good publications.

FURTHER READING

1. DeRenzo E. *Writing Clinical Research Protocols: Ethical Considerations*. Academic Press 2005 New York, New York.

Hypothesis Testing – Part One

"*The man who smokes two pocks of cigarettes a day for twenty years and develops lung cancer faces problems which deserve our sympathy but he has no grounds for asking 'How could God do this to me?*' "

Harold Kushner (1).

INTRODUCTION

The progress of human scientific understanding of the universe rests not only with observing nature and creating theories, but with actually testing the validity of those theories. This comes under the general heading of hypothesis testing and will be covered in this chapter.

HISTORIC PERSPECTIVE

The history of how we really understand anything in nature revolves around the development of the Scientific Method, and the philosophical processes in its development. This historical process, although interesting, goes far beyond the purpose of this chapter – and the reader will therefore not be subjected to the full story. We will however review a brief segment of the process to put the discussion of hypothesis testing into context.

Human understanding of the world in which we live has undergone a number of progressive changes throughout the history of our species. In very ancient times human understanding was based on ideas about the world, observed or not, and the creation of mental concepts as to the causes of events. Such concepts constituted much of the ancient belief systems and usually attributed natural phenomena to the activities of spirits or gods., Such belief systems dominated the behaviors of most human societies (and still do to a large degree today) since before recorded history. These beliefs were rarely if ever tested for validity in any way as to do so was considered heresy.

RATIONALISM

Some of the first known attempts at changing this approach to human understanding began with the development of "the doctrine of rationalism" – mainly from ancient Greece and its precursor civilizations (2). Rationalism in its purest form simply holds that knowledge of the world (or the Truth regarding the world) comes through Reason rather than simple observation. An example of this process is the following;

The statement is made that "All Robins have red breasts," and one considers it to be true.
One then tells the rationalist "This bird has a red breast."
The rationalist at once infers from the above information that the bird in question is a Robin.

Although this is an improvement over simple observation, one can easily see potential and serious logic problems that can develop. This is due to the fact that other birds (such as Grosbeaks) also have red breasts. A rationalist could never make this differentiation, at least not without a lot more information.

EMPIRICISM

The next major change in the formal acquisition of knowledge came into play in the 1700s with the advent of Empiricism. In this setting one uses not only the rational process, but strict observation in a process termed *Inductive Inference*. Inductive inference requires one to develop a specific theory or possible explanation regarding a particular phenomenon based on direct, detailed, and repeated observation of that phenomenon. As example, regarding the Robins, the empiricist observes 100 Robins and notes that all 100 birds have red breasts. The empiricist therefore generates a theory stating "All Robins have red breasts." This theory is quite different than the Rationalists idea about all birds with red breasts being Robins. The empiricist makes a rationally inferred

statement regarding something they have observed, without venturing into areas that they cannot base on those observations. Because the statement is based on direct observation of Robins, it applies only to Robins and does not extend to Grosbeaks or other birds with red breasts. The best generalized response of the Empiricist to the statement "this bird has a red breast" would be "it could be a Robin." The specific Empiricist response to the statement "This bird is a Robin," would be "it has a red breast." Clearly the empiricist cannot observe all Robins in existence to confirm their theory, but they could "test" their theory by observing another 100 Robins to see if it holds constant. The more times this gets repeated, the greater confidence one has in the "Truth" of the theory. However, it is impossible to ever completely "prove" that the theory is true – not without actually observing all Robins in existence in nature. The empiricist approach however does represent an improvement over strict rationalism.

CRITICAL RATIONALISM

By the 1930s problems with the empiricist method began to escalate as attempts to understand nature became more complex. Karl Popper, an Austrian philosopher living in Great Britain, challenged the inductive empirical approach to science in favor of a so-called *Critical-Rationalism* using a deductive approach. Popper's premise is that knowledge accumulates by a deductive process of reasoning through falsifying hypotheses that one has generated in their mind. That is, one creates a hypothesis about nature in their mind (based on observations or not) then sets about testing that hypothesis in order to prove it false. This makes more sense because one can actually prove a hypothesis as false (it only takes one Robin with a breast color other than red to prove the hypothesis that all Robins have red breasts as false) (3).

The difference may seem subtle but it has a profound impact on how we actually conduct scientific investigations into nature and how one may apply statistical analysis to research. With some modifications of the strict Popperian approach, we arrive at the method of hypothesis testing used by most investigators today. It is important to note that not all research conducted actually requires hypothesis generation. This is particularly true when one is interested in empirical research into a phenomenon in which information gathering about the phenomenon is the goal. However when one is attempting to prove a specific point, relationship, or causality – a hypothesis is required.

THE HYPOTHESIS IN HYPOTHESIS TESTING

Based on the above discussion most research conducted proceeds as follows. One develops a theory about how something exists or works in nature – eg. "All Robins have red breasts." The opposite theory to this would be "All Robins do not have red breasts." Because the theory that all Robins have red breasts can never actually be proven, we denote it as the *Alternate Hypothesis* (H_A). We can prove the H_A false however by finding a single Robin with a breast color other than red – which is much more likely to happen than that we will actually account for every Robin on the planet and see that it has a red breast. The other theory - that all Robins do not have red breasts - is called the *Null Hypothesis* (H_O) and represents where we started before we came up with the H_A. It does not matter if the H_O is "true" or not, it simply represents the baseline or current state of understanding of the universe before we suggest the H_A. We will collect evidence about the H_A that will either support or not support it. If the evidence is not strong enough to support the H_A we will stay with the H_O. If the evidence is strong enough for the H_A, we will reject the H_O in preference of the H_A. Because the H_A is not totally provable, we only accept it over the H_O by some degree of confidence and probability, and only when there is sufficiently strong evidence (usually set at a 5% chance that we might be wrong in rejecting the H_O for the H_A). The level of risk we allow ourselves for making an error and the sort of errors that might be made are covered in the next chapter.

Before we can further discuss how a hypothesis is tested, we need to cover some basic principles of how information (data) is graphed and the distribution of such data on graphs. This will allow us to discuss the application of statistics to accepting or rejecting a particular hypothesis.

THE NORMAL DISTRIBUTION OF DATA

Graphing the frequency of occurrence of data values in a frequency distribution graph is one of the most efficient and intuitive methods to observe and evaluate the nature of a parameter collected on a population. Figure 1 demonstrates a generic frequency distribution plot in which the value of the parameter measured (x) is plotted on the horizontal – x axis and the frequency of occurrence is plotted on the vertical – y axis. The figure demonstrates a *Normal*, or *Gaussian* distribution of data in which the data are equally distributed on either side of the mean. This composes the classic "Bell" curve of data that is rarely seen in perfection in nature but commonly approximated. However, most observations in population studies will very closely approximate a normal curve of distribution – which becomes important in the analysis and inference that can be made regarding these data. Data in a generally normal distribution allow one to greatly simplify the mathematical analysis of data. A term for statistical tests that manage normally distributed data is *Parametric*. Data distributions that are "lopsided" around the central tendency require a different statistical management and are termed *Non-parametric*. Note that the measures discussed in later chapters regarding central tendency (mean, median, mode) are evaluations of the way the data collect about a central region on the graph while the measures of data dispersion (Range, Quartiles, Deviations from the mean, Variance, and Standard Deviation) are all concerned with the way the data spread out from the mean.

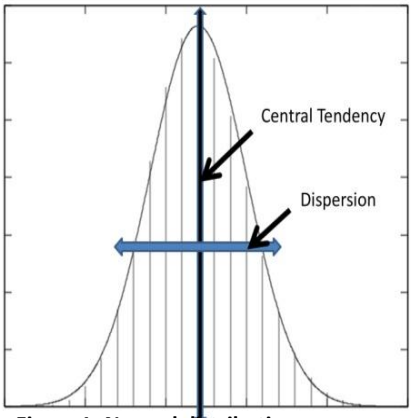


Figure 1: Normal distribution

CENTRAL LIMIT AND NORMAL DATA DISTRIBUTION

The Central Limit Theorem simply states that;

"Where a sufficiently large number of data values (each with a finite mean and variance) are collected, the frequency distribution of these values will closely approximate a normal distribution."

Because almost all data sets collected on population parameters have a finite variance, the central limit theorem accounts for why the appearance of nearly normally distributed data is so common in nature. Because of this tendency towards normal distribution, we can make some valid assumptions about the populations behind the data, and thereby simplify the analysis of such data. This is not to imply that all data will be sufficiently normal in distribution, but this will be discussed later.

LINKING HYPOTHESIS TESTING TO STATISTICS

When researchers seek to understand something in nature, they formulate a research question which eventually takes the form of a hypothesis termed the Alternate Hypothesis (H_A). When we think about the nature of research "questions" expressed in general language, they take on a reasonably straight forward and understandable format, such as;

"Do antibiotics reduce the wound infection rates in simple traumatic wounds in the diabetic population?"

What such a research question really does is make a statement about the underlying nature of some population outcome (parameter) measured under some specific experimental condition. As example, if we look at the

research question above, the implication is that if we measure the mean infection rates of diabetic patients with wounds not treated with antibiotics (μ_1) and the mean infection rates for those treated with antibiotics (μ_2) – and we further assume that there will be a difference between the two means (which is what the H_A states) then this can be expressed as;

$$\mu_1 \neq \mu_2$$

More specifically if we believe that the patients treated with antibiotics will have a lower overall rate of infections, we may predict that;

$$\mu_1 > \mu_2$$

It is in this manner that the general language of the research question can be made to relate to the underlying statistical nature of the issue being investigated. It should be noted here that the ability to relate the research question to the underlying mathematical relationships is a direct function of the type of study and the specific study design being used.

If one looks at what a hypothesis "asks" from a statistical point of view, the topic of hypothesis testing becomes easier to understand. Therefore, the nature of the hypothesis is as follows;

(a) Statistically a hypothesis is a statement about how the data, when measured, will be distributed on a frequency graph.
(b) The hypothesis usually makes a statement about some underlying parameter of the test population, such as "μ_2 = 10."
(c) The hypothesis also makes some statement about the relationship of the two means and the probability distribution of the data. That is; that there is or is not a statistical relationship between some specific outcome and variables involved in causing the data distribution.
(d) The hypothesis also makes some statement about the population parameters such as $\bar{\mu}_1 = \bar{\mu}_2$, i.e. that something is true or false.

When researchers engage in research, and hypothesis testing – what they really do is test the evidence for or against supporting the H_O. The H_O states that there is no difference, or no relationship, between the frequency distributions of the population parameter measured of the test groups. This is analogous to the presumed innocence of an individual charged of a crime under the United States Constitution. Under the H_O, any differences between groups that are observed (such as differences in $\bar{\mu}_1$ and $\bar{\mu}_2$ in the wound study mentioned previously) are attributed to CHANCE only.

The Alternate Hypothesis (H_A) is in some sense a contradiction of the H_O. This is analogous to the charge being made by the prosecution in a jury trial. That is, just as the prosecutor believes that the individual charged with a crime is guilty of the crime, the researchers think that there **IS** a difference between study groups. The researchers then are required to devise a proper experimental process and collect data in order to provide "evidence" that the H_O cannot logically be supported and that the H_A is the most likely alternate explanation of any differences seen in study groups. This is again analogous to the prosecution in a jury trial providing properly collected evidence to support their claim. In this setting, the prosecution must provide sufficiently strong evidence to meet the criteria set in the minds of the jury in order to convince them that the H_A is more likely than the H_O. Statistically then, this would mean that any difference noted between study groups (e.g. $\bar{\mu}_1 \not\approx \bar{\mu}_2$) is real and not due to random chance (4).

DEFINING A HYPOTHESIS TEST

A hypothesis test is a decision making process that allows one to examine the set (or sets) of data collected by some research design process. On the basis of what one expects to find if the H_O is true (which can be statistically estimated) one can make a decision as to whether or not to reject the H_O. If the evaluation of the evidence makes us reject the H_O, then we are left by default with the H_A as the explanation of any observed differences between the test groups. Technically one never "proves" the H_A as the "truth" but rather accepts the H_A the most likely explanation of the differences observed if there is insufficient evidence to support the H_O.

GRAPHICAL PRESENTATION OF HYPOTHESIS TESTING

As noted before, various parameters of a given population may be demonstrated in a frequency distribution graph. Such a display may be used to demonstrate how statistical analysis in hypothesis testing is actually linked to the data obtained during research.

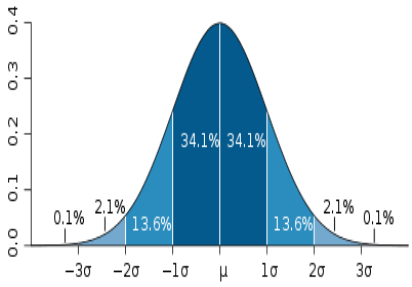


Figure 2: Data Distributions

First we need to take a look at what the *normal frequency distribution* graph tells us about the population it represents (see figure 2). When researchers base their analysis of a hypothesis on the presumption of a normal distribution of data, we can make a number of assumptions about the distribution of data on that frequency curve. In the figure above, the Standard Deviation (a statistical measure of dispersion) is plotted on the x-axis. In a normal distribution we can assume that about 68% of the population data will lie within 1 standard deviation of the mean, 95% within two standard deviations, and 99.7% within three standard deviations. Knowing these values (68-95-99.7) actually allows one to perform several important statistical tests on data that closely approximate a normal curve. These values actually serve as some of the major assumptions that allow one to use simplified parametric tests.

SUMMARY

Understanding the logical limits of human reasoning in the development of hypotheses, makes the process of hypothesis testing more understandable. The nature of data distribution and its tendency to accumulate around some central tendency such as means, is critical in our ability to analyze, infer and make predictions about parameters in populations. The next chapter will cover a number of statistical terms and concepts.

REFERENCES

1. Kushner H. *When Bad Things Happen to Good People*. New York Avon Books 1981.
2. Flew A. *Dictionary of Philosophy*. Atesbury Great Britian, Laurence Urdang Associates Ltd. 1979.
3. Popper KR. *The logic of Scientific Discovery*. New York. Basic Books 1959 (Reprint of the 1935 edition).
4. Aschengrau A, Seage GR III. Essentials of Epidemiology in Public Health (2nd Edition) Jones & Bartlett Sudbury, MA 2005

Hypothesis Testing – Part Two

"There are three kinds of lies: lies, damned lies and statistics."

Mark Twain (1)

INTRODUCTION

Statistical tests are usually thought of as very objective and scientific methods of answering research questions. In reality, they are no more or less objective than any other estimation method used in science. This section provides a brief review of the major mathematical relationships and terms used in most statistics.

PARAMETERS AND ERRORS

PARAMETERS

Any numerical characteristic of a population is called a parameter.

The target of any scientific investigation is to evaluate a population for some specific characteristic of interest (parameter). Examples of parameters would be the mean systolic blood pressure of a population or the proportion of a population that responds favorably to a specific pharmaceutical agent (2).

If we were able to measure the blood pressure on all members of a population (say the entire population of the State of Oklahoma) we would have no issues with arriving at accurate parameter measurements. We would also be able to determine if some entity (a drug or treatment) altered a parameter in the population from the known mean. Unfortunately it is impossible to measure the blood pressure of the over 3.5 million individuals in the state. Therefore we are limited to taking only a sample of the overall population in such a way so as to be able to generalize by inference our findings to the larger total population. This will be discussed more when we cover sampling techniques in research.

ERRORS IN STATISTICS

Just as the legal system may convict an innocent man or free a guilty man, statistical tests can (and do) make mistakes. When the researcher uses statistics to make comparisons of groups – they evaluate some parameter of those groups for differences between each other. The statistic used, generally provides a probability value that any difference seen between the groups – is due to random chance only, and not a real difference when the probability values is low enough (usually set at 5%) then researchers generally reject the idea that the difference seen is due to chance in favor of the difference being "real." From a statistical point of view, researchers consider such probabilities as being "significant." This idea of significance and the concept and limitations of "p" values will be discussed in greater detail in the next chapter. For now let us continue and consider with kinds of errors can be made in this context. There are four decisions that may be arrived at with hypothesis testing. Two of these decisions are correct, and two of these decisions are errors (3).

TYPE I AND TYPE II ERRORS

If a mistake is made such that one believes that statistical significance is attained from data, when in fact there is NO statistical significance, this is known as TYPE I ERROR or alpha error. This is analogous to convicting an innocent man by jury trial. Conversely, if a mistake is made and one does not obtain statistical significance with a statistic, when in fact there IS statistical significance, this is termed a TYPE II ERROR or beta error. This is analogous to releasing a guilty man after trial by jury.

As noted before, most tests are designed to limit Type I (alpha) Error to 5% or 1% probability. This is done by the researcher at the start of the research protocol when they decide what probability of error they will accept during the data collection process.

There is no customary level of Type II (beta) error for research however the Type II error is known to be linked to several factors in a research process (3).

- (a) How much is going on? That is, how serious is the crime being committed? How dramatic or large of a change or impact are we looking at in this study? If there is a large impact involved, then the chance for a large Type II error becomes very small.
- (b) How much noise (variability) is there in the data? This is analogous to the quality of the evidence collected for the jury in a trial. Poor quality data or difficult to collect data with wide variance results in higher levels of Type II Error.
- (c) What is the size of the study? Just like the amount of evidence collected for a jury, the more or larger the study – the less likely it is for a Type II Error to occur. Therefore, the larger the sample size, the less likely the chance for a Type II Error.

Remember that Statistical Significance is not Synonymous with Practical Importance.

CENTRAL TENDENCIES IN DATA COLLECTION

When one collects data on some parameter of a population, the values of the parameter measured have a tendency to collect around some common value. The more individual data points on a parameter that are collected in a study, denser the clustering of these values will be. This is termed the "*Central Tendency*" of data. Measuring the central tendency in data is usually the first step in data analysis (2, 3, 4).

THE MEAN ($\bar{X}$) OR ($\bar{\mu}$)

Given any data set with the size of n,

$$\{x_1, x_2, \ldots, x_n\}.$$

The mean of all the x values is represented by the symbol ($\bar{x}$). Means relating to populations are usually represented by the mu symbol ($\bar{\mu}$) which will be used in this text. It is computed by summing all the x's in the set and dividing the sum by n.

$$\bar{\mu} = \frac{\sum x_n}{n}$$

The mean is the most commonly employed measurement of statistical central tendency. It is easy to calculate and lends itself to more advanced and detailed statistical methods. The term mean, is used ot differentiate this measurement from the "average" for the following reason. If the set of data values in the analysis consists of the full set of collected observations or measurements – and these values are not a statistical sample of a larger population, then the calculation may be referred to as a simple average. However, if the data values represent some statistical sample of a larger population – the term mean should be used.

THE MEDIAN

If one divides the data set into ascending or descending order, the median is the number that divides the data set into equal halves. If the number of observations (n) is odd, then there will be a unique median; the ½(n+1)th number from either end in the ordered sequence. If n is even, there is "strictly" no middle observation, but the median is defined by convention as the average of the two middle observations – the ½(n)th and ½(n-1)th from either end of the set.

The advantage of using the median measurement is that it is less impacted by extreme observations. It is however less sensitive in that it does not account for the precise magnitude of the observations. If two groups of data are combined, the median of the combined group cannot easly be expressed in terms of the medians of the two component groups. Again, this results in a loss of resolution or information about the nature of the data in the groups. Finally, in large data sets, finding the median requires more work to calculate than the mean – and it is not of much use in complex statistical calculations.

THE MODE

This is the third measure of central tendency commonly used. The mode is the data parameter value most commonly found in the data set (ie. the value at which a frequency polygon of the data reaches its peak). The mode is not commonly used except for its descriptive nature. It can be highly ambiguous because fluctuations in the data often create spurious modes.

MEASURING DISPERSION WITHIN DATA SETS

Once a mean (or other measure of central tendency) has been determined for a set of observations, one then measures the degree of variation or dispersion around the mean. That is – are all the values tightly grouped close to the mean, or are they dispersed widely? This measure of dispersion is important not only for descriptive purposes, but also because it plays an important role in the further statistical analysis and inferences of the data set (2, 3, 4).

THE RANGE (R)

This is the difference between the largest and the smallest value in the observation set. The value of the range is determined by only two of the entire set of observations. The value of the range depends (due to statistical reasons) on the number of observations made.

PERCENTILES (QUANTILES)

Percentiles represent the percent of data values below a selected point in the data table distribution – when the data are arranged in descending order.

INTERQUARTILE RANGE

This measure is specifically defined as the range (distance) between the 75^{th} and 25^{th} percentiles. This is usually smaller than half the range of all the values because of the tendency of the values to cluster around the central values.

DEVIATIONS FROM THE MEAN ($\mathbf{x - \mu}$)

This is an alternative to the range in describing the dispersion from the central tendency. It also serves as the basis for the development of several other measurements of dispersion that are extensively used in statistical analysis. The deviations from the mean are determined by the following relationship (2, 3, 4);

$$\boldsymbol{x - \mu}$$

The greater the variation in the data set, the larger the magnitude of these variations will be.

Taking the mean of the deviations is useless because;

$$\sum (x - \bar{\mu}) = 0$$

One can take the mean of the absolute values of the deviations but it is more difficult to manage mathematically. The real use of deviations from the mean come into play in the determinations of variance and standard deviation.

VARIANCE (σ^2)

From the deviations from the mean above, the variance (σ^2) can be computed by squaring each deviation, adding them, and dividing their sum by one less than n;

$$\sigma^2 = \frac{\sum(x - \bar{\mu})^2}{(n-1)}$$

Using *n* – 1 in the equation is not really that important with very large data sets, but becomes very important with small values of *n*.

Variance becomes a critical element in the use of several statistical processes in data analysis and hypothesis testing. It should however be noted that σ^2 is measured in the square of the units in which the x's of the data set are measured (eg. if x is in seconds, then σ^2 is in seconds-squared). Therefore, when one expresses variance, it would be convenient to have the measure of variation expressed in the same units as the x's. This can be done by calculating the other major measure of dispersion from the central tendency – the standard deviation.

STANDARD DEVIATION (σ)

The standard deviation (σ) is calculated by taking the square root of the variance;

$$\sigma = \sqrt{\frac{\sum(x - \bar{\mu})^2}{n-1}}$$

As with the σ^2, σ is a key component of statistical tests commonly used on data in research. This is particularly true for the sorts of analysis and research done in Emergency Medicine.

SAMPLE SIZE IN RESEARCH

One of the most frequent and important questions asked when a research protocol is being designed is "how many patients do I need in this study?" This is critical because if there are insufficient cases in the analysis, one will be unable to see any true differences in groups above all the noise. Fortunately we have some methods that can be used to estimate the number needed before the study starts. Although we tend to think of statistics as a "hard science" they are in reality only estimates of the truth and provide us with answers in terms of probability only. As such, you will see that even these determinations of the needed sample size are also based on "educated guesses" about the population and the issue being studied (3).

STEPS IN SAMPLE SIZE CALCULATION

In general, there are four primary steps taken to estimate the sample size needed for a particular study. They are as follows;

1. Select the values you want for alpha and beta errors. These should be selected after considering the clinical or scientific context of the question and the relative importance of committing a type I or type II error.
2. Determine the size of the difference between the treatment groups you want to seek in the study (this is equivalent to the alternative hypothesis). The size of the difference depends on the clinical and scientific context of the study and on the practical considerations regarding the size of the population available for study.
3. Using the guidelines above, determine the appropriate statistical test for the type of data that will be obtained by the study.
4. Look up the required sample size in reference tables or calculate the sample size from available formulae. Reference tables are most readly available for the t test, chi-square test, and Fisher's exact test. Even when the requirements of these tests are not strictly met, the sample size tables for them may be used to gain a rough idea of the required sample size.

If the study and statistical test needed for the study is somewhat unusual, sample size tables will probably not be available. These require more complex evaluation and one should probably consult a statistician for assistance. However even if one intends to consult a statistician one should still go through the first three steps as this information will still be needed.

ALPHA, BETA AND SAMPLE SIZE

There is a trade-off between the alpha value (the risk of a type I error) and the beta (the risk of a type II error) and the sample size in a research study. For any given sample size, the smaller the alpha, the larger the beta. To balance this relationship (that is to get the alpha and beta to acceptable values) one can only increase the sample size. It is this relationship that is published in tables for various types of statistical tests or through calculations. If the impact of a particular exposure or treatment is small (as example a new blood pressure medicine changes the blood pressure by only a few mm/Hg) then the sample size needed to see that impact on a test population will have to be large. We almost always set the values of alpha at 0.05 and beta at 0.20. But these values are totally arbitrary and are selected simply because of tradition. They do however represent reasonably good values of allowable error when we do studies (2, 3, 4).

It is important to understand that most studies are actually biased such that you are less likely to make a false positive conclusion and more likely to make a false negative conclusion. The critical point here is that alpha, beta, the impact of the treatment or exposure and sample size MUST be estimated BEFORE the study is initiated. This should be part of the protocol submitted to an institutional review board (IRB).

THE REALITY OF ESTIMATING SAMPLE SIZE

In practice, estimation of the sample size can be difficult. For reasonably small studies with basic statistics, most resident researchers can make their own size estimations using available resources. For anything much beyond this, a biostatistician should be consulted.

SUMMARY

Understanding the nature of data distributions around a central tendency and how such tendencies are described and managed statistically is important to the basics of data analysis and testing a hypothesis.

References

1. Twain M. Mark Twain's own Autobiography: The Chapters from the North American Review (Wisconsin Studies in Autobiography) Kiskis MJ (Editor). Madison, WI. University of Wisconsin Press. 2nd Edition 2010.
2. Aschengrau A, Seage GR III. Essentials of Epidemiology in Public Health (2nd Edition) Jones & Bartlett Sudbury, MA 2005.
3. Le CT. 2001. Health & Numbers: Problems based introduction to biostatistics 2nd Ed. Wiley-Liss New York, NY. Pp. 56-79.
4. Jekel FJ, Katz DL, Elmore JG (2001) *Epidemiology, Biostatistics, and Preventive Medicine* (2nd edition) W.B. Saunders Co. Philadelphia, PA

Statistics - Sampling and the "P" Value

INTRODUCTION

In the process of data analysis, one wishes to make the strongest case (positive or negative) with regard to the hypothesis under investigation using the limited amount of data one has available. There are two major problems that the researcher must deal with in order to approach the "truth" in any collected data these are - the impact of biological variability and problems with experimental imprecision. It is human nature to see patterns in randomness. If you stare long enough at the random static on a television screen when it is receiving no signal, it is well known that one will eventually start to see coherent images in flashes across the screen. The same phenomenon is seen in the interpretation of research data, particularly if the data has some degree of emotional connection to the researcher. We want to see patterns and relationships in nature, irrespective of the contribution of random variability. It is only through the use of statistical analysis that one may minimize the impact of the human propensity to see such relationships when they in fact do not exist.

In the same context however, one needs to understand the limitations of statistical analysis of data. There are times when a simple descriptive review of data are so profoundly obvious, that performing a statistical test on the data are simply not needed. As example, the relationship between cigarette smoking and lung cancer has never been "proven" statistically. However, simple observation of the data demonstrates an overwhelming association that does not require statistical validation (1). This supports the aphorism: if your data speak for themselves, don't interrupt. It is in the setting of small differences (usually because the impact on a population of a particular exposure or disease is relatively small in proportion to the entire population at risk) that one must use statistics to expose the effect of interest.

SAMPLING ISSUES

When the researcher approaches the analysis of any data, there are several logical assumptions that are made. These are (1) that the data one has collected is randomly collected, (2) that any association found in the study population is to be applied only to that study population. If these two assumptions were true, most of the problems with data analysis would be solved. The reality of research is however, qite different. Most of the time the sample population used for data collection is not random (although there are statistical methods available to decrease the impact on nonrandom sampling to some extent). In addition, one always wants to be able to apply any results from the sample population to a wider population within the community as a whole. This desire to generalize the results to a larger population is what makes it so important that steps be taken to ensure that the population of individuals on which the data are collected (the study population) be as much like the larger population group as possible (2).

INDEPENDENT SAMPLES

Another important point about sampling in data analysis is that each of the individuals of the study population is sampled or tested independently from the other members of the study population. The importance of independent sampling may be thought of with the following example.

Suppose that one is measuring blood pressures in test subjects. They are divided into two groups of 5 individuals each. One group is being given a drug that is thought to lower blood pressure while the other is not. Each individual in each group has their blood pressure taken three times. This gives 15 blood pressure values for each group. It is important to note, that these are NOT 15 independent blood pressure readings (samples) from each group. This is due to the fact that the three readings taken on a single individual are much more likely to be very similar in value than any readings taken between two different individuals – as such they are NOT independent

values. One can however take the mean value of the three readings on each individual – which can be used as an "independent" mean sample and treated as such.

This concept of independent sampling can also be looked at in the context of the sample population. As example, suppose researchers are conducting a study in multiple Emergency Departments. In this study, 20 participants are enrolled from an inner-city ED and another 20 from a suburban ED in an affluent neighborhood. The blood pressure values taken from the inner-city cohort are much more likely to be similar to each other than they are to the cohort from the suburban ED (and vice versa). As such, these are not really independent samples and bias is going to be introduced into the study. Again, there are some statistical methods that can help reduce the impact of this non-independence, but only to a limited degree.

THE "P" VALUE

The p value is a commonly used result of several statistical processes in the analysis of data. We will review the definition, use and limitations of this statistic.

P VALUE DEFINITION

The "P" value is defined as a probability expression ranging from 0 to 1 which expresses the probability that random chance would lead to the same level of difference in some value you are measuring between groups in a study population (3). As example, suppose one takes the mean values of systolic blood pressures in two groups from the previous blood pressure study example and find that they are somewhat different. The p value calculation tells the researcher what the probability is that random chance would lead to a difference between those blood pressure means the same or larger than the result that has been observed.

CALCULATION OF THE P VALUE

As with most statistics there are a number of ways to calculate a p value. The issue of primary importance for the resident researcher is not the mechanics of the calculation, but rather the selection of the proper statistical test to calculate the p value. This selection will be determined by the type of data collected, the distribution of the data (parametric or nonparametric), the size of the population sample and how the data are to be expressed (simple values, survival times, etc). The most common tests used by researchers to calculate p values are demonstrated in Table 1 (4). Each test will be briefly discussed in chapter 12.

Number of Groups Compared	Independent Samples	Paired Samples
Groups of Nominal Data		
2 or More	Chi-Square Test Fisher's Exact Test	McNemar's Test
Groups of Ordinal Data		
2	Wilcoxon rank-sum test Mann-Whitney U test	Wilcoxon signed-rank test
3 Or More	Kruskal-Wallis test	Friedman one-way ANOVA
Groups Continuous Data		
2	Student's t test Wilcoxon rank-sum Mann-Whitney U test	Paired t test Wilcoxon signed rank test
3 or more	ANOVA or F test Kruskal-Wallis test	Repeated-Measures ANOVA Friedman one-way ANOVA

Table 1. Common Statistical Tests for Significance

In modern research, most data are entered into computerized relational data bases. Most of these computer programs have intrinsic methods available to calculate basic statistics such as p values. A such, the raw details regarding how to calculate p values are beyond what the resident researcher wants, or needs, to know. What is important is understanding what method of significance testing to select depending on the data available.

INTERPRETING THE P VALUE

Consider the following. In the example of the groups measuring blood pressure previously used, suppose one group is given some drug that is thought to decrease blood pressure while the other group is not given the drug.

Now we measure the blood pressures and take the mean value for each group. We note that the mean blood pressure values of the two groups are somewhat different. We calculate a p value to see if the difference we are seeing between the two mean blood pressure values is "significant" or not. The result obtained is;

$$P = 0.03$$

This p value means that there is a 3% chance of observing the same or greater difference that we are observing between the two groups due to random chance if we run the experiment again on identical individuals from identical groups.

ERRORS IN P VALUE INTERPRETATION

The researcher must not fall into the mind-set of thinking that a p value of 0.03 means that there is a 97% chance that the difference observed between the groups is "real" and that there is a 3% chance the difference is the result of random chance. The p value does not and cannot tell the researcher this. What it actually does say is that if you repeat the experiment randomly from an identical study population, one will find a difference smaller than was observed 97% of the time and a difference the same or larger than observed only 3% of the time due to random chance (3, 4). This boils down to rather the researcher wants to "bet" on a 3% coincidence that there is no real difference in the groups. Most individuals would not take such a bet. As such we would usually consider the p value of 0.03 as "significant."

HOW SIGNIFICANT IS SIGNIFICANT

As was noted in chapter 10, researchers usually set their level of significance (or α) at 5% or a p value of 0.05. This then means that any p value < 0.05 is considered significant by definition. It would seem intuitive that a p value of p = 0.0001 should be "more" significant than p = 0.04. This is however not the case. The p value can give the researcher information about the study parameter in terms of probability of getting some particular level of difference between two groups. It cannot tell the researchers anything about how "likely" something is to happen as a result of a particular drug or exposure (4). This is one of the limitations of the p value statistic and it is important to understand. Although some authors will modify or tag various p values as significant, "very significant" or "extremely significant" based on the p value, there is really no scientific justification for this terminology. The p value is simply significant or not significant according to the definition we impose on the test (the α of 0.05). Other statistics must be used to determine the nature of the impact of a particular issue on the differences observed between groups. Some of these methods will be discussed in later chapters.

ONE AND TWO TAIL P VALUE TESTS

Without delving into the statistical theory behind the "tails" associated with p value tests, one should none the less understand the following. When we are looking at the difference of some parameter between two groups, there is usually one of the groups that the researchers expect to manifest the change or difference. In the example of the blood pressure study we have been using, this would be the group being given the drug and it would be expected to lower the blood pressure in that group. In this setting, the researchers would know ahead of time which group will manifest the change, and in what direction the change will be.

When a researcher selects a statistic to calculate the significance (p value) of an observed difference, the option of a two-tailed or one-tailed test is usually available (and often presented as a selection choice in computer programs). The two-tailed test answers the question: Assuming the null hypothesis (that there is no difference between the two groups) what is the probability that randomly selected measurements from the two groups would have means as far apart as observed – with either group having the larger mean? The two-tailed p statistic provides a p value without consideration of which group manifests the change or what direction the change is in.

The one-tailed p statistic however requires that the researcher predict which group will have the larger mean BEFORE any data are collected. Therefore, the one-tailed p value answers the question: Assuming the null hypothesis (that there is no difference between the groups) what is the chance that randomly selected measurements from the two groups would have means as far apart as observed knowing that one of the groups (which was specified before data were collected) will have the larger mean. In this setting, again without going into the details of the math, the primary advantage of using a one-tailed p value test is that it is considerably less complicated to perform the calculations. What advantage this actually presents for the researcher in this day of computerized statistical calculations is dubious at best. This is particularly true when the disadvantages and potential problems associated with one-tail tests are considered.

There is probably little good reason to use a one-tailed p value test currently (3, 4). This is due to the potential problems and limitations of the one-tailed tests and the advantages provided by use of the two-tailed tests. As example, when one calculates confidence intervals (discussed later) the relationships between the p value and confidence interval is more clear when the two-tailed p value statistic is used. In addition, when one is calculating p values in comparison of more than two groups, the concept of one or two tails no longer has any meaning (as mathematically the groups have "multiple" tails). In this setting, the p value will more closely mirror reality if the two-tailed test is used.

One more consideration regarding the one or two-tailed p value test is as follows. Suppose one thinks that a particular group will be the one impacted by the issue being studied and that the impact will be in a particular direction. However, when the data are analyzed, the researcher finds that the direction of the impact is opposite that anticipated. The temptation here would be to simply switch to a two-tailed test or actually reverse the direction of the experimental hypothesis. Neither of these options is acceptable for multiple scientific, statistical and ethical reasons. The only option is to assume that the change seen – no matter how large – is the result of random chance. It is best to avoid this situation by avoiding the use of a one-tailed statistic from the start.

SUMMARY

A commonly calculated component of statistical data analysis is the p value. The p value provides the researcher with a probability that any observed differences between groups is due to random chance – when properly calculated. Understanding the limitations of the p value can prevent errors in the statistical interpretation of study data.

REFERENCES

1. Witschi H. A Short History of Lung Cancer. Toxicol Sci 2001;64(1):4-6
2. Le CT. 2001. Health & Numbers: Problems based introduction to biostatistics 2nd Ed. Wiley-Liss New York, NY. Pp. 56-79.
3. Aschengrau A, Seage GR III. Essentials of Epidemiology in Public Health (2nd Edition) Jones & Bartlett Sudbury, MA 2005.
4. Jekel FJ, Katz DL, Elmore JG (2001) *Epidemiology, Biostatistics, and Preventive Medicine* (2nd edition) W.B. Saunders Co. Philadelphia, PA.

Select Statistical Concepts and their Usefulness in Medical Research

INTRODUCTION

With the advent of computerization and readily available relational database programs such as Microsoft Excel, the performance of most simple statistics is a matter of placing the database in the correct format, selecting the data to be analyzed and telling the computer what test to perform on the data. Large statistical packages such as Statistical Analysis System© (SAS) software can cost many thousands of dollars and be difficult to use. Such software provides a high level of flexibility and capacity for professional statisticians, but is rarely needed for the average researcher in medicine, particularly at the resident level. Simple statistical programs are available at the hundreds of dollar level that can often meet the needs of the resident researcher. As example, the Centers for Disease Control and Prevention sponsors a free software program downloadable from the CDC web site that can perform almost all basic statistical functions needed in epidemiological research called Epi-Info© (1). The author uses the statistical package available in Microsoft Excel and a Statistical add-on package called Analyze-It© for a total cost of under $1000. Most statisticians will have their favorite statistical software – which is usually the one they trained on.

The point of the above is that few statistical tests are carried out with a calculator and stubby pencil in the current era. In addition, most medical researchers can easily be familiar with using a simple statistical package of some sort to perform the basic data analysis on their own study. This does not obviate the need for a statistician, but under such professional direction, can streamline the process of data analysis and interpretation.

This chapter will briefly cover a few select statistical tests often used in medical research. The reader will be spared the mathematical derivation and proofs of these tests in the interest of brevity. In addition, as noted above, one does not necessarily need to know the mathematical process to use the test due to the availability of modern statistical software. Those interested in the details are directed to the references for this chapter for further reading.

PARAMETRIC VERSES NON-PARAMETRIC TESTS

Recall that data that are grouped in a reasonably “Normal” distribution around the mean are termed parametric – allowing one to use greatly simplified statistical tests – termed Parametric tests. Data that are not “Normally” distributed and skewed one way or the other around the mean, are termed non-parametric and require more complicated statistical management under a set of statistics termed Non-parametric tests (2, 3).

PARAMETRIC TESTS

THE STUDENT’S T TEST (PARAMETRIC)

This test is used to determine if the means of two sets of data are the same (within the margin of error). The t test assumes that the data are normally distributed and that the data from both treatment groups have a nearly equal variance. There are different forms of the t test available which should be used depending on the design of the study (i.e. if the two data sets or groups are paired or not).

THE PAIRED T TEST (PARAMETRIC)

The paired t-test is used when there is a single outcome variable to be measured, and there are two nominal variables that can impact it. One of the nominal variables must have only two possible values. This is usually used in the context of one nominal variable representing different individual subjects in the study and the other being “before” and “after” some treatment or exposure. In this setting the measurements are “paired” to a single

individual through time in the before and after setting. The pairing may be spatial as well as in comparing a left and right extremity or some individual inside or outside of some structure.

An example would be the performance of residents on a test of manual dexterity before and after drinking an energy drink. For each resident, there would be two observations, one before the drink and one after. One would expect the residents to vary widely in their performance, so if the drink increased their mean performance by only 5% it would take a large sample size to detect the difference if the data were analyzed using a student's t test. Using a paired t-test has more statistical power when the difference between groups is small relative to the variation within groups, and this is the primary advantage to its use (4).

CHI-SQUARE TEST (PARAMETRIC)

This statistic is usually used with categorical variables in testing the null hypothesis to determine if there is a difference in the outcome of groups based on some treatment. The test assumes that there are at least 5 observations of each combination of treatment and outcome in the analysis. It is most commonly used with two treatments and two outcomes (in a two by two contingency table. The Chi-square can be generalized to almost any number of treatments and outcomes.

FISHER'S EXACT TEST (PARAMETRIC)

This test is analogous to the Chi-square test but it can be used when there are fewer than 5 observations of each combination of treatments and outcomes. It is considerably more complicated than the Chi-square test, but fortunately most statistical packets will calculate it for the researcher.

ONE-WAY ANOVA (PARAMETRIC)

The one-way analysis of variance (One-Way ANOVA) is a test of significance used for three or more sets of continuous data drawn from samples with equal means, assuming the data are normally distributed and that the data from all groups have identical variances. The One-Way ANOVA can be thought of as a t test for three or more groups.

THE F TEST (PARAMETRIC)

The F test gets its initial from Sir Ronald A. Fisher who developed the test in the 1920s. It is really an analysis of the differences between the variance calculated for study groups. It is given by a simple process of;

$$F = \frac{S_1^2}{S_2^2}$$

The F test can only determine if there is a significant difference between the variance of groups. As such it is used instead of the student's t test only when this is the desired result.

NON-PARAMETRIC TESTS

MANN-WHITNEY U TEST OR WILCOXON RANK-SUM TEST (NON-PARAMETRIC)

The Mann–Whitney U test (also called the Wilcoxon rank-sum test or Wilcoxon–Mann–Whitney test) is a non-parametric test of significance for assessing whether two independent observations have equally large values. It is one of the best-known and most commonly used non-parametric significance tests. The Mann-Whitney U test is virtually identical to performing an ordinary parametric two-sample t test on the data after ranking over the combined samples.

KRUSKAL-WALLIS (NON-PARAMETRIC)

This is a non-parametric test analogous to the One-Way ANOVA test. However, no assumptions are made regarding the normality of distribution or the variance of the data. The data may be continuous or ordinal, as the test relies on the ranks of the data, not the values themselves. The Kruskal-Wallis test may be thought of as a Mann-Whitney U test for three or more groups.

MCNEMAR'S TEST (NON-PARAMETRIC)

McNemar's test is a non-parametric test that is used to compare two population proportions that are related or correlated to each other. McNemar's test is also used when one analyzes a study where subjects are accessed before and after the study intervention. McNemar's test is applied by a 2×2 contingency table with the dichotomous variable. In addition, if a researcher wants to determine whether or not a particular drug has an effect on a disease, then the individuals are listed in a table as + or – (alternately 0 or 1) before and after being given the drug. McNemar's test can be applied to make statistical decisions as to whether or not a drug has a statistically significant effect on the disease.

WILCOXON SIGNED-RANK TEST (NON-PARAMETRIC)

The Wilcoxon signed-rank test is a non-parametric statistic of significance for the case of two related samples or repeated measurements on a single sample. It can be used as an alternative to the paired student's t test when the population is not normally distributed. Like the paired t-test, the Wilcoxon signed-rank test involves comparisons of differences between measurements, so it requires that the data are measured in intervals. However it does not require normal distribution. It should therefore be used whenever the distributional assumptions that underlie the t-test cannot be satisfied.

ANALYSIS OF VARIANCE ANOVA (NON-PARAMETRIC)

Analysis of variance (ANOVA) is a collection of statistical models and their associated procedures, in which the observed variance of a set of samples is partitioned into components due to different sources of variation. In its simplest form ANOVA provides a statistical assessment of whether or not the means of several groups are all equal, and therefore performs the same task as the student' two sample t test but for more than 2 groups. ANOVAs are helpful because they possess a certain advantage over a two-sample t-test. Doing multiple two-sample t-tests would result in a substantially increased chance of committing a type I error. This is because every time one performs the t test on a sample set with an alpha of 0.05, the errors are additive. That means, as example if one performs the paired t test on 10 groups, the chance of committing a type I error is actually about 50%. For this reason, ANOVAs are useful in comparing three or more means.

KRUSKAL-WALLIS TEST (NON-PARAMETRIC)

The Kruskal-Wallis statistic is a non-parametric test that evaluates the median values of some study population parameter. The values of the data are actually replaced by their ranks and the variance (σ^2) of the group medians are then compared. In essence the test is like the one-way ANOVA with the data for the groups replaced by ranks. As such, it is also analogous to a Mann-Whitney U test when one has 3 or more groups to evaluate.

THE FRIEDMAN ONE-WAY ANOVA (NON-PARAMETRIC)

The Friedman analysis of variance by ranks is an alternative to one-way repeated measures ANOVA, because it does not require the dependent variable to follow a normal distribution. Using the Friedman (or any other non-parametric test), however, it is nearly impossible to achieve $p < .05$ with a sample size smaller than 12, because the tests are more exacting than their parametric counterparts. Because of this, one must have a reasonably large sample size to use this test.

SENSITIVITY AND SPECIFICITY

The idea of sensitivity and specificity may be best understood by reviewing their association with diagnostic test interpretation. In this setting the association of the sensitivity and specificity of a test is in calculation of the probability that a patient has a disease in the context of a specific test result (2, 3, 4).

A 2 by 2 table may be used to demonstrate this process.

Label the table with the test results on the left side and the disease status on top as shown here:

	Disease present	**Disease absent**
Test positive	*True positives*	*False positives*
Test negative	*False negative*	*True negatives*

Sensitivity is the proportion of patients *with* disease who test positive. In probability notation:

$$\mathbf{P(T^{+}|D^{+}) = TP / (TP+FN).}$$

This is read as "*The probability that a patient will test positive (given the condition that) the patient actually has the disease tested for – is equal to the true positives (TP) divided by the True Positives (TP) plus the False Negatives (FN).*"

Specificity is the proportion of patients *without* disease who test negative. In probability notation:

$$\mathbf{P(T^{-}|D^{-}) = TN / (TN + FP).}$$

This is read as "*The probability that a patient will test negative (given the condition that) the patient does not have the disease tested for – is equal to the true negatives (TN) divided by the true negatives (TN) plus the false positives (FP).*"

Several concepts and terms are needed here. ***Pretest Probability*** is the likelihood of the patient having the disease before the test is done. It is the same thing as **prior probability**. The pretest probability is simply estimated from what is known about the disease in the population you are studying. If a defined population of patients is being evaluated, the pretest probability is equal to the **prevalence** of disease in the population. The ***Prevalence*** of a disease is the total number of cases of a disease thought to be in a population at any one time. This is also an estimate as it is impossible to really know the number of true cases of a disease in a population. Estimations of incidence may be found in surveillance reports available through state departments of health or the Centers for Disease Prevention and Control (CDC). In addition, estimates of prevalence may be found in reports in the literature and textbooks. This is distinct from the ***Incidence*** of a disease which is the number of new cases diagnosed of a disease over a specific period of time (usually a year). Prevalence of a disease in your study

population may be determined from the 2 X 2 table above as the proportion of total patients who have the disease, in probability notation this is:

$$\mathbf{P(D^+) = (TP+FN) / (TP+FP+TN+FN).}$$

Sensitivity and specificity describe how well the test discriminates between patients with and without disease. Sensitivity and specificity address a different question than we want answered when evaluating a patient, however. What we usually want to know is: given a certain test result, what is the probability of disease? This is the **predictive value** of the test.

POSITIVE AND NEGATIVE PREDICTIVE VALUES

Predictive value of a positive test is the proportion of patients with *positive* tests who have disease. In probability notation: $\mathbf{(D^+|T^+) = TP / (TP+FP)}$. This is the same thing as ***posttest probability*** of disease given a positive test. It measures how well the test rules in disease.

Predictive value of a negative test is the proportion of patients with *negative* tests who *do not* have disease. In probability notation: $\mathbf{(D^-|T^-) = TN / (TN+FN)}$. It measures how well the test rules out disease.

SUMMARY

The applications of most statistical tests to data are dependent on their distribution around the central tendency (means) of the data sets. A symmetric "bell curve" distribution or "Normal" distribution indicates that a parametric test may be used. Asymmetric data distributions require a non-parametric test. There are a number of "short-cut" methods that can occasionally allow nearly normal data distributions to be considered normal – gretly simplifying the mathematical management of such data sets. Calculations of sensitivity and specificity as well as positive and negative predictive values are vital in the analysis of many diagnostic and assessment processes in clinical medicine.

REFERENCES

1. CDC. EPIINFO HTTP://www.cdc.gov/epiinfo/ (accessed 10/14/2010)
2. Le CT. 2001. Health & Numbers: Problems based introduction to biostatistics 2nd Ed. Wiley-Liss New York, NY. Pp. 56-79.
3. Aschengrau A, Seage GR III. Essentials of Epidemiology in Public Health (2nd Edition) Jones & Bartlett Sudbury, MA 2005.
4. Jekel FJ, Katz DL, Elmore JG (2001) *Epidemiology, Biostatistics, and Preventive Medicine* (2nd edition) W.B. Saunders Co. Philadelphia, PA.

Confidence Intervals

INTRODUCTION

Recall from chapter 10 the brief discussion of standard deviation (σ) which is a measure of the variability of individual measurements of a data set. The standard error (SE), is a related statistic but rather than demonstrating the variability of individual measurements, it is a measure of the variability of the means of several data sets. As such the SE provides an estimation of how variable a single estimation of the mean value from one set of research data is likely to be. The general formula for SE is as follows;

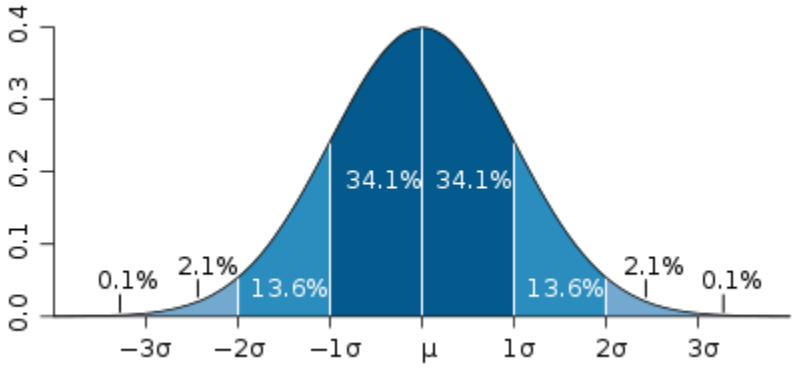
$$SE = \frac{\sigma}{\sqrt{N}}$$

N = The sample size of the mean

Obviously the smaller the SE, the better the estimate of the mean is.

CONFIDENCE INTERVALS AND DATA DISTRUBUTION

With the above discussion as background, the question is often asked of the researcher, how much confidence does one have that the mean measured is accurate? An estimation of such accuracy can be made through the calculation of 95% confidence intervals. Recall the bell curve distribution of data, When data are distributed on a frequency graph as demonstrated, about half of the measurements fall above the mean (μ) and half below. If we measure out 1 standard deviation above and below the μ, we find that 68.2% (34.1% above and 34.1% below) of the measurements fall within this area of the curve. Going out two standard deviations (2σ) we find that an additional 27.2% (13.6% above plus 13.6% below μ) of measurements are found for a total of 68.2% + 27.2% = 95.4% of the measurements at plus or minus two standard deviations from the mean. To make this number easier to deal with, drop the 0.4% (by going out only 1.96 standard deviations instead of 2 above and below the mean) and one now has the range within which 95% of the measurements will fall on either side of the mean.



This range having been identified (at 1.96 standard deviations either side of the mean) one may now substitute the standard error (SE) for the σ. This provides the range within which 95% of the measured values of the means of the data sets will fall. If the value for the mean, plus or minus 1.96 standard errors is known – the one can calculate the *95% Confidence Interval* (95%CI).

THE 95% CONFIDENCE INTERVAL

The 95% Confidence Interval (95%CI) is defined as the range of mean values in which the researcher can be 95% confident that the true mean lies – (in reality the 95%CI means that if we repeat the measurement of the mean, we are 95% confident that the new calculated mean will fall within the 95%CI range, but the subtle difference in meaning adds little to the basic understanding of the concept for most non-statisticians) (1).

EXAMPLE CALCULATION OF A 95% CONFIDENCE INTERVAL

Suppose one carries out a study measuring the systolic blood pressures of 26 adults. Evaluating this data, we determine the following;

Number of Observations (N) = 26

Mean Systolic Blood Pressure (μSBP) = 113.1mmHg

Standard Deviation (σ) of μSBP = 10.3mmHg

First Calculate the Standard Error (SE);

$$SE = \frac{\sigma}{\sqrt{N}} = \frac{10.3}{\sqrt{26}} = \frac{10.3}{5.1} = 2.02mmHg.$$

Next Calculate the 95% Confidence Interval;

$$95\%CI = \mu \mp (1.96)(SE)$$

$$95\%CI = \mu SBP \mp (1.96)(SE)$$

$$95\%CI = 113.1 \mp (1.96)(2.02)$$

$$95\%CI = 113.1 \mp 3.96$$

$$95\%CI = Between\ 113.1 - 3.96\ and\ 113.1 + 3.96$$

$$95\%CI = 109.1, 117.1mmHg.$$

INTERPRETATION OF 95% CONFIDENCE INTERVALS

The more narrow the confidence interval, the better. A wide range means that although we are 95% sure the value is in that range, the range may be too wide to be of much use (2).

In addition, interpretation of a CI depends on what parameter has been measured. As example, if one has calculated a risk ratio or odds ratio, we know that a value of "1" indicates no difference in odds or risk between two variables. Therefore, if a calculated risk ratio value of 1.7, has a 95% CI of between 0.92 and 2.70 – it would not be significantly different from a value of 1 – because the confidence interval contains the value of 1. However, if that same risk ratio of 1.7 had a CI of between 1.02 and 2.06, we would consider the ratio significant because the value of 1 is not in the CI range. Note that this line of thought is important for risk ratios, odds ratios and likelihood ratios, but other calculated statistics such as means will (and actually must) be contained within the confidence interval range.

SUMMARY

Confidence intervals help to provide a deeper insight into the "significance" of data analysis beyond simple "p" values by providing a range of confidence within which a calculated statistic such as means or ratios should fall. As such, they should almost always be included in the reporting of such data analysis.

REFERENCES

1. Jekel FJ, Katz DL, Elmore JG (2001) *Epidemiology, Biostatistics, and Preventive Medicine* (2nd edition) W.B. Saunders Co. Philadelphia, PA
2. Normal GR, Streiner DL. *Biostatistics: The Bare Essentials* 2nd edition BC Decker Inc. 2000 Hamilton Ontario, Canada.

Statistical Associations between Variables

INTRODUCTION

Evaluation of two or more variables for cause and effect is a critical part of statistical analysis. What researchers often want to know is if some variable in a process is changed, does it impact or change some other variable? In the simplest terms, one picks some outcome of interest to measure (this is called the outcome, dependent or response variable). Next one selects some variable that is known (or at least suspected) to have an impact on that outcome variable (this is known as the independent variable). Next, one sets up a condition that allows one to change the independent variable to some degree and measure the outcome variable for any change. When properly designed and carried out, this process allows one to determine if there is a relationship or association between a variable and the selected outcome, and to some degree what the nature of that relationship is. Note that this does not necessarily indicate that the variable is the "cause" of the outcome, just that it is associated with the outcome in some way.

EXAMPLE OF AN ANALYSIS OF ASSOCIATION

A research protocol was done to evaluate Community Acquired Methicillin Resistant *Staph aureus* in abscesses presenting to an Emergency Department. Among the data collected were ages of the cases, and the time in days it took for the abscess to heal after incision and drainage in the ED. One question that was asked is whether or not there is an association between the age of the patient with the abscess and the time in days it took to heal. The basic statistical test results performed on the data are noted below in table 1.

n = 263

X Age of Cases in Years	Y Days to Heal
Mean = 26.94 years	Mean = 9.20 Days
Variance = 220.06	Variance = 36.49
STDEV = 14.86	STDEV = 14.86
SE(x) = 0.92	SE(y) = 0.37

Table 1. Data summary for MRSA Study

Next, one may use the data to create a scatter plot with Age (x) over Days to Heal (y) as in figure 1. Many computerized statistical packages will generate such scatter plots as part of the routine analysis process of data.

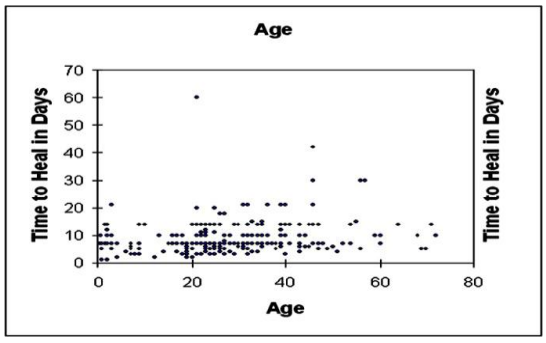


Figure 1: Scatter plot from MRSA Study, Time-to-Heal/Age

The scatter plot strongly suggests that there is no correlation between the dependent variable (age x) and the independent variable (days to heal y). One could try taking the log of the days to heal but it is unlikely to this change much the view of the relationship much. Next the Pearson's Correlation Coefficient statistic is performed on the same data. The closer the result (termed "r" in statistical parlance) is to zero, the less likely there is to be a correlation between the variable (age) and the outcome (time to heal). This provides an r = 0.176. Because this is close to zero, it is unlikely that there is a correlation between the two variables (at least not a linear correlation).

Finally one determines the Coefficient of Determination which is defined simply as the square of the Pearson's Correlation Coefficient (r). The result of the Pearson's Correlation Coefficient test is;

r^2 = 0.031 or 3.1%

The Coefficient of Determination serves as an index of the impact of the independent variable on the dependent variable or outcome measured. In this case, there is very little impact if any. At best, the age (x) accounts for 3.1% of the variance in the time to heal (y). So in this case, the analysis supports NO association between age and time of an abscess to heal. At this point, no further analysis is warranted as it is unlikely that these two variables are associated.

WHEN A STATISTICAL ASSOCIATION – IS NOT

It is impossible to understand all relationships associated with a specific dependent variable in research. However, it is important that as many potential relationships as possible be looked at from the point of view of biological plausibility as well as statistical inference. What this means is do not be "blinded" by the appearance of a mathematical-statistical result, when in fact the result does not make sense, or there is a much stronger logical explanation for the apparent association. The following is an example of an apparent relationship that puzzled researchers for a number of years.

EXAMPLE OF A FALSE RELATIONSHIP BETWEEN VARIABLES

The majority of studies of the relationship between the number of prenatal visits and the health of the child at birth have consistently demonstrated a strong positive correlation. That is, the larger the number of prenatal visits before birth, the healthier (by any outcome measure) the child is. This correlation seems to hold true even when other factors such as socioeconomic status, race, and education are controlled for. The following (table 2) is a set of data that demonstrates the number of prenatal visits, birth weight of the child (under 2,500 grams is considered low birth weight), and the weeks of gestation at delivery for 12 women.

Num Prenatal Visits	Birth Weight in Grams	Weeks of Gestation
13	3895	40
13	3555	40
12	3250	40
12	2900	40
11	2850	39
10	2458	39
10	2588	38
9	2425	38
8	2410	36
8	2025	36
8	2155	36
8	2000	36

Table 2. Prenatal visits, birth weight, and gestation data on 12 women.

First one creates a scatter plot as in the MRSA example to get a feel for the data and possible relationships between variables. The number of prenatal visits is the variable of interest so it will be made the dependent variable with the other factors serving as independent variables. The plot is of birth weight over prenatal visits, and it seems to demonstrate a strong positive association with a tightly grouped series of data points about the normal line. The up slope to the right indicates a "positive" association suggesting that the more prenatal visits, the better (higher) the birth weight of the infant.

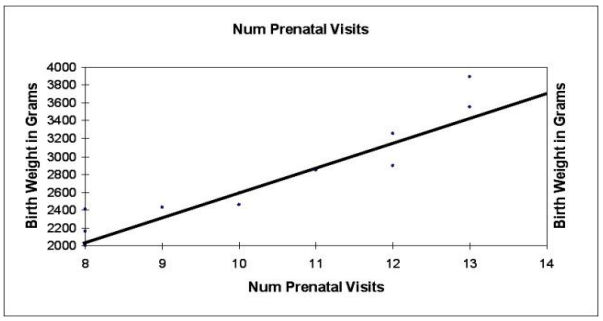


Figure 2: Prenatal Visits and Birth Weight

A Pearson's Correlation Coefficient is performed for this pair with an r = .944 and a Coefficient of Determination r^2 = .892. Both of these statistics support a strong positive correlation between the number of prenatal visits and higher birth weights. Despite the math, the logic error should become obvious before going any further.

One can go even further than in the MRSA example and do the next commonly done test by placing this data into the test for independence under the simple linear regression model. If this is done one will get further results consistent with a high degree of correlation between prenatal visits and birth weight. The problem is that the two variables are not actually directly connected. There are other issues (confounders) that are actually causing this apparent relationship.

Prenatal visits early in pregnancy generally have a frequency of about once monthly if there are no complications or complaints. As pregnancy progresses and the date of delivery comes near, the frequency of prenatal visits increase. The frequency may rise from monthly to bimonthly and finally to weekly during the last month.

Children that are born preterm – say at 36 weeks – have on average less robust outcome indicators such as birth weight, APGAR scores, as well as higher duration of hospital stay, and frequency of complications. In addition, when pregnancy terminates early the mother will have – on average – fewer prenatal visits than a pregnancy that continues to term (4 to 5 fewer visits in this case due to the increased frequency of prenatal visits during the last month). That is, if assessment is made of a relationship directly between low birth weight and the number of prenatal visits, there will likely be a strong positive correlation noted. However, no direct association actually exists between the two variables. The issue that modifies both of these variables is actually the cause(s) of preterm delivery. This is demonstrated in Figure 3.

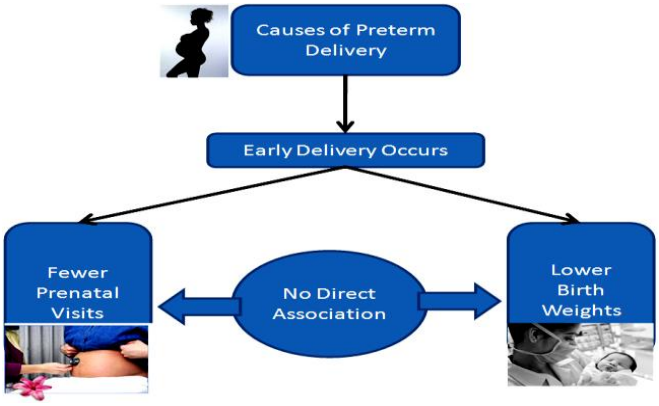


Figure 3: Lack of Link Between Prenatal Visits and Birth Weight

So in this case, the variables of number of prenatal visits and birth weight are indeed related to one another, but they have no direct effect on each other. They are related in that there are a number of factors (confounders) that were not known or considered in the analysis that actually cause preterm delivery. Once preterm delivery happens, the number of prenatal visits is cut short, and the infants are born with lower birth weights – but these outcomes are NOT actually influenced by each other.

The key here is that it makes no logical sense that more prenatal visits should so strongly influence birth weight. Any time one is presented with any apparent association between a variable and an outcome, the researcher must take the time to determine if the association is actually logical in the context of the clinical process. If not, confounders and other associations must be sought.

SUMMARY

Statistical tests of association are intended to determine if there is a statistical correlation between the value of one variable as another variable changes. However, the potential association between such variables must be

considered in the logical and "real world" context of the process being investigated. This includes evaluation for other factors (confounders in this case) that can be involved in the relationship.

ADDITIONAL READING

- Motulsky H. *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking* Oxford University Press, 2010. New York, NY.

- Le CT. (2001). *Health & Numbers: A Problems-Based Introduction to Biostatistics* (2nd ed.). New York: Wiley-Liss.

A Theory of Justice 5, 30, 41
ABSTRACT **4**, **16**
Abstracting Charts 65
Alpha, Beta, and Sample Size 93
Alternate Hypothesis 16, 46, 85, 86, 87
American Research Abuse 33
Analysis of variance (ANOVA) 101
Belmont Report 36
Beneficence 36
Bias 4, 5, 8, 9, 23, 24, 25, 26, 28, 61, 70
Case Definition 45
case-control study 69
Central Limit Theorem 86
Central Tendency 90
Chart review 63
Chi-Sauare test 100
Clinical Equipoise 36
Cohort studies 66
Confidence Interval 13, 105, 106
Confounding 4, 5, 9, 23, 26, 27, 28, 70
Consequentialist Ethical Theory 5, 30
Continuous data 4, 20
Critical-Rationalism 85
Crossover Study 76
Darth Vader 31
DATA **7**, **46**
Declaration of Helsinki 35
Deontology 5, 30
Department of Health and Human Services 34, 42
Deviations from the Mean 91
DHHS 34
dichotomous data 21
Discrete data 4, 20
Ecological Fallacy 26
Emergency Exemption Consent Waver 39
Emergency Medicine Research Consent 39
Empiricism 84
ethics 5, 29, 30, 31, 37, 42, 74
Exclusion Criteria 10, 45, 81
Experimental Approach 55
F.I.N.E.R 43
FDA 34, 76
Federal Food and Drug Administration 34
Federal Regulation *45 CFR 46 subparts B-E* 81
Fisher's Exact Test 100
Flow of Information Problem 55
Formal Literature Search 49
Friedman analysis of variance 102
generalizable **20**, **77**
Guttman 8, 58, 59
Hawthorne Effect 25
Healthy Worker Effect 24
Henderson Act 33
Historical Approach 54
IACUC 36
Incidence 103
Inclusion Criteria 45, 81
Independent Samples 95
Index Medicus 50
index of accomplishment 17
Informed Consent 37
Institutional Animal Care and Use Committee 36
Institutional Review Board 9, 36, 63, 65
Interquartile Range 91

interval data 21
Interviews 56
IRB 6, 36, 37, 39, 40, 42, 44, 63, 65, 74, 79, 81, 82, 83, 93
Japanese Research Abuse 32
Jeremy Bentham 30
John Mill 30
John Rawls 30
Justice 36
Kant 30
Karl Popper 85
Kruskal-Wallis 102
Likert 8, 58, 59
Mann-Whitney U Test 101
Mcnemar's Test 101
Medical Librarian 53
Medical Subject Headings 50
MeSH 50, 51, 53
Meta-Analysis 71
METHODS **4**, **10**, **17**, **80**
National Center for Biotechnology Information 50
National Library of Medicine 50
National Research Act of 1974 35
Nazi Research Abuse 32
NCBI 50, 51
NLM 48, 50, 51
nominal data 21
Non-parametric 12, 86, 99
normal frequency distribution 88
Normal, or *Gaussian* distribution 86
Null Hypothesis 16, 46, 85
Numerical Descriptor 8, 58, 60
Nuremburg Code 35
Obi-Wan Kenobi 31
One-Way ANOVA 100
Opportunistic Data Torturing 26
ordinal data 21
Outcome Measure 46
Outcome Measures 16
p value 96, 97, 98
Paired t test 100
parameter **20**, **86**, **87**, **89**, **90**, **91**, **97**, **102**, **106**
Parametric 12, 86, 99, 100, 101, 102
Percentiles 91
PHI 65, 80, 82
Pilot testing 62
posttest probability 103
Predictive value of a negative test 103
Predictive value of a positive test 103
Pretest Probability 103
Prevalence 103
prior probability 103
Private Health Information 65, 80
Procrustean Data Torturing 26
PubMed 7, 48, 50, 51, 53
Questionnaires 56
randomization **20**, **28**, **66**, **73**, **74**, **77**, **81**
Randomized Clinical Trial 73
Ratio data 21
Rationalism 84
Reasonable Patient Standard 38
Reasonable Physician Standard 38
Research Protocol **9**, **10**, **15**, **64**, **79**
Research question **16**
Research Question 42
Research Regulation 34
Respect for persons 36

sample population **17**, **20**, **22**, **95**, **96**
Sample Size 92
scatter plot 108
Semantic Differential 8, 58, 59
sensitivity and specificity 102
Standard Deviation 92
statistic **20**, **89**, **96**, **97**, **98**, **100**, **101**, **102**, **105**, **107**, **108**
Statistics **11**, **16**, **86**, **89**, **95**
Student t Test 100
STUDY POPULATION **7**, **45**
Subjective Standard 38
Survey Monkey 62
syphilis 33
The Difference Principle 31
The F Test 100
The Liberty Principle 31
The Mean 90
The Median 90
The Mode 91
The Range 91
The Survey 54
Third Reich 32
Tuskegee 33, 35, 41
Type I and Type II Errors 89
United States Atomic Energy Commission 33
United States Public Health Service 33
Utilitarianism 5, 30, 41
Variables **13**, **16**, **28**, **54**, **71**, **108**
Variance 92
Visual Analog 8, 58, 60
Volunteer Effect 25
Wilcoxon rank-sum test 101
Wilcoxon signed-rank test 101