Osteopathic Research Center
Fostering nationwide collaborative OMM research

MANUAL OF BASIC TOOLS

FOR RESEARCH IN OSTEOPATHIC MANIPULATIVE MEDICINE

Second Edition
April 3, 2007

The University of North Texas Health Science Center
Texas College of Osteopathic Medicine

MANUAL OF BASIC TOOLS FOR RESEARCH IN OSTEOPATHIC MANIPULATIVE MEDICINE

TABLE OF CONTENTS

FORWARD

This 2nd edition of the Manual of Basic Tools for Research in Osteopathic Manipulative Medicine is a product of the national Osteopathic Research Center. Original contributions and source materials were used that are cited below or in the text. It is offered as a guide for students, residents and faculty conducting research in Osteopathic Manipulative Medicine. The manual has been produced in collaboration with many research education partners. Without their dedication it would not have been possible.

The primary author and editor of this research manual is des Anges Cruser, Ph.D., MPA. Dr Cruser used the pronouns she, he, her and him independently throughout the document so as to avoid using s/he, or she or he, or him or her repeatedly. Either pronoun applies to both genders as used. Dr. Cruser welcomes all comments and input to this manual to ensure it is useful for students, resident physicians and new researchers. She can be reached at dcruser@hsc.unt.edu

This manual was produced with the most generous editing assistance of Krista Gordon, MPH, Cathy Kearns, BA., and Jackie Williams, M.S.

This Manual is a guide. You will want to use other sources to gain additional information and guidance for your specific project. On this page is a list of source materials used.

Source documents

The Louisa Burns Osteopathic Research Committee Research Manual

The AOA Research Manual

Manuals created by Colleges of Osteopathic Medicine for use by their students including:

- A.T. Still University, Kirksville College of Osteopathic Medicine
- Western University College of Osteopathic Medicine of the Pacific
- The national Osteopathic Research Center at the University of North Texas Health Science Center, Texas College of Osteopathic Medicine and Graduate School of Biomedical Sciences

On the following page is a list of individuals you can contact who are active in OMM research.

Key Contacts and e-mail addresses

Michael A. Seffinger, D.O.
mseffinger@westernu.edu

Michael L. Smith, Ph.D.
msmith@hsc.unt.edu

Michael B. Kuchera, D.O.
michaelkuc@pcom.edu

Michael Patterson, Ph.D.
email

Scott T. Stoll, D.O., Ph.D.
sstoll@hsc.unt.edu

Jane Johnson, M.S.
JJohnson@atsu.edu

Hollis King, D.O., Ph.D.
hking@hsc.unt.edu

Brian Degenhardt, D.O.

John C. Licciardone, D.O., MBA, MS
licciar@hsc.unt.edu

Karen Steele, D.O.

Kendi Hensel, D.O. (Ph.D. student)
khensel@hsc.unt.edu

Daisha J. Cipher, Ph.D.
dcipher@hsc.unt.edu

Marty Knott, Ph.D. (D.O. student)
eknott@hsc.unt.edu

Heath D. White, D.O., M.S.
hdwhite@swmail.sw.org

David Russo, D.O., M.S., MPH
russoda@ohsu.edu

Donald Noll, D.O.
email

William C. Meeker, D.C., MPH
Email

Paul Standley, Ph.D.

David Yens, Ph.D.
email

other EBSC members listed as such separately

This page intentionally not numbered

CHECKLIST FOR DETERMINING THE APPROPRIATE RESEARCH PROJECT FOR ME

The following questions will guide your thinking about the feasibility of a research project. These are not the only questions you should ask. These questions are a beginning.

A. WHAT KIND OF RESEARCH SHOULD I DO?

- □ Examine existing data for trends; explore differences in populations; analyze and describe in a correlational study the relationship between presenting problems, diagnostic findings, and treatment interventions?
- □ A clinical trial? (use a treatment group only or a controlled trial?)
- □ A case control study?
- □ A review of the literature in a specific disease process or treatment?
- □ An arm of an existing research project?
- □ Replicating another ongoing study or a portion of an ongoing study?
- □ Basic science or clinical science?
- □ Other? _____

B. TIME AND RESOURCES?

- □ Do I have easy access to data or patients?
- □ Who are the available experts or mentors with the time and experience to guide and support my research?
- □ What will I need and expect from that person or persons?
- □ How long do I have or want to take to conduct this research?
- □ Other: _____

C. DATA ACCESS OPTIONS

- □ Am I generating pilot data?
- □ Are pilot data available that can guide my research design development?
- □ Is there pilot data that can be analyzed in a different way?
- □ Is there existing data in an electronic form that I can analyze such as a national survey or other national or local data base?
- □ Other _____

D. END POINTS AND PRODUCTS

- □ Will I publish or present the findings?
- □ What is the benefit or value to osteopathic medicine or health care consumers?
- □ Is there a cost-benefit issue to address for this research project?
- □ What do I expect of myself during and after the project?
- □ Other _____

E. OTHER RESOURCES

- □ Do I have a biostatistician available to consult on data management and analysis?
- □ Am I trained in Human Subjects protection issues, IRB procedures, and/or have access to a clinical research coordinator to assist in data collection, and study reports/reviews?
- □ Other _____

F. MY OWN SKILLS

- □ Do I want to do this?
- □ Do I have or can I acquire the basic skill and knowledge to accomplish my goals for this research project?
- □ Am I committed to doing this research project?
- □ Other _____

The next step is the design phase and assistance with this can be found in the Worksheets at the end of Chapter I and in other source materials.

Note: The most often cited reason for not completing a study is lack of an adequate number of subjects. Rarely do published reports state that all subjects enrolled completed the trial. Furthermore some studies do not achieve their enrollment goals. Be very careful about this. It is a significant investment of time and other resources, and a potential source of frustration.

This page intentionally left blank and not numbered

Chapter I
THE RESEARCH PROPOSAL

SECTION 1
INTRODUCTION

Chapter I, "The Research Proposal" includes one section for each part of a traditional research proposal. It is based on the NIH-Gold Standard for research proposals of all kinds.

No matter what type of research you chose to conduct you will need a research proposal. This is the document used to communicate your plan for a research project to all relevant audiences.

The main sections always include at least the Specific Aims, the Background and Significance, (your own, if any, Preliminary Studies), and Methods.

For research using any information from or about human subjects, Chapter II provides guidance on standard human subjects' protection issues.

Multiple sources of guidance on the content and format of, and the process of developing each section of a research proposal for small and large studies are available in textbooks, on the world-wide-web, and at your institution. As a student you should learn to seek out these sources and use them to their maximum benefit. Research project development is not a solitary process. Although writing a research proposal is ultimately the responsibility of one person, it always benefits from team, consultant, mentor, and reviewer input to shape a well crafted proposal.

This principle applies also to the logistics of the research. Great research ideas and proposals must be feasible. Research must be "doable" in the real world regardless of the topic or the research design. Every stage and step in any research project, no matter how small or how large, must be carefully developed, considered, and communicated within the proposal so that it can be translated into an operational plan that will ultimately serve the goal of the project from beginning to end.

Objectives of a Research Proposal

A research proposal has two primary objectives.

1. To define and describe the research project
2. To establish the plan for the conduct of the project

Research with humans cannot be conducted without IRB approval, exempt, expedited, or full board review. Your research proposal will either be attached to your IRB forms or proposal, or it will be used to copy and paste into relevant sections of an IRB proposal. This depends on what your institutional guidelines require.

Organization of this Chapter

This chapter is organized around the major components used by the National Institutes of Health (NIH) for research proposals.

For more information, refer to the instructions for PHS 398 forms at the following url.
http://grants1.nih.gov/grants/funding/phs398/phs398.html
You may also look at a specific program announcement by any potential sponsor including your own school. This chapter does not cover all possible information you may want to develop a research proposal. It includes the essentials. You may have unique needs. Your university or OPTI may have other guidelines. Follow the guidelines most appropriate for your research.

This chapter provides

- A description of the purpose of each section of a research proposal,
- Prompts to guide the writer and
- Examples of strong and weak material.

Full proposal samples are available from the ORC with the permission of the Principal Investigator.

SECTION 2
GOOD HABITS AND TIPS

You should be able to create a research proposal in no more than about five or six pages. A research **grant** proposal would be longer.

Appendices are discouraged for several reasons, one of which is that if a researcher cannot say in the allotted pages what she wishes to communicate to the audience, revisions are needed. Always ask about unnecessary or necessary appendices, especially if the instructions are unclear or the appendix is required or requested by the sponsor.

All data collection forms (called DCFs – this acronym will be used again in this manual) or questionnaires you plan to use in the study must be thoroughly described, cited, justified, and when permitted or required, the DCFs will be an appendix to the Research Plan. If a data collection form (DCF) is described in the body of the proposal the form itself may need to be in an Appendix. For an IRB proposal you will normally include all DCFs. This is the way the IRB will determine the level of review required and will see what information you plan to collect from records or from individual subjects.

Some tips for conserving space in your proposal include such strategies as:

- Use a single space rather than double spaces between a period and the beginning of a sentence (yes this IS permissible),
- Use tables instead of vertical lists of items,
- Limit modifiers, and
- Choose words and sentence structures carefully.

Parsimony is important in scientific writing. Parsimonious means to be unusually or excessively frugal; and is herein applied to the use of words. Attempt to utilize the simplest assumption in the formulation of a theory or in the interpretation of data, especially in accordance with the rule of Occam's razor. Occam's razor is also called the *principle of parsimony*. These days it is usually interpreted to mean something like "the simpler the explanation, the better" or "don't multiply hypotheses unnecessarily." Select cautiously from among theories with equal explanatory power, and when giving explanatory reasons for something, avoid positing more than is necessary.

Good Habits For Scientific Writing

- Read other scientific material or proposals with a critical eye.
- Use active voice vs. passive.
- Use “this study” or “this proposed study” in reference to your proposal, and “that study” or “their study” in reference to another you are citing.
- Do <u>not</u> use the term “etc.” Be precise. If there is too much to list in a sentence use a table or refer to a list elsewhere. On occasion you can say “such as” or “including but not limited to”. Do not make the reader guess what you mean or have the impression that you are vague.
- Use the terms “valid” and “reliable” wisely. There are many different types of validity and reliability. Refer to a glossary of research terms. Ensure you are correctly using these and other research terms.
- A finding is either significant or not significant. Some authors erroneously use modifiers such as “highly” or “greatly”. This is not sound research language.
- Always use correct grammar. Check and have others check your spelling. Do not rely solely on the computer for spelling or grammar checks.
- “Account for” means that a variable or a factor explains a portion of the effects of the independent variables on the dependent variable in a correlation analysis. For example, “age” may account for a portion of the difference between the time men and women take to fall asleep in a study of sleep aids.
- You may indeed make good use of the first person such as I or we.
- Keep it simple and logical.
- Get at least two detail-oriented individuals to read the proposal for you.

SECTION 3
THE RESEARCH PROPOSAL AND THE INSTITUTIONAL REVIEW BOARD (IRB) PROPOSAL

There is a difference between the Research Proposal and an IRB proposal. The IRB functions to protect the subjects in a study, and to protect the organization (the university, the hospital, and such) by ensuring ethical and legal practices in all research conducted by its faculty. The research proposal usually accompanies the IRB proposal that focuses on inclusion/exclusion of protected groups of persons, risks, and protections for subjects in a study. An IRB outline and sample clinical trial proposal is included in Chapter II.

The Institutional Review Board (IRB) Proposal

The Osteopathic Research Center (ORC) has a designated liaison to the local IRBs and communicates with IRBs at other institutions. As a researcher you will be expected to work with someone at your institution or you may contact the ORC to help to develop and review your IRB proposal BEFORE you go to the IRB staff.

One reason for this is to have the luxury of an informal pre-review performed so that you experience as few delays as possible in satisfying the IRB requirements. The typical IRB WANTS research to proceed, and has the responsibility to protect the institution and the research subjects.

SECTION 4.A.
THE OUTLINE

The main components of a research proposal are listed below. These are major section headings that reflect the gold standard endorsed by the NIH and most published texts on writing research proposals. These sections are to be followed in the order presented. Although each major section may vary in the order in which sub-sections are presented, there are essential, commonly accepted elements that should be included. If a sponsor or institution has no requirements, you cannot go wrong with this formula. It is endorsed by the finest texts in clinical and social science research in the world.

RESEARCH PLAN STANDARD OUTLINE BASED ON NIH FORMS AND FORMAT

A. Specific Aims
B. Background And Significance
C. Preliminary Studies Or Research (Only Your Own)
D. Research Design And Methods
E. Human Subjects (If You Plan Animal Research There Is A Separate Section For That)
F. Literature Cited
Appendices

Each of the above components is restated below with:

- A description of the component
- Key questions to guide your thinking
- Prompts to guide your writing
- Examples of strong and weak material

(NOTE: some examples are paraphrased to give a brief sense of the tone of the component and for brevity. Complete examples are available upon request.)

SECTION 4.B.
THE ANNOTATED OUTLINE

Margins and Fonts

You may use any font that is easily readable. The preferred fonts are Helvetica 12 or Arial 11. You may use either 1" margins all around, or the NIH guidelines. NIH margins are ½" all around and font is specifically recommended in the instructions. CHECK YOUR MARGINS!

A good font to use for the NIH proposal is Arial 11. You do not need to keep that size in tables or figures, but nothing is to be smaller than 10 point font. Also, using a consistent font type/style is helpful to the reader. You may bold or italicize for emphasis or separation of key points.

This is Arial 11. This is Arial 12. This is Times Roman 12. This is Bookman Old Style 12.

The Research Plan

A. Specific Aims

Use this section to establish and briefly describe the rationale for the study. For the Master's Thesis Research Proposal, keep it short – about two paragraphs. For NIH and other research funding proposals the standard is one page.

This section is where you present in a logical, organized fashion the thought process that leads the reader to the same conclusion that you make in your statement of specific aims. This must "grab" the reader, make her interested in reading further, and be the most parsimonious statement of your intent that you can possibly make.

Always present your specific aim/s and question/s or hypotheses on the first page. You will probably begin with a longer version of this section and then edit and reshape it to fit.

This section must always contain:

1. First a statement of the broad, long range goal of the proposal
2. A few statements that justify that goal and describe the focus or topic of the research
3. One or two primary or the primary and secondary aims of the research project that will lead directly into the primary research question or hypotheses for the proposed study

The hypotheses may be linked specifically with each aim and you can restate each aim with its associated hypotheses in the Methods section.

Above all: KEEP THIS SECTION SIMPLE, STRAIGHT FORWARD AND LOGICAL. This is your chance to take the reader by the hand and lead him from nothing to the end of the page "Ah-Ha". The reader should be able to determine the likelihood of approving your proposal this early in the reading. All else will be measured and linked back to this page.

Lastly, you will usually find some "redundancy" in the proposal. This means you will repeat what is stated here later in the paper.

Prompts:

- Tell the reader what you are going to do.
- Tell the reader why.
- Tell the reader how you will do this.
- Tell the reader what you told them.

Key Question: What is the focus of the study?

Example

(strong) *The overall goal for this proposed exploratory/developmental clinical trial is to evaluate the efficacy of a conservative, biomechanical, non-surgical treatment for carpal tunnel syndrome (CTS).*

(weak) *This proposal seeks to describe how OMT might improve carpal tunnel syndrome*

Example

(strong) *Osteopathic manipulative medicine (OMM) is based on scientific principles of the relationships among the musculoskeletal, fluid, and neurologic structures and functions of the human body. OMT directs itself to the systemic processes associated with disease and dysfunctional conditions to promote the body's self-regulatory and self-healing abilities. Despite the plausible good fit between OMM principles and the clinical and functional characteristics of CTS, there are few known studies of OMT and CTS.*

(weak) *The focus of this study is how manual medicine techniques influence the person's ability to strengthen fluid and other body structures in a self-healing process. No one has studied this, or published any findings about it.*

Example

(strong) *Diabetes mellitus is the fourth leading cause of death and disability in the United States. The diabetes epidemic imposes a tremendous public health burden on our population and economy.[21] The combination of repeated insulin administration and a strict dietary regimen is the only available strategy to treat diabetes, and even these measures only delay rather than prevent progression of the disease. The paucity of effective treatments for diabetes prompted the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) to issue a program announcement (PA-01-112) seeking applications to study the use of antioxidants to mitigate diabetic complications. The purpose of this proposal is to generate foundational data to support applications to the National Institutes of Health in response to PA-01-112.*

(weak) *Many people have diabetes in the US. It costs a lot of money to treat them. I want to collect data to help others study new treatments for this disease.*

Thinking Exercise: How can you improve this example? Does it say what you think the person really means? Are there any errors in thinking in this example?

There are still many unanswered questions in the research of OMT and the elderly. In particular, there is a void of data on the relationship between somatic dysfunction and chronic conditions. This study is intended to help narrow the gap of knowledge and improve the osteopathic physician's ability to diagnose and treat the older individual. By understanding the characteristics of older patients that seek OMT and by learning more about the common chronic conditions and areas of somatic dysfunction that are associated with this group of individuals, the physician may come to better anticipate, diagnose and treat the structural and visceral dysfunctions in the elderly population.

Key Question: Why is this area the focus?

Example

(strong) *CTS results from compression of the Median nerve within the carpal tunnel of the wrist. Confirming a diagnosis of CTS involves both a clinical exam and electrodiagnostic data. CTS may cause parathesias (e.g. tingling, burning) in the area of the hand and fingers and may have associated symptoms in the arm and shoulder. CTS affects at least 10% of the U.S. adult population, and costs over $1 billion annually in medical care alone, but the true prevalence is unknown (Papanicolaou, 2001). A variety of treatments are available for CTS. Conservative management can cost less than surgery (Osterman, 2002). Also, surgical risks may include injury to nerves, excessive scaring, and loss of strength or sensation. Up to one-third of all post-surgical patients report recurring pain and range of motion problems (Rosenbau, 2000). For these and other reasons, clinical researchers continue to search for viable, non-surgical alternative treatments for CTS.*

(weak) *Up to 10% of the people in the US have CTS. We don't know how many people actually have the condition. Surgery puts patients at risk but alternatives might be better. We all continue to search for better ways to treat CTS. This proposal should contribute to this knowledge.*

Key Question: What therefore will I do?

Example

(strong) *The primary aim is to determine if OMT with and without stretching exercise will improve electrophysiologic conduction of the Median nerve and decrease symptoms of CTS.*

Secondary aim one is to evaluate the effects of OMT with and without stretching exercise on grip strength and hand sensation in persons with CTS.

Secondary aim two is to examine the extent to which OMT with and without stretching exercise can increase the dimensions of the carpal tunnel and decrease the water content (edema) of the carpal tunnel.

(weak) *Therefore this study will collect data about how OMT impacts CTS in the area of the Median nerve and other symptoms. Secondarily it will determine whether stretching, added to OMT makes a difference in the functioning and edema and dimensions of the carpal canal.*

Example

(strong) *The following aims will enable us to validate the diabetic dog model, demonstrate the investigative team's proficiency in the proposed experimental techniques and procedures, and provide an experimental foundation for extramural applications to investigate the impact of antioxidant therapies on diabetic sequelae.*
Aim 1. *To test the hypothesis that persistent hyperglycemia increases myocardial stiffness and decreases left ventricular compliance in diabetic dogs.*

Aim 2. *To test the hypothesis that diabetes increases the energy costs of cardiac performance, lowers myocardial oxygen utilization efficiency, and exacerbates myocardial energy depletion during coronary underperfusion by forcing the myocardium to consume fatty acids as its principal energy substrate.*

(weak) *The aims for this study include testing whether hyperglycemia increases stiffness and changes ventricle compliance in dogs who have diabetes, and also studying whether diabetes affects the energy costs or oxygen efficiency in the consumption of fatty acids as a principal substrate.*

Example

(strong) *The principal objective of this proposed research project is to test the reproducibility of earlier findings in this line of research in a manner that will clarify the relative contributions of touch and the placebo effect relative to the effects of OMT with a multi-center randomized clinical trial.*

(weak) *This proposal proposes to use a multi-center design to study previous findings and clarify the ways touch and placebo contribute to the effects of OMT.*

Key Question: How will I achieve the aims?
(By answering the questions posed as hypotheses)

Example

(strong) *We propose to conduct a randomized, controlled clinical trial of six weeks duration with a three-month follow-up assessment with four treatment groups, to test six hypotheses, two for each of the three stated aims.*

Hypothesis 1. *OMT plus stretching exercise will most effectively improve the Median motor and sensory distal latency and amplitude over a sub-acute interval.*

Hypothesis 2. *OMT plus stretching exercise will be the most effective intervention for decreasing negative symptoms of CTS*

(weak) *This study will help determine if these hypotheses are true, that Median motor and sensory distal latencies and amplitudes will improve in CTS patients who get OMT and do stretching exercises, and negative symptoms of CTS will decrease when patients include stretching with OMT.*

Example

(strong) *The two main hypotheses to be tested with this proposed research project are as follow.*

Hypothesis 1. *Osteopathic manipulative treatment will be, whereas the sham treatment and conventional care only will not be associated with decreased health care resource utilization for elderly patients hospitalized with pneumonia.*

Hypothesis 2. *Osteopathic manipulative treatment will be, and sham treatment and conventional care only will not be correlated with improved health status in elderly patients hospitalized with pneumonia. For this proposal, health care resource utilization is defined as hospital length of stay, duration of antibiotic use in the hospital (intravenous and oral), in-hospital complications (ventilator dependent respiratory failure and transfers to special care units), and re-admission rates to the hospital. Also, for this proposal, health outcomes are defined as time to clinical stability specifically for systolic blood pressure, heart rate, respiratory rate, temperature, oxygen saturation, the ability to eat, and improved mental status, and as time to overall clinical stability; change in leukocyte count; mortality; patient satisfaction; and the rate of symptomatic and functional recovery. Additional data will be collected on somatic dysfunction, confidence in treatment, side effects of treatment, and adequacy of sham treatment to blind subjects to group assignment.*

(weak) *The hypotheses for this study are:*

1. *OMT will improve the way the elderly use health care resources in the hospital.*
2. *Sham and conventional care will not be related to how fast the elderly patient improves his or her health status.*

We define the measurements as length of stay, use of antibiotics, readmission, blood pressure, heart rate, and other medical indicators of health. We are also interested in mortality, satisfaction, somatic dysfunction, and the placebo-effect.

B. Background and Significance

In the Background and Significance section of the proposal you need to describe the existing research literature or scientific body of knowledge that builds the rationale for your proposal. This is one of the most difficult sections to write succinctly, coherently, and powerfully. You will write and re-write it, and re-write it again. Frustration is normal. Your mentor may see something on the third edition she did not see on the first – that is how it evolves – that is why you will have milestones, product deadlines and regular meetings with key individuals to write your proposal – or later your report.

Discuss how the current scientific knowledge in your specific area of focus supports your hypothesis or reveals important gaps in the scientific knowledge that you will address. The Background and Significance section should be written in such a way that it leads the reader to agree that there is either evidence from previous research that supports the direction of your proposal or that there is insufficient relevant research on the subject although the area is an important one to study. This is the section to present the theoretical framework in which your research is grounded.

If you are replicating a study, this is the place to say whose study, why, and what the original study did, and how you will improve knowledge by replicating it. If you are developing a new instrument for measuring outcomes of an intervention or chemical changes from a laboratory experiment, make your case for its value in this section. You can divide this section into each major aim, hypothesis, outcome, or strategy that you will use by using separate headings such as: OMT as an intervention in pneumonia, or outcome studies of surgical interventions in CTS. You might use existing data no one has mined with your hypothesis or question.

This section should logically lead the reader into your Methods Section and have them say “now I see” why or “oh wow,” or “yes!”

Key Questions to guide your thinking are as follow.

- What is the scientific or clinical theoretical framework for this research?
- What do we currently know about this area?
- What are the strengths of the current body of knowledge?
- What are the gaps in the current body of knowledge?
- How will this proposed study build on the strengths, correct the limitations and fill the gaps (or begin to do this)?

This Background and Significance section may be organized many different ways. A good rule-of-thumb is to use it to logically support each of your aims and hypotheses. Include information that refutes your hypotheses also, to show that you know about these and will address them. If you are writing about non-surgical interventions, discuss what is pertinent about surgery. If you are writing about women’s health, discuss similar conditions in the male population and how women are different. Use comparison groups and related subjects to provide a full and rich picture to the reader.

This section is normally three to five pages in a funding proposal of 25 pages. For your thesis proposal it may be only two to three pages at most.

Examples follow.

Background and Significance Examples

Example

(strong)

B.1. Introduction

The first published report of Carpal Tunnel Syndrome (CTS) was in 1856, and is now the most commonly reported neuropathy. It is found most frequently in women. In men and women it is most prevalent between the ages of - - -. Risk factors include - - - -. CTS is reported more in white than non-white populations. (Becker, 2002) However, we do not know if this is because non-white men and women with symptoms do not present to the health care system for evaluation and treatment. Persons with CTS may have pain all night… severe thenar atrophy.[2] The National Center for Health Statistics reports that CTS accounts for the highest average number of lost workdays compared to- - - . The study we propose may generate data that would (do what - - -).

B.2 Clinical Diagnosis of Carpal Tunnel Syndrome

CTS is an entrapment neuropathy (now define it). Specific tests measure the electrical response of a nerve to an electrical stimulus and any delays in the conductance … (latency period) between the stimulus site …. NCS are used to establish a quantitative baseline assessment against which to measure the outcomes of therapy.[2]

B.3. The literature

In reviewing the literature using terms associated with OMT, CTS, NCS, electrodiagnosis (EDx), symptoms, …and outcomes of CTS interventions we identified N articles on post-surgery outcomes, and we relied primarily on the most recent studies. Although the literature provides ample information on the clinical course of CTS, we found no clinical trials assessing the effectiveness of manual medicine using a rigorous prospective, randomized, blinded, controlled research design. None of the studies we reviewed considered edema as an outcome measure, even though edema may impact CTS symptoms. The following section is organized around the primary and secondary outcomes - - - ,and how we will measure them.

Symptoms and Functioning In 1993 David Levine (did what). The resulting Symptom Severity Scale (SSS) and Functional Status Scale (FSS) have been found to be reliable, internally consistent, and correlated with sensory testing. (Levine, Patterson, Padua, Katz). We therefore will ….

B.4. Summary

From a review of the available literature we have determined that all published research on manual medicine and CTS are flawed in several ways. These include - - - -. This proposed study will correct these flaws by - - - -.

(poorly constructed)

B.1. Introduction

Carpal Tunnel Syndrome started in the 1800's and is most commonly reported more in women than in males. It is most prevalent between 30 and 40, with risks more in black or Hispanic or asian than caucasions. No one knows why this is true. CTS presents all night pain and severe atrophy for some people. Furthermore The National Center for Health Statistics says that it causes the most lost workdays compared to …. This proposal will create information to better understand this.

B.2 Diagnosing CTS

CTS is found by measuring the electrical response of a nerve to an electrical stimulus and any delays in the conductance. This establishes …a baseline to use in determining how therapy affects…...

B.3. The literature

When we reviewed the information published related to CTS and our outcomes we found few articles about them, so we relied primarily on the most recent studies. Even those studies were not directed at effects of OMT, and did not randomize subjects. Edema did not get included in any of them. But we think edema is important from signal intensity. Next we discuss our outcomes to be measured in this study.

Example

(strong)

1. Scope of the problem

In the United States about 600,000 persons each year are hospitalized with pneumonia. Community-acquired pneumonia is estimated to be the sixth leading cause of death and - - - - - - . Because of the increased frequency of pneumonia, longer hospital stays, and - - - - , persons over age 65 are particularly likely to benefit from the application of adjunctive therapies such as OMT.

2. Why osteopathic manipulative treatment is well-suited for pneumonia

In order to understand why OMT is particularly suited for treatment of pneumonia, it is useful to understand how manipulative medicine specialists use manual forces to open up the free flow of the vessels, thus helping the body heal itself. We suggest that better structure means better function and better function means better host defense. Use of OMT for pneumonia and other pulmonary disorders is advocated by the osteopathic medical profession. However, applications of OMT have not yet been tested using modern clinical research techniques.

3. Literature supporting mechanisms of action

One mechanism theory for the effect of OMT in treatment of pneumonia is that OMT works through … mediated through a viscerosomatic reflex. [18] It is theorized that, … in the thoracic paraspinal muscles, …However, few clinical trials have been conducted to test this theory. Ortley et al. reported that "the heart rate tended to decrease in a number of subjects" in study of six healthy young adults. This study found (what). Other studies suggest a second mechanism should be considered to be - - - - . Allen and Pence reported using spirometry to measure the effects of OMT in pulmonary disease. They studied (describe the subjects and the methods used).

4. Clinical trials of osteopathic manipulation for pneumonia

There are several small, poorly controlled studies, published in a non-peer reviewed journal, … the time period to recovery was 9.2 days in the OMT only group, 8.3 days in the antibiotic only group, and 6.4 days in the OMT plus antibiotic group.

5. Relevant non-osteopathic literature

Intensive bottle breathing has been shown to shorten the length of stay in patients hospitalized with pneumonia. …Elderly patients are known to have decreased chest wall compliance and diaphragmatic mobility. [45 46]

6. Significance of the proposed study

As outlined in the introductory paragraphs, pneumonia is a major health problem, particularly in the elderly. Because the problem is so large, any new adjunctive modality that is proven effective has the potential to relieve suffering, lower health care costs, and gain widespread acceptance. …If OMT does shorten hospital stays, reduce the duration of IV antibiotics, and/or lower …, then OMT is likely to prove to be an economical and practical modality.

(weak)

This study and others like it will have an enormous impact upon the osteopathic profession.

AVOID STATEMENTS LIKE THIS. IT DRAWS A CONCLUSION THAT IS NOT SUPPORTED. YOU CAN SAY "This proposed study may have…" Not "will have".

C. Preliminary Studies

This is the section to describe your own preliminary data from research on this topic. It may not apply to your project. If your mentor has preliminary data and is the PI, this section should be written using her data.

There is another possible use for this section, even if there is no preliminary data from your own independent or collaborative work.

You may use this section to discuss your other experiences or knowledge to show your competence to conduct this study. Or you may want to demonstrate competence among the research team in the clinical topic or to show how the institution or research team has special unique properties making it most likely to succeed.

D. Research Design and Methods

The Research Design and Methods Section is usually the largest section of the proposal. It can be organized in many different ways according to the nature of your research. This section describes the following in carefully detailed steps:

- How you will achieve the specific aims
- What activities you will undertake to test the hypotheses for your study
- What you will do in your experiment or what you will do to the human or animal subjects
- How you will do it
- When
- Where
- How
- How often
- Why
- What will you use to determine if the action you propose will generate the reaction or the result/outcome you hypothesize that it will generate – your measurements.

In this section you must address:

1. Subjects: Who will be in the study? How will you select them by inclusion and exclusion criteria, and how will you recruit them? (NOTE: Human subjects (risk and exclusion justification) issues are covered in a separate section E and do not go here)
2. The independent variables: What you are observing or what you expect to change – the outcomes or measures of change
3. The 'Power' of the study: See guidelines on when and how to obtain a power analysis to estimate sample size
4. The dependent variables: What you are using as an intervention or treatment
5. Other data such as categorical variables
6. The hypotheses to be tested, question to be explored, or the product to be developed
7. The design of the study: randomized block, repeated measures, longitudinal, retrospective, prospective, cross-sectional, existing data, controlled, blinded, and the like.
8. The instrumentation to collect data, and to measure the experimental change, the clinical outcomes, the validity of the theory, or the utility/validity of the product
9. The statistical tests you plan to use to test the hypotheses or question

(NOTE: Examples are divided by plausible but not required sub-sections)

D.1. Overview

Example

D.1. Overview

The overall goal for this proposed exploratory/developmental phase II clinical trial is to build a scientific body of knowledge concerning the impact of osteopathic manipulative treatment (OMT) on CTS and establish a plausible model to test in larger more definitive studies. This study will use a randomized, controlled, blinded design with a three-month follow-up to examine whether OMT with and without stretching exercise provides a more effective treatment for CTS than non-surgical standard care, or a sub-therapeutic ultrasound placebo.

OMT and OMT plus stretching may favorably impact CTS by several mechanisms including 1) mechanical expansion of the carpal tunnel resulting in decreased mechanical compression of the Median nerve; 2) decreased tension within upper extremity fascial strain patterns that may be compressing low pressure lymphatic and venous vessels and contributing to fluid congestion and edema within the carpal tunnel; and 3) decreasing minor upper extremity neural impingements at the nerve roots, brachial plexus, Median or Ulnar nerves that may be causative of other upper extremity symptomatology and may be contributory to the "Double Crush Syndrome" of the Median nerve at the wrist.

Based on this understanding of the mechanisms of OMT and stretching exercise on CTS, six hypotheses will guide this study. Each hypothesis links to an outcome measure.

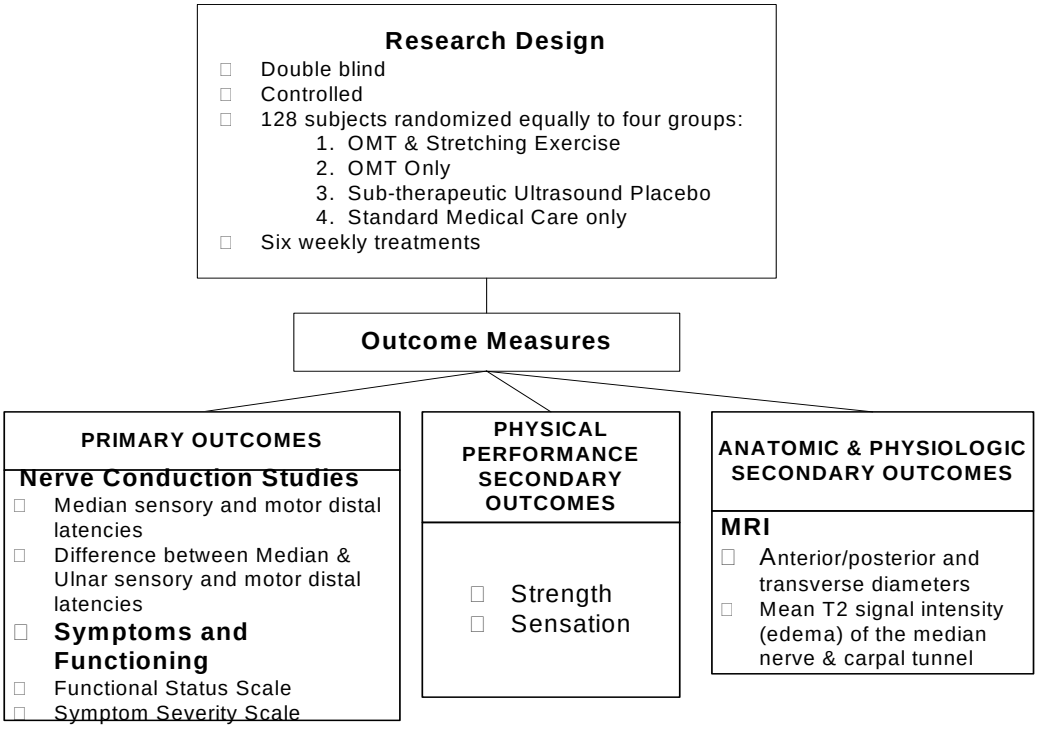


Figure 3. Research Design

We propose to collect two primary outcome measures:
1. Median motor and sensory distal latencies by electrodiagnostic testing, specifically nerve conduction studies (NCS)
2. Symptom severity and functioning (using the Levine Symptom Severity and Functional Status Scales)
We will also collect four secondary outcome measures: 1. Physical strength (grip and pinch), 2. Sensation quantified by measuring current perceptual threshold (CPT), 3. Change in the cross-sectional area of the carpal tunnel (including ...

D.2. Subjects

D.2.1. Power Analysis

Example

In order to allow for subgroup analyses and for stratification on the multiple sites a total of 360 subjects, 120 in each group, will be recruited into the study. Each site will recruit 90 subjects, 30 in each group, over the course of the study. Censored values (those dying in the hospital and those placed on a ventilator) and dropouts are estimated to account for no more than 20 percent of the total number of subjects. Based on preliminary data on length of hospital stay, when the sample size of non-censored values in each group is 96, an exponential maximum likelihood test of equality of survival curves with an α=0.05 two-sided significance level will have power of 0.80 to detect the difference between a median length of hospital stay of 6 days versus 9 days. Therefore, recruitment of 120 subjects per group should be sufficient to detect the hypothesized differences in hospital length of stay. Figure 7 on the following page illustrates the projected enrollment timeline for the 360 subjects.

D.2.2. Subject Selection and Recruitment

Example 1

Inclusion Criteria

Inclusion criteria for participation in this study are:
1) age 21 to 70
2) clinical diagnosis of carpal tunnel syndrome
3) nerve conduction studies consistent with CTS: median nerve sensory distal latency greater than 2.2 ms, a difference between median and ulnar sensory distal latency greater than 0.3 ms, median nerve motor distal latency greater than 4.2 ms, and/or a difference between median and ulnar motor distal latency greater than 1.5 ms
Inclusion nerve conduction studies (NCS) will be performed immediately after obtaining informed consent to verify that subjects meet the electrodiagnostic inclusion criteria for CTS.
Those subjects who do not meet the inclusion criteria for NCS will be immediately excluded from the study.

Exclusion Criteria
Exclusion criteria for participation in this study are:
1) severe CTS that has progressed to muscle atrophy
2) pregnancy
3) previous wrist surgery
4) systemic disease which includes but is not limited to: diabetes mellitus, thyroid disorders, rheumatoid arthritis, Paget's bone disease, gout, myxedema, multiple myeloma, acromegaly, hepatic disease, dialysis patients, and other disease in which peripheral neuropathies are common
5) secondary causes of CTS including, but not limited to, ganglion cyst, mass, or accessory muscle.

Exclusion criteria for participation in the MRI portion of the study protocol are:
1) cardiac pacemaker
2) hip prosthesis
3) metallic foreign body in the immediate vicinity of affected wrists
4) other conditions which serve as a medical contraindication to MRI as determined by the radiologist/MRI staff.

Example 2

Inclusion Criteria: *Subjects must be 60 years of age or older hospitalized with acute community-acquired or nursing home-acquired pneumonia. Pneumonia is defined by. . .. . and the presence of at least two of the following signs and symptoms compatible with pneumonia: (1) new or increased cough; (2) new or increased sputum production; (3)…. (4)…. (8) worsening of mental or functional status.*

Exclusion Criteria: *Patients with lung abscess, advancing pulmonary fibrosis, established bronchiectasis, pulmonary tuberculosis, and underlying lung cancer will be excluded from the study. Patients unable to cooperate with or tolerate the study protocol treatments will be excluded. Those with metabolic bone diseases (i.e. Paget's disease, hypoparathyroidism) that put them at risk for pathologic bone fractures will be excluded. Should the original diagnosis of pneumonia change to a different diagnosis during the course of the hospital stay, these patients will be withdrawn from the study. Subjects unable to complete the study for whatever reason will continue to have outcomes data collected to facilitate intention-to-treat analysis, unless declined by the subject or guardian.*

D.3. Methods

Begin with a statement such as "This section describes how we will collect the data pertinent to this proposed study. It is divided into (your number) sections: 1) study protocol, 2) experimental groups and interventions 3)……."

D.3.1. Study Protocol

Example

Figure 4 is the study protocol schedule. All subjects will be enrolled in the study via the level one screening and consent process by a clinical research coordinator (CRC). Level two screening data will be collected by the CRC with consultation and review by the PI or Co-I. All treatment and outcome visits will occur in the Osteopathic Manipulative Medicine (OMM) clinic at UNTHSC/TCOM. MRI tests will be done at the Monticello Diagnostic Clinic, described in the Resources Section of the proposal. Subjects will receive $10 for their travel time and expenses for each study visit. All forms (Demographic Data Sheet, SOAP Note Form, Symptoms Severity Scale, Functional Status Scale, Data Collection Form) are contained in Appendix A.

D.3.2. Experimental Groups and Interventions

Example

The experimental design for this study is a two-factor design, with repeated measures on one factor, and is illustrated in Figure 6. The independent factor is the treatment group that consists of three levels: OMT group, sham treatment group, and conventional care only group. The repeated factor is time which consists of a variable number of levels, dependent on the individual subject and the time until s/he reaches one of the conditions for stopping the treatments. A total of 360 subjects, 120 in each group, will be recruited into the study. After informed consent is obtained, the study participants will be stratified (as described in Section D.4) within each site and then randomized into one of three treatment groups. Measurements will be taken on all subjects preceding the initial treatment (as described in Section D.5). Following the pre-treatment measurements, subjects will receive … During the study period, all subjects will continue to receive ….and the extent of this care will be monitored.

D.4. Methods

D.4.1. Outcome Measures

Example (adequate but could be stronger by eliminating the rationale statements that should be elsewhere in the proposal)

Hypothesis 1: *Persistent hyperglycemia increases myocardial stiffness and decreases left ventricular compliance.Rationale. Hyperglycemia causes oxyradical-mediated crosslinking of myocardial interstitial collagen.[4] As these crosslinks, termed advanced glycation endproducts (AGEs) accumulate, the ventricles become … Although myocardial stiffening has been well characterized in diabetic rats[24] it has never been examined in large diabetic mammals including dogs and humans, and the impact of antioxidants on myocardial stiffness has not been tested in any species.*

Protocol and interpretation. Left ventricular chamber stiffness will be assessed from end-diastolic pressure:volume relationships obtained in conscious dogs. Left ventricular pressure and volume will be measured … These studies will be conducted weekly to monitor the …..and compared with pre-alloxan baseline and values obtained in non-diabetic dogs.

D.5. Data Analysis Plan

Example 1

Secondary aim one is to evaluate the effects of OMT with and without stretching exercise on strength and nerve sensation in persons with CTS.

Hypothesis 3*. OMT plus stretching exercise will most effectively improve strength.*

Grip strength will be measured with the Jaymar Dynomometer™. A pinch gauge by Baseline® will be used for pinch strength. Adjustments will be made as necessary for males and females, and for age. To test this hypothesis grip and pinch strength will be analyzed using repeated measures ANOVA between groups, and t-tests for gender and other relevant classification variables.

Hypothesis 4*. OMT plus stretching exercise will most effectively improve sensation.*

Sensation data will be in the form of microampere sensation thresholds for Median and Ulnar nerves at the three frequencies tested for each nerve (2000, 250 and 5 Hz). Median and Ulnar sensation thresholds will be compared using repeated measures analysis of variance (ANOVA) between treatment groups to determine which groups had the greatest improvement in sensation as measured by CPT.

Example 2

Statistical Analysis

To test the hypotheses regarding group differences on health care resource utilization and health outcomes, four methods of analysis will be used. Statistical methods that stratify by site will be used for analyses of data from more than one site. To compare the three groups on length of hospital stay, duration of IV antibiotic use, duration of oral antibiotic use in the hospital, and time to clinical stability, stratified Cox proportional hazards models will be employed. [65] The differences in the effects of the three groups on continuous outcome measures where repeated measurements are made (e.g., leukocyte counts, rate of symptomatic and functional recovery, somatic dysfunction burden) will be examined using general linear mixed-effects models (marginal models) with the subjects and the sites treated as random effects and the group as a fixed effect. [66] For repeated categorical outcome measures (e.g., in-hospital complications), generalized estimating equations (GEE) will be utilized. [67, 68] Analysis of categorical outcome measures that are measured only once (for example, reason for ending study treatments, re-admission rates) will be performed using logistic regression models. [68]

INTENTIONALLY BLANK CONTINUED ON NEXT PAGE

D.6. Study Timetable and Management Plan

A research proposal should always contain a timetable and management plan, however brief it may be. It lets the reader know that you have considered the logistics of carrying out your research plan. The shape of this depends on your research, particularly if it is sequential, or multi-center, or involves a data safety and management board (DSMB) or other advisory groups. It lists the key steps involved in the project from start-up to final report and marks the weeks or months during which these will occur. It may be similar in format to the protocol time schedule and assures the reviewer that you have a plan in mind even though it may be modified later.

Example

This phase III clinical trial will follow the general plan presented in Figure 5 below to achieve the aims of this proposed study.

Major Activities	Grant Year 1												Grant Year 2											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Complete study Manual	X	X																						
Training on intervention reliability		X	X					X				X				X								
Subject Recruitment			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Enrollment & screening/baseline measures			X		X		X		X		X	X		X										
Interventions			X	X	X	X	X		X	X	X		X	X	X		X	X	X		X	X	X	
Evaluation of progress and protocol								X				X				X				X				X
Six month follow-up											X	X		X	X		X	X		X	X			
Study analysis & final report																								
Publications & possible R-01 application																								

Major Activities (cont.)	Grant Year 3											
	25	26	27	28	29	30	31	32	33	34	35	36
Complete study Manual												
Training on intervention reliability												
Subject Recruitment	X	X	X	X								
Enrollment & screening/baseline measures												
Interventions	X	X	X	X								
Evaluation of progress and protocol				X								
Six month follow-up	X	X	X		X	X						
Study analysis & final report								X	X	X	X	X
Publications & possible R-01 application							X	X	X	X	X	X

If this proposed exploratory study suggests that changes do occur as a result of OMT with or without stretching, we intend to pursue and participate in a larger multi-center study that would target specific etiologies of CTS. Regardless of the outcomes of the study, completion of this protocol, its analysis, and publication of the findings will have provided information to the osteopathic physician in clinical practice, and contributed to the training …..

***Figure 5**. Project Management Timetable*

E. Human Subjects

This section has a separate chapter.

G. Literature Cited

You may use number references in the text and number the literature list, OR use names in parentheses and alpha-order the literature list.

Example

Abate C, Patel L, Rauscher FJ III, Curran T (1990) Redox regulation of Fos and Jun DNA-binding activity in vitro. *Science* **249:** 1157-1161.

Andrews NC, Faller DV (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* **19:** 2499.

THIS PAGE LEFT BLANK ON PURPOSE

Osteopathic Research Center
Fostering nationwide collaborative OMM research

Appendix 1 for Chapter I
RESEARCH PROPOSAL WORKSHEETS

GUIDE FOR WORKSHEETS

The sections in this appendix will help you write statements and draw figures to think though specific parts of your research project. Completing these questions will help you make connections between the parts of the research plan. You will literally sketch out your ideas, steps, and activities you will need to guide you from idea through formulation of hypotheses and design to implementation. Your responses to these items and prompts should take you from the first step on day one to the last step on the day that you will write your report.

SECTION I
THE QUESTION

This section will help you formulate your primary research question or hypothesis. You should finish this section with one sentence. The problem can be stated as a question (e.g. Is there a relationship between Somatic Dysfunction and chronic health problems in persons over age 65 receiving continuing care from a family medicine clinic?) or as a hypothesis (e.g. The worse the chronic health condition a patient has, the more pronounced will be the Somatic Dysfunction. OR if you have two groups you might say Somatic Dysfunction will be more pronounced in persons over age 65 with chronic health problems than in those with no chronic conditions.)

1. Statement of problem of interest:
Name the condition or disease.
Name the target population of interest.
Name the comparison group if any.

2. Is this problem or question of long-term interest like scoliosis, or a does it address conditions related to a temporal event like asthma?

3. Write a brief statement of why the condition of interest is a problem? (This is the "so what" answer. e.g. Restrictions in movement in persons over 65 may exacerbate a chronic health problem leading to overuse of medications for pain or limiting adequate oxygenation). Consider for example whether this is a public health concern, or a cost burden to society.

4. Write one sentence for your main research question or hypothesis. It must withstand the test of being focused and answerable.

5. To answer your question or test your hypothesis what research design will you use?
Experimental ____ Number of groups _____, placebo control or standard care control? _____
Non-experimental ___ e.g. retrospective cohort, cross-sectional prospective, cross-sectional case-control, other

6. What exactly will you use to measure your expected results or answer your question?

SECTION II
TASKS AND TIME LINES

7. Right now, as you conceive of it, draw a conceptual flow chart or diagram of exactly what you will do in your research. This helps form the methods section of your proposal. You may come back and change this later, but start now.

8. Right now, as much as you can, associate times such as days and weeks with the tasks needed to achieve your research goals. Start with when it will end and you are writing your report, and work backwards. Include, for example, where you will go, what you will get or see, what tools you will need to collect information and where you will get them, and what your subjects will do and when. Think about who else will be involved from the IRB to your major professor and any staff involved or touched by this project

WHAT/TASK	WHEN	WHO
Power analysis to determine sample size using information from published research on effect size – amount of change or difference anticipated	When I have my question or hypothesis finished	A biostatistician

SECTION III
OUTCOMES – MEASUREMENTS

9. What are the outcome measures of interest? These are your variables.
Specifically what do you expect to change or to happen as a result of the intervention? Or what are you going to analyze as a measure of change or difference? How will you define your outcomes of interest? i.e. in the example used say: For this study Somatic Dysfunction will be the total number of lesions documented in the chart at the time of the first visit. For a non-experimental cross sectional descriptive study, consider for example what you want to examine such as asking if a certain group of patients show medical or health status changes following OMT for a particular condition such as back pain?

10. How will you know the answer to your question? _____

_____

11. What information will you collect?
How will you measure each of your variables? Think about how you will be consistent especially if different source documents use different measurements (i.e. inches compare to inches, palpatory findings must be consistent or you must consider how many different practitioners' measurements you will use and consider this in the power analysis)
Define the outcome: i.e. don't just say blood pressure, say health status will be measured by the following criteria. 1. sputum will be collected as …… x times….. and tested using….. the values to be included in this study are ……, 2. blood pressure will be recorded at five minutes after the rest period following collection of heart rate…….

OUTCOME VARIABLE	DEFINITION	SOURCE

12. If you are using health indicators are they in a record or do you have to collect them? If you are asking questions about functionality is there an existing measurement instrument or questionnaire or do you have to create one?

List the data collection forms (DCFs).

DCF	SOURCE AND OTHER INFORMATION

13. Name the Independent (what will not vary or change) and Dependent Variables (your outcome variables of interest measured as change in a condition, a number of actions or events, or a status)

14. Are you comparing change in different age groups with a particular disease who used an OMT clinic or those who got OMT and those who did not, for a particular condition?
List any grouping variables that will be independent variables or factors such as age, gender, presence or absence of another condition or quality.

15. List the variables you will test as predictor or criterion variables, such as a number of prenatal visits could predict birth weight of baby for women who had OMT

SECTION IV
INCLUSION/EXCLUSION CRITERIA

16. List the characteristics or criteria by which you will select your records or subject for the study (inclusion criteria).
List those by which you will exclude records or subjects (exclusion criteria). (e.g. age, health condition, lab value, time periods, services criteria, locations)

INCLUSION	EXCLUSION

SECTION V
ACQUISITION AND RANDOMIZATION

17. How will you **acquire the sample** of records or persons? Where will you get the information or subjects, how will you recruit or access them?

18. If this is an **experimental design**, what are the experimental groups? How will you make random assignment?

SECTION VI
REVIEW AND REVISE

19. REVIEW: Look back at the question or hypothesis you wrote on page one of the worksheets.

REWRITE one sentence for your main research question or hypothesis. It must withstand the test of being focused and answerable

20. Ask someone to look at the question you wrote and ask you questions about it. Have a discussion to refine these key items.

After revising or re-writing your primary question or hypothesis, review the drawing of your research question, methods and outcomes and the time-table in items 7 and 8. REVISE THESE IN THE SPACE BELOW IF YOU NEED TO DO THAT NOW.

(7) Flow chart or diagram of exactly what you will do in your research.

(8) Time-table

WHAT/TASK	WHEN	WHO
Power analysis to determine sample size using information from published research on effect size – amount of change or difference anticipated	When I have my question or hypothesis finished	A biostatistician

SECTION VII
Data Management and Analysis

21. REVIEW AND REVISE THE DCF LIST AND USE IT TO ANSWER THESE QUESTIONS

21.1. What will your data collection forms look like? Attach samples. If you develop your own you must test them and adjust them before you begin the data collection.

21.2. Are the collection instruments in the public domain, or is a cost associated? Is a license required? (It is important to not accept someone's opinion that you can use an instrument. You would want credit given to you if you developed an instrument. Also, there are many good public domain or other-student-developed instruments available.)

21.3. How will you document validity and/or reliability of the measurement instruments?

21.4. Will you attempt to "generalize" to a larger population or other populations?

21.5. Which computer software will you use to enter data?

21.6. How and where will data be stored?

21.7. How will you protect the data from errors or tampering?

21.8. How will data be guaranteed as confidential or anonymous?

21.9. Identify specific data analysis approach and appropriate statistical tests of the data.

21.10. Identify data characteristics that may influence the analysis methods.

Osteopathic Research Center
Fostering nationwide collaborative OMM research

Appendix 2 for Chapter I
RESEARCH DESIGNS FOR OMM RESEARCHERS

SECTION 1
INTRODUCTION

This appendix describes several different types of research designs. For each design we have listed some of the benefits and limitations. We have also provided some of the key words essential to your understanding of these designs.

Knowing what types of research designs are most commonly used in OMM research as well as other types is vital to your competencies in research. This will assist you in your discussions and your readings of published research. Certain designs are companions to certain research questions. The question drives the design and the design drives the methods section and data analysis.

Key Concepts in Research Design

Random *selection* (sampling) – in choosing participants for your study, each person in the population has an equal chance of being selected.

Random *assignment* – when assigning participants to the various groups, each person has an equal chance of being placed into one particular group.

Validity: *There are many types of "validity" in research design.*

> ***Internal* Validity** – control for all outside influences that may influence the characteristics or status of the groups being studied, except of course the variable of interest. True internal validity has no extraneous factors affecting it. This is difficult in the real world. For example, can you tell a patient not to get any other treatment except the study treatment in an OMM study? No. But you can control for extraneous variables. Those variables can become part of the study such as medication use.
>
> Random assignment helps control factors that might contaminate internal validity but nothing controls for it completely. Causal inference from a study finding is typically NOT DONE. Is there anything you can think of that causes something to happen and is accounted for as the ONLY thing that makes "it" happen?
>
> ***External* Validity** – determines whether or not the research can be generalized to the larger population. External validity depends on the power of the study, the method and criteria for selection of records or participants. It may also depend on the measurement instruments selected. Some instruments are designed to measure pain and others are designed to measure back pain specifically.

Reliability – Can your results be repeated by other researchers using your same methods?

Experimental Design

This would be a research project in which you utilize an intervention or observe a population under a certain controlled condition. For these studies you may elect to conduct pilot research using small numbers of subjects. This type of research is particularly valuable to discover and uncover MECHANISMS of action related to manual/manipulative medicine treatments

SECTION 2
TYPES OF NON-EXPERIMENTAL RESEARCH DESIGNS

Five Research Designs

1. Causal Comparative
2. Correlational
3. Cohort Studies
4. Cross Sectional
5. Case Control

Each of these research designs can be conducted by a resident or a student. These are "publishable" as posters or short manuscripts. These studies all require Institutional Review Board approvals and in some cases will require something called a "HIPPA" waiver. Any research involving information about a person's protected health information will require IRB approval and some method of addressing consent to participate. There are many web-sites that help make decisions about human subjects' protection assurances related to a particular type of research design and specific populations such as pregnant women or children.

1. Causal Comparative Studies

Are you studying two groups from the same population that might be different on a critical characteristic or condition? When it is not feasible to use an experimental design or to "manipulate" a variable or the environment or use an intervention, you might consider this design.

Compare outcomes by provider or by age group or by some other control variable by collecting S/D information, numbers of treatments, treatment modalities, and results in the records of several osteopathic physicians, grouping patients by a disease condition.

Compare the number of health complaints over a specified period of time –

or compare those with the same diagnosis who did and who did not receive OMT for utilization of other medical services or use of NSAIDS or use of psychiatric medications. Collect information from the records of a group of patients diagnosed with fibromyalgia who receive OMT and those without that diagnosis matched for age, body mass index and social history or other health indicators.

Earaches or otitis media events in children who do and who do not receive OMT would be another question that would use this design.

In this design you are asking about the possible "causal" relationship between an event or pattern of events and the observed behavior. However you cannot infer cause, only the strength of the relationship.

Hypothesis or question
- **Base on your own 'theory' with grounding in the literature to support the theory logically**

Groups
- **Ensure one has the characteristic and one does not: all other descriptors the same (matched)**

Methods (and data management)
- **Carefully select measurements**
- **Parametric**
- **Validity threats and strengths**

Limits
- **Never prove 'cause'**
- **Type I error**

Strengths
- **Builds pilot data for future research and may be theory-building**
- **Exploratory Developmental**
- **Fun**
- **Relatively easy**
- **Can be prospective (longitudinal) or retrospective (historical records, interviews)**

2. Correlational Studies

In a "correlational" study we compare two or more variables (factors or characteristics or qualities) in one group. For example, "what factors contribute to selection of osteopathic physician PCP for women compared to me?" or "What factors characterize mothers who bring children to DOs for manipulative medicine treatments?"

A study of the relationship between chronic illness and somatic dysfunction in elderly patients would be a correlational study.

Hypothesis or Question
- **Base on your own 'theory' with grounding in the literature to support the theory logically**
- **Shoe size and earned income**
- **Not a strong correlation (using two different measures of independent functioning)**

Group
- **Only one: Precise definitions of the characteristics of interest (may subset)**

Methods (and data management)
- **Scatter Plots are useful**
- **Linear and curvilinear relationships (Circumplex model of family systems)**
- **Multiple statistical tools – care in choosing**

Limits
- **In context**
- **Type I error**

Strengths
- **Leads to larger research**
- **Theory building**
- **Exploratory Developmental**
- **Fun**

3. Cohort Study

This type of design is used to compare an outcome of interest between two groups over time, usually those with a given characteristic and those without it.

Hypothesis or Question
- **Base on your own 'theory' with grounding in the literature to support the theory logically**
- **Reoccurrence of a condition over time in a group with and without OMT**
- **Compares incidence of a condition**

Group
- **Two groups are usually matched on some characteristic (gender, age, condition-receiving and not receiving a treatment)**

Methods (and data management)
- **Analysis of variance and Multiple analysis of variance, t-tests**
- **Survival analysis**
- **Multiple statistical tools – care in choosing**

Limits
- **In context**
- **Type I error**
- **Time, expense, attrition**

Strengths
- **Leads to larger research**
- **Theory building**
- **Exploratory Developmental**
- **Fun**
- **Can provide insight to risk factors for conditions**

4. Cross Sectional Study

A cross sectional study design is used to describe a population on several dimensions. It captures a condition at a specific moment in time. You can determine or estimate the prevalence of a condition in a population. You can use this to generate questions for future research. You may have a new clinic that has started using specific OMM treatment strategies such as cranial manipulation, or a population with COPD being seen in an OMT clinic. If there are adequate numbers to power a study you could use exploratory data analysis tools to richly describe the osteopathic findings and treatments by disease severity to document whether practitioners approach different populations with the same health condition differently when using OMT.

You could be innovative with the statistical analysis and use a discriminant analysis function to determine if age of disease predicts treatment techniques or number of visits or if the clinic uses the St. George Respiratory Questionnaire you can examine outcomes in a singe population from baseline before treatment to xth treatment.

5. Case Control Study

A case-control study is basically a case report comparing those with and those without a particular condition. If you can get enough information on scoliosis patients from an OMT provider and a non-OMT provider you could trace clinical indicators for the two groups. Surgeries, for example could be a variable to determine if the OMT recipients had fewer surgeries. Headaches would be another possible topic if you can match the cases and track/record from the record use of prescription pain relievers or number of visits for other health complaints.

These studies are typically of not much value unless the condition is rare or the treatment is experimental or novel as applied to a particular condition. If you had access to records from another clinic in another school you could compare treatment strategies for a given condition. Carpal Tunnel Syndrome is a good topic because there is not one standard protocol for relieving pain or improving functioning. Remember research is not clinical practice. In research there must be ONE protocol tested against another one not different protocols for each patient individualized and customized to their own needs – not until the mechanism and the efficacy is first demonstrated scientifically.

Use of Questionnaires

When developing a survey ALWAYS look and look again, and then check again to determine if anything at all exists that would satisfy your needs. It is FAR better to use an existing instrument in a novel way, with a new population than to develop your own questionnaire.

Also seek out expert advice and guidance. It is not silly to ask questions about questionnaires. There are psycho-metricians who make their living designing valid and reliable questionnaires. Sampling is of critical importance and bias is another concern in the design and use of questionnaires.

Mail surveys are inexpensive unless you include cash compensation for the person's time. You should always include a pre-paid return envelope. You will need to get consent procedures approved by an IRB.

Limitations: response rates are notoriously low. Ensuring understandability and readability is critical and so you must pre-test the questionnaire – any questionnaire on a group representative of the population of interest.

Face-to-Face Surveys

You can conduct a survey with a group. For example you can distribute a survey at a clinic and provide a box for deposit of the completed survey and still provide cash compensation. Be careful using student groups. This is a protected class and requires IRB approval with specific precautions for confidentiality. It is also a biased group.

You can conduct a personal one-on-one interview which is time consuming but inexpensive. You can use the telephone for this or recruit subjects – you can purchase lists of names and addresses.

Chapter II
GUIDE FOR HUMAN SUBJECTS SECTION OF THE RESEARCH PROPOSAL

SECTION 1
INTRODUCTION

This Guide for the Human Subjects Section of the Research Proposal is based on the NIH guide for PHS398, the packet for applications for research funding of clinical trials. This format also applies to educational research if information is requested of or taken from students' files or face-to-face interviews or mailed surveys.

The Human Subjects' section of research can be intimidating. It can seem like a lot of redundant work, and just a lot of work plain and simple, for any research.

However, think of yourself as a "research subject" and consider how you would like to have your information protected.

Your own IRB may have a specific outline they prefer. There is not a single standard but there are common principles. These principles can be found in Title 45 of the Public Welfare act, part 46 revised periodically. The website at NIH will help you. The Food and Drug Administration has responsibility for safety in pharmaceutical and device studies.

http://irb.jhmi.edu/Guidelines/FDAvsOHRP.html

At the sites below you will find guidance on criteria for levels of review, exempt, expedited, or full board review. You will find important definitions and criteria by which to determine risk levels and how to address protections for special populations.

You will need to complete the human subjects' research training required by your institution before your proposal can be reviewed.

http://www.hhs.gov/ohrp/

http://grants2.nih.gov/grants/peer/tree_glossary.pdf

This site will give you guidance about the Health Insurance Portability and Accountability Act (HIPPA) of 1996 and addresses how researchers may be affected by the Privacy Rule in CFR Title 45 sPart 160, and Subparts A and E of Part 164.

http://privacyruleandresearch.nih.gov/pr_02.asp

The full text of the Privacy Rule can be found at the HIPAA Privacy Web site of the Office for Civil Rights (OCR):

http://www.hhs.gov/ocr/hipaa.

SECTION 2.A.
THE OUTLINE

The following sections are the standard components that must be addressed for research using protected health information.

E. HUMAN SUBJECTS

E.1. PROTECTION OF HUMAN SUBJECTS

E.1.1 RISKS TO THE SUBJECTS

E.1.1.1 Human Subjects Involvement and Characteristics

E.1.1.2. Sources Of Materials

E.1.1.3. Potential Risks

E.1.2. ADEQUACY OF PROTECTION AGAINST RISKS

E.1.2.1. Recruitment And Informed Consent

E.1.2.2. Protection Against Risk

E.1.3. POTENTIAL BENEFITS OF THE PROPOSED RESEARCH TO THE SUBJECTS AND OTHERS

E.1.3.1. Potential Benefits of Research to the Subjects and Others

E1.3.2. Risk Benefit Analysis

E.2. COMPOSITION OF THE PROPOSED STUDY POPULATION

E.2.1. INCLUSION OF WOMEN AND MINORITIES

E.2.2. INCLUSION OF CHILDREN

E.3. DATA AND SAFETY MONITORING PLAN (This is used only in clinical trials but you should be familiar with the terminology and purpose of such a plan and a Data and Safety Monitoring Board)

F. VERTEBRATE ANIMALS

E. HUMAN SUBJECTS

E.1. PROTECTION OF HUMAN SUBJECTS

E.1.1 RISKS TO THE SUBJECTS

E.1.1.1 Human Subjects Involvement and Characteristics

Example

Men and women between the ages of 20 to 65 years for a total of 144 subjects recruited to protect against a 10% potential attrition rate will be eligible for the study. Women and minorities will be included as part of natural referral population – see the Targeted/Planned Enrollment Table.

Women and minorities will be included in this study as shown on the Targeted/Planned Enrollment Table (page 50). Pregnant women will not be included because of the multi-factorial nature of pregnancy in relationship to the etiology of CTS. CTS is often a transient condition during pregnancy.

Subjects for this study must meet three levels of screening.

Level One Screening

If the subjects have any of the following conditions they will be excluded from the study.

Any medical conditions that are plausible secondary causes of CTS, including but not limited to

- *severe CTS that has progressed to muscle atrophy;*
- *pregnancy;*
- *previous wrist surgery on the wrist to be studied;*
- *a systemic disease or condition including but not limited to diabetes mellitus, thyroid disorders, rheumatoid arthritis, Paget's bone disease, gout, myxedema, multiple myeloma, acromegaly, hepatic disease, dialysis patients, or other diseases or conditions in which peripheral neuropathies are common.*

Level Two Criteria:

Each level one qualified and consented subject will be given a baseline NCS. For the purpose of this study, the electrodiagnostic values utilized for inclusion criteria are from Kilmer and Davis.[15] Often, electrodiagnostic findings do not correlate with patients' reported symptoms. Therefore, for the purposes of this exploratory study, we will use a combined screening method that includes primarily electrodiagnostic value with a summed cut-point on the SSS and FSS for inclusion of subjects at this level. The summed symptoms and function cut-point will be a combined raw score of 36 or a mean score of 4.0 (ideally at least 2.30 on the SSS and 1.70 on the FSS). These correspond to the low range of scores found by Levine et al. in the seminal study for these questionnaires.

For the NCS portion of the screening, each subject must have NCS values with at least one of the following:

- *Median motor nerve distal latency greater than 4.2 ms;*
- *A difference between ipsilateral Median and Ulnar motor nerve distal latency greater than 1.5 ms;*
- *Median nerve sensory peak latency greater than 2.2 ms;*
- *A difference between Median and Ulnar sensory peak distal latency greater than 0.3 ms[15]*

Level Three Criteria:

Subjects who qualify under Level Two Criteria screening will be given an MRI exam of the wrist being studied.

Subjects with any of the following conditions will be excluded from the MRI portion of the study:

- *cardiac pacemaker;*
- *hip or knee prosthesis;*
- *any metallic foreign body in the immediate vicinity of affected wrist(s).*

Example Continued…

It is anticipated that the incidence of these conditions in the study population is so low that it should not affect recruitment or threaten the power of the analysis by impacting the number of subjects in each group. For the MRI portion of the study, these subjects may be randomized to the group that does not get the MRI measurement.

If a MRI shows a plausible secondary cause of CTS such as a ganglion cyst, mass, or an accessory muscle, that subject's case will be evaluated by the PI and the consulting radiologist for referral, retention, or exclusion from the study.

Subjects will be equally and randomly assigned to four treatment groups.

E.1.1.2. SOURCES OF MATERIALS

Example

The proposed project will obtain demographic information directly from the subject without using previously existing medical records. Subjects will be interviewed to obtain self-reported medical history and physical information. Subjects will be asked to provide information about the number of times they performed home stretching, took medications, what type of medications, and if they received or did anything outside of the trial to address their CTS symptoms.

We will collect information from non-invasive medical tests (no needles will be used, no blood or other body fluids will be collected). No x-rays or laboratory tests will be used. No existing specimens, other previously existing records, medical charts, or other data will be used. Data collected on project participants will be used only for research purposes.

We will use neurodiagnostic tests using a The NeuroMax 1002, a portable 2-channel EMG system for diagnostic electromyography studies.

We will perform Magnetic Resonance Imaging in a Hitachi Altaire high-field performance, open design that combines a 0.7T field strength and advanced magnet technology.

We will take measurements of grip and pinch strength. Grip strength will be measured on a Jamar Dynamometer which measures up to 200lbs (91kg). The pinch gauge will have a range from 0 to 30lbs (14 kg). Measures of sensory nerve conduction will be taken using a Neurometer® CPT by Neurotron, Inc.

The subjective measures will be the Levine Symptom Severity Scale and the Functional Status Scale, paper and pencil tests that the subject completes in the presence of the research coordinator or other designated key personnel blinded to the group assignment.

E.1.1.3. Potential Risks

Example

The proposed study carries minimal risk to the participant. Any risks are not significantly increased over that ordinarily encountered in daily life for a person with carpal tunnel syndrome who might be treated with physical therapy. Potential for physical and psychological injury from the OMM and sham treatments is minimal. HVLA will be used and subjects are informed that there is a very slight chance of fracture of a bone. There should be little to no legal, social, or other risk due to loss of confidentiality because all records will be kept under lock and key, and will be processed in either a coded or generalized manner to fully protect anonymity. Additionally, electronic storing and transmission of data will be performed in a secure manner with no direct identifying information included.

The osteopathic intervention being used in this study is an accepted standard practice. Therefore there are no notable difficulties anticipated. For this same reason, there are no alternative procedures/therapies warranted other than non-participation.

If the MRI indicates a clinical condition previously undetected, that may need the attention of a physician, the PI and the consulting radiologist will determine if the subject should be offered a referral. Certain exclusions for the trial are described in Section B.

Participants may feel some discomfort during the nerve conduction studies due to mild electrical pulse stimulation of the median and ulnar nerves. The lowest intensity will be used to get the maximal reading. There are no major risk factors associated with nerve conduction studies.

Example continued

There are no known risks or side effects to MRI and those subjects who meet the exclusion criteria for the MRI protocol will be excluded from this portion of the study. Subjects will need to remain still for short periods (minutes) while images are being taken. The images will be taken on an open MRI thus reducing discomfort from enclosed spaces.

Subjects will be asked not to receive OMT or other manual therapies during the eight months of the trial. Should the subject decided to pursue a surgical treatment, the subject will be removed from the study.

Information from the study not revealed to the patient – if conditions are identified that might, in the opinion of the OMM specialist or the radiologist reading the MRI, risk or compromise the health of the subject the physician team will confer and determine if the subject should be referred for treatment.

E.1.2. ADEQUACY OF PROTECTION AGAINST RISKS

E.1.2.1. Recruitment and Informed Consent

Example

Subjects will be recruited from several sources, using several recruiting methods. One source will be the Internal Medicine and Family Medicine clinics at the University of North Texas Health Science Center (UNTHSC)/Texas College of Osteopathic Medicine (TCOM). The Internal Medicine and Family Medicine clinics at UNTHSC/TCOM had approximately 170 patients with CTS during the past year (April 2002-March 2003). Other medical referral sources will include the Veterans' Medical Center and John Peter Smith Health Network.

In addition to these sources, referrals will be recruited by newspaper, internet and posted flyers. A special effort will be made to reach the Asian community to ensure inclusion of Caucasians, Blacks, Hispanics, and Asians in the study. No difficulty is anticipated with recruiting the required numbers of subjects over the three-year study period.

Informed consent will be obtained from the subjects by the clinical research coordinator (CRC) or if necessary, a designated alternate research team member. The CRC will inform subjects of the nature of the study, inclusion/exclusion requirements, their right to withdraw at any time without affecting their medical care.

E.1.2.2. PROTECTION AGAINST RISK

Example

To the best of our medical knowledge the risks for adverse effects from any of the non-invasive procedure in this study is minimal. Persons with certain medical conditions are excluded from the study. Persons with certain conditions are excluded from the MRI.

Participants may feel some discomfort during the nerve conduction studies due to mild electrical pulse stimulation of the median and ulnar nerves. The lowest intensity will be used to get the maximal reading. There are no major risk factors associated with nerve conduction studies.

If the MRI indicates a clinical condition previously undetected, that may need the attention of a physician, the PI and the consulting radiologist will determine if the subject should be offered a referral. There are no known risks or side effects to MRI and those subjects who meet the exclusion criteria for the MRI protocol will be excluded from this portion of the study.

Trained and licensed osteopathic physicians will conduct all intervention protocols in order to guarantee an absolute minimal possibility for injury or other adverse reaction. Should adverse events occur, the proper protocol for reporting these adverse events to the Institutional Review Board will be immediately implemented. This protocol is outlined in the UNTHSC Institutional Review Board Investigator's Manual.

This application does not include activities involving human products or unidentifiable patient data. No test article (investigational new drug, device or biologic) will be utilized in this study

All records and medical information will be kept as confidential as possible under current local, state, and federal law. All subjects' study charts and all data collected in the study will be maintained and protected in their respective place of collection, the OMM clinic and the offices of the Osteopathic Research Center at UNTHSC in the office of the Research Coordinator in a locked filing cabinet. These files will only be accessible to the Key Personnel of this study, the UNTHSC Institutional Review Board, and regulatory agencies. The databases will contain no personal identifiers. All subjects will be given an

identification number for the database. All reports and potential publications will report aggregate information only with no personal identifiers.

If any individual is injured or suffers adverse effects while participating in this study, s/he will be immediately referred to the most appropriate health care resource. Trained and licensed osteopathic physicians will conduct all intervention protocols in order to guarantee an absolute minimal possibility for injury or other adverse reaction. All other data gathering procedures will be performed by either trained health care professionals or trained project personnel who are either certified or deemed fully competent in the particular area in which they are functioning.

The potential study subjects will be informed of the potential risks and benefits of the proposed research and clearly informed that their decision or refusal to participate will in no way affect their ability to receive the standard best available medical care.

Participation in this clinical trial is completely voluntary and the patient is under no obligations by this study. At any point during the study, any subject may withdraw without consequence. If a subject experiences unexpected side-effects secondary to the protocol or no longer meet the inclusion/exclusion criteria with the exception of NCS, they will be discontinued from the study. If at any point a subject is excluded from the study, the reasons for discontinuation will be noted for analysis as a possible complication variable.

E.1.3. POTENTIAL BENEFITS OF THE PROPOSED RESEARCH TO THE SUBJECTS AND OTHERS

E.1.3.1. Potential Benefits of Research to the Subjects and Others

Example

There are limited benefits to the subjects who participate in this study. All subjects will continue under the supervision of their primary care physician. Participants in the OMM groups will be asked not to have any additional OMT or manual therapies during the six weeks of treatment and the six months period before the last NCS and MRI, and study data collection visit. The subjects in the OMT and OMT and stretching groups could potentially benefit from these interventions as an adjunctive modality to measurably alter the usual clinical course of the recovery; although this benefit is what is being tested in this research. Information about the results of tests will not be given to the subjects during the study, but may be available to them after the study.

The potential benefits for patients in this study include: (1) possible increase in functional status of upper extremity with carpal tunnel syndrome; (2) possible decrease in symptoms associated with carpal tunnel syndrome; (3) possible improvement of carpal tunnel syndrome through a conservative, non-surgical treatment; (4) possible improvement in strength of hand effected by carpal tunnel syndrome; (5) MRI of wrist area which is not offered as normal standard care.

E1.3.2. Risk Benefit Analysis

Example

The far-reaching benefit would be providing an effective non-surgical alternative treatment for Carpal Tunnel Syndrome in order to reduce medical costs for the patients. The potential for showing the efficacy osteopathic manipulative treatment for carpal tunnel syndrome could potentially reduce the need for invasive surgical procedures. This in turn would affect the current economic burden in treating carpal tunnel syndrome. To the best of our medical knowledge the risks for adverse effects from any of the non-invasive procedures in this study is minimal and therefore the benefits outweigh the risks. The potential for showing the efficacy osteopathic manipulative treatment for carpal tunnel syndrome could potentially reduce the need for invasive surgical procedures and change the current economic burden in treating carpal tunnel syndrome.

Trained and licensed osteopathic physicians will conduct all intervention protocols in order to guarantee an absolute minimal possibility for injury or other adverse reaction. Should adverse events occur, the proper protocol for reporting these adverse events to the Institutional Review Board will be immediately undertaken. This protocol is outlined in the UNTHSC Institutional Review Board Investigator's Manual.

E.2. COMPOSITION OF THE PROPOSED STUDY POPULATION

E.2.1. INCLUSION OF WOMEN AND MINORITIES

Example

Women and minorities will be included in the study. CTS occurs more frequently in females than in males, and more in white populations than in non-white. However, based on the ethnic demographics of the Tarrant County, (Fort Worth) Texas area we will make a specific effort to include non-white Hispanic and Black populations. We will do this with Spanish language ads and flyers in places including but not limited to grocery stores, unemployment/employment offices, city health department clinics, physicians' offices, and directly contacting physicians who serve more of the Hispanic population at other UNTHSC family practice clinics. There are too few Asian and Native American or Pacific Islander residents to include in adequate numbers in this pilot study. Page 50 contains the Targeted/Planned Enrollment Table.

E.2.2. INCLUSION OF CHILDREN

Example

Children under age 18 and adults between 19 and 21 will not be included in the study. Statistically CTS is rare in persons under the age of 20. The purpose of this study is to learn more about adults with a likelihood of having CTS to lead to a larger trial with groups that are associated with particular occupations.

E.3. DATA AND SAFETY MONITORING PLAN

Example

The proposed clinical trial carries minimal risk potential to the subjects. Data and safety monitoring will be conducted by two methods. First the investigators will perform continuous, close monitoring activities to ensure to the extent possible, the safety of subjects involved in this research study. Adverse reactions or side effects from the protocol treatments will be evaluated and recorded daily by the study protocol treatment provider. This information will also be assessed daily by the site PI or Co-investigator. Any adverse events will be reported as soon as possible to the Institutional Review Board (IRB) Chair.

A Data Safety and Monitoring Board will be established by the PI in consultation with the epidemiologist and the radiologist. The board will consist of at least three members with the PI and the Co-I as ad hoc members. The board will include an ethicist, a specialist in carpal tunnel syndrome (a neurologist) and a statistician.

F. VERTEBRATE ANIMALS

Not applicable.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Chapter III
REVIEWING THE RESEARCH LITERATURE

SECTION 1
INTRODUCTION

What is the purpose of the Background and Significance section of a research proposal and how and why do I review the literature.

The purpose of the Background and Significance section is to describe the scientific knowledge, research recommendations, and the identified gaps in the scientific knowledge as communicated by published peer-reviewed or textbook materials relevant to the topic of interest.

This section provides the rationale for the proposed research and summarizes the expected contribution your research will make to the field.

In this section you will:

- Make a compelling case for your proposed research project. Why is the topic important? Why are the specific research questions important? How are the researchers qualified to address these?
- Establish that you are knowledgeable in the topic.
- Avoid outdated research.
- Use citations not only as support for specific statements but also to establish familiarity with all of the relevant publications and points of view, not just those that support your hypothesis.
- Make sure the citations are specifically related to the proposed research. Cite and paraphrase correctly and constructively.
- Highlight why research findings are important beyond the confines of a specific project, i.e. how the results might be applied to further research in this field or related areas.
- Stress any innovations in experimental methods (e.g., new strategies, research methods used, interventions proposed.

The Art of the Literature Search

The literature search is the process of finding research reports related to your research topic. Much information can be gained by doing this preliminary work and there are various ways to perform it. This section reviews some important reasons for conducting a literature search and four possible approaches to a literature search.

Why Perform a Literature Search?

1. To review the existing published research on the topic
2. Discover how and why the topic has been studied
3. Determine what has changed between the first and most recent published information
4. Focus your research by reviewing the methods and definitions used in prior studies
5. To learn the names of the prominent researchers in the area of interest
6. To help shape study design by learning from and improving on prior studies
7. Identify sources of valid, reliable measurements and procedures and invalid and irrelevant measures as well.
8. Determine if the study is important and useful

What is a review of the literature?

Quotes excerpted from http://www.utoronto.ca/writing/litrev.html

"A literature review is an account of what has been published on a topic by accredited scholars and researchers. The purpose is to convey to your reader what knowledge and ideas have been established on a topic, and what their strengths and weaknesses are. As a piece of writing, the literature review must be defined by a guiding concept "

Skills You Need

- Ability to scan the literature efficiently, using manual or computerized methods, to identify the materials you will use and cite
- Ability to apply principles of analysis and critical thinking to identify valid, reliable research reports and identify and describe the gaps in the literature pertaining to your topic
- Organize your material around and related directly to your topic
- Synthesize results into a summary of what is and is not known
- Cogently discuss areas of controversy in the literature
- Formulate questions that need further research

Section 2
CONDUCTING A LITERATURE REVIEW

Checklist of Questions To Ask Yourself

- ☐ What key words have I found in the first three articles I have read?
- ☐ If these are the most recent articles on the topic, what references have they used?
- ☐ Do I need to include theory?
- ☐ Are there any existing "reviews of the literature" on my topic?
- ☐ Do I need to include Text Books in addition to published peer review articles?
- ☐ What unpublished material may help my review?
- ☐ Have I checked the literature cited by authors to ensure I have covered a broad scope? (when you begin to see citations repeated in the articles you review, you may have exhausted the field.
- ☐ I have over 100 articles, how can I address this large scope?
- ☐ Have I used a written list of the specific outcome measures or population characteristics, or other criteria to cull out articles that are extraneous?
- ☐ Do I have a checklist to assess strengths and weaknesses of the materials?
- ☐ Have I cited and discussed studies contrary to my perspective?
- ☐ Will the reader find my literature review relevant, appropriate, and useful?

QUESTIONS TO USE IN REVIEWING ARTICLES AND TEXTS

- ☐ Has the author formulated a problem/issue?
- ☐ Is it clearly defined? Is its significance (scope, severity, relevance) clearly established?
- ☐ Could the problem have been approached more effectively from another perspective?
- ☐ What is the author's research orientation (e.g., interpretive, critical science, combination)?
- ☐ What is the author's theoretical framework (e.g., psychological, developmental, feminist)?
- ☐ What is the relationship between the theoretical and research perspectives?
- ☐ Has the author evaluated the literature relevant to the problem/issue?
- ☐ Does the author include literature taking positions she or he does not agree with?
- ☐ Does the author provide complete information or how to access more information on the basic components of the study design
- ☐ Are the measurements valid and reliable for this topic?
- ☐ Is the analysis of the data accurate and relevant to the research question?
- ☐ Are the conclusions validly based upon the data and analysis?
- ☐ Is there scientific evidence underlying the reasoning, or is the author promoting a perspective that is theoretical only?
- ☐ How does the author structure the argument? Can you "deconstruct" the flow of the argument to see whether or where it breaks down logically?
- ☐ In what ways does this book or article
  - o contribute to our understanding of the problem under study,
  - o provide information for practice?
- ☐ What are the strengths and limitations?
- ☐ How does this book or article relate to the specific thesis or question I am developing?

How to Conduct a Literature Search

A well-organized, goal-oriented approach to a literature search will save a lot of time in the library. Draft a literature search outline with the questions pertinent to your research. Faculty members, biostatisticians, and librarians can be of great assistance in this area. There are many different ways to conduct a literature search. A combination of the following four approaches can be used to avoid potential limitations in the search.

Ancestry approach

Begin with most recent reference and trace its bibliography

Pros: describes evolution of the idea or technique, places study in an historical context

Cons: unveils many older references vs. more recent references

Descendency approach

Locate the classical reference and identify each subsequently published study that refers to the classical reference (via Science Citation Index or Social Science Citation Index)

Pros: identifies a large number of articles in sometimes unlikely journals

Cons: unveils articles of peripheral issues, may bypass opposing arguments in search

Database search (e.g. MEDLINE)

Databases cross-indexed by topic, content, author, year of publication

Pros: intelligent control of study with manageable number of references

Cons: out-of-date compared to original references

Network

Contact an author directly

Contact an association to learn of contacts

Pros: primary source, state-of-the-art information, may provide additional contacts, may identify students willing to assist with the study

Cons: can be restricted to a single point of view, information may not have been subjected to peer review

EXAMPLE of a literature search using online resources can be modified to yield any, or a combination of any of the four basic approaches to literature searches.

Enter Keyword/s or author

Refine as needed in combinations or single key word or author searches

Limit your choices – rigorously and focus on the latest research

Consider

a) whether the results cover topics such as the fundamental principles or philosophy on which your question is based and

b) whether these articles will likely include citations to earlier work that you will want to consider.

Then read the abstracts to determine if you want the entire article (pdf preferable but not always possible in osteopathic medicine research)

Refine your keywords if necessary

Select whole articles

READ, don't just "review" articles; marking them and making notes on them

Use a three ring binder to organize the materials

Review references cited by the authors you have selected

Trace back to the seminal research in the area

Include foreign articles – you can get translations if you need any – European osteopathic research is very useful – refer to Dr. Michael Kuchera as a member of the Federation Internationale de Medicine Manuel

Acquire additional articles from this process

INCLUDE those that you do not agree with – and from other professions – and commentaries as needed to elucidate the relevant debate

Include reviews of the literature and meta-analysis articles

Do you have another strategy? – write it here:

Final Notes

There is value in the old “index card” system of keeping notes on articles or use excel files so you can search later by key words from your own literature data base.

Example

Topic
Author
Title
Journal, year, volume, and pages
Aim of article/research
Primary question/hypotheses
Results
Methods
Future research questions recommended
Comments: unique features, stimulating facts

A literature review is a piece of **discursive prose**, not a list describing or summarizing one piece of literature after another. It's usually a bad sign to see every paragraph beginning with the name of a researcher. Organize the literature review into sections that present themes or identify trends, including relevant theory. You are not trying to list all the material published, but to synthesize and evaluate it according to the guiding concept of your thesis or research question.

If you are writing an **annotated bibliography**, you may need to summarize each item briefly, but should still follow through themes and concepts and do some critical assessment of material. Use an overall introduction and conclusion to state the scope of your coverage and to formulate the question, problem, or concept your chosen material illuminates. Attempt to group the literature under important medical or clinical sub-sections. This process will help you indicate comparisons and relationships.

Appendix 1 for Chapter III
GUIDE TO PRESENTING A RESEARCH REPORT/JOURNAL ARTICLE

(This material is excerpted from the OMM Pre-Doctoral Clinical Research and Education Master of Science Degree program at the University of North Texas Health Science Center developed by Dr. des Anges Cruser former Graduate Advisor and Fellowship Director)

Purpose

A. The purpose of reviewing, presenting and discussing a research article is to better understand the role of science and research in evidence-based medicine. The Fellow will lead a discussion of a full length research article associated with Osteopathic Manipulative Medicine using generally accepted criteria for discussion of a research article.

B. This presentation enables all who are involved to improve their knowledge of OMM research and achieve a better understanding of the science and research design and methods involved in the study.

Article Selection Criteria

It is preferable that the article be a research article. If it is not a research article per-se, but is approved by the student's mentor or preceptor, the discussion should center on the research aspects of the topic and how this article helps the profession regarding medical education and clinical research or patient care.

1) A research article reports findings from an original scientific or clinical study.
2) A research article may also be one that reviews other clinical or basic science research in the OMM field, such as a focused review of the literature or a meta-analysis of published research data.
3) The research may be in an area of interest of the fellow, or one about which s/he wishes to learn more in terms of the science of the technique, the underlying physiological or biochemical principles of the disease response to OMM, or its relevance to medical education or clinical practice.

Role of the student or resident in selecting, obtaining approval for, and preparing the presentation of the research article

- Two weeks before the scheduled event: The presenter will personally provide a copy of the article to the student's selected mentor/instructor (in person) so the faculty member can review the article and the presenter still have time to find another should that one not be suitable.
- The faculty member is expected to respond within 24 hours to the request for approval.
- Ensure that the faculty member who approved the article or another one who is knowledgeable in research design assists in developing the presentation.
- Provide copies to all who will attend the session
- Notify the Academic Coordinator to announce the presentation at least one week prior.

Role of the Faculty in the Journal Article Presentation

The faculty selected by the Fellow will:

1) Approve the article or assist in a different selection within 24 hours of receiving the article personally from the student or resident
2) The article is to be a research article – OR – if it is not research per-se, the discussion should center on the relationship between the subject, and purpose and conclusions of the article in relationship to future research, teaching, or patient care.
3) Read the article
4) Assist the student in the critique
5) Attend that research presentation event
6) Summarize the discussion at the end of the journal club session – or assist the student or resident to summarize the discussion.

Five Points to Cover in the Presentation and Discussion of the Article

1) Why you selected this article

   For example: a) Findings contradict existing knowledge in this area, b) Article is seminal, c) Study has a unique design, d) First prospective randomized study in that area, e) area of interest to me
2) State a) the hypothesis or main question and b) the findings or results.

   For example: a) Does the Spencer Technique, given specific circumstances, increase ROM in elderly patients post-surgery? b) The group who received this treatment increased ROM in 50% less time than those without it.
3) Is the study methodologically sound and how do you know or what tells you that it is or is not?
4) What did this study contribute to the field of OMM and/or the Osteopathic medical profession?
5) Where and how can this study be taken further?

Please do not **ONLY** answer these questions.

Please think about the research.

If something seems to be missing present this in the form of a question to the group: 'I think XYZ is missing from this article because……. What do you think?'

JOURNAL CLUB PRESENTATION

1) Why I selected this article (e.g. Findings contradict existing knowledge in this area, Article is seminal, Study has a unique design, First prospective randomized study in that area)

2) Hypothesis and findings:

A. The research hypothesis or main question of this article was:

B. The findings or results of this research were:

3) Is the study methodologically sound and how do I know?

4) What did this study contribute to the field of OMM and/or the Osteopathic medical profession?

5) Where and how can this study be taken further?

Fellow's Name_____ Date: _____

Faculty Advisor_____ Date_____

dAC: TCOM 3-27-03 and revised 7-1-04

SECTION 1
INTRODUCTION

Research Design and Methods

Now you are actually talking about what you are going to do to conduct this research.

This is the section you drew in the worksheets after Chapter I.

When you write this section, you will begin to understand that you may need to revise your background and significance or that you want to "tweak" your question. This section should flow naturally from the specific aims.

Keep it simple.

How will you answer your question? A clearly organized Method section is critical. Many researchers are confused about the distinction between design and methods, leading them to lump the two together. This is not just an issue of semantics. This is the DESIGN and METHODS section. In fact the NIH changed the name of this section to include the word design because many proposals omitted this sub-section.

Present the design and methods sub-sections separately.

Design is the way in which you conceptualize your research project.

Methods are detailed tasks, steps, stages, and procedures you will use to conduct the research.

The design discussion is relatively brief, particularly when compared to the methods narrative. However, it should be inherently interesting. It can be very creative. For example, you could have designed your project in any number of ways, but you chose this one. Why?

Methods, by contrast, are straightforward. It is critical to describe exactly what you will do to conduct the study from beginning to end. Provide sufficient detail for a complete evaluation of your work. This is particularly important if you are proposing to develop a new methodology or a new technique.

In the methods section you will include a discussion of the strengths and limitations of your study. Discuss any technical problems that may arise and what alternate plans you may implement. This is a sophisticated part of the proposal, that should be reviewed carefully with a mentor.

End the research design and methods section with a timeline. It is important to convey that your project can be conducted within the proposed time and that you have a logical, well conceived plan to implement.

There is no one standard way to organize the Research Design and Methods section.

Recommended Outline

D. Research Design and Methods
- D.1. Introduction
  - D.1.1. Restate the research aims and or question/s or hypotheses
  - D.1.2. Describe how this section is organized
- D.2. Power Analysis
- D.3. Inclusion and Exclusion/Selection Criteria
- D.4. Outcome Measures – variables of interest
- D.5. Data management and analysis
- D.6. Project management plan and time lines

SECTION 2
ANNOTATED OUTLINE

Tasks Associated With Research Design and Methods Section

Introduction

The reader has just finished either the background and significance in which you made your case that this project is supported by the literature and/or fills a gap in the literature, or the preliminary studies section in which you discuss your or your mentor's previous work or evidence of the ability to conduct this study.

Begin the Research Design and Methods section with something like:

"The purpose of this study is to . . .."

Example

The overall goal for this proposed exploratory/developmental phase II clinical trial is to build a scientific body of knowledge concerning the impact of osteopathic manipulative treatment (OMT) on CTS and establish a plausible model to test in larger more definitive studies. The ultimate goal is to evaluate the efficacy of a conservative, biomechanical, non-surgical treatment for CTS. This study will use a prospective, randomized, controlled, blinded design with a three-month follow-up to examine whether OMT with and without stretching exercise provides a more effective treatment for CTS than non-surgical standard care, or a sub-therapeutic ultrasound placebo.

Then tell the reader how you have organized the material.

Example

The following material includes first a discussion of the research project design, next a discussion of the power of the study and the inclusion/exclusion (selection) criteria. Next we provide information about the outcome variables of interest and how we propose to manage and analyze the data to address the study questions/hypotheses. Last, we provide a project management plan and time-lines for completing the project.

Key Questions to guide your thinking are as follow.

What is the scientific or clinical theoretical framework for this research?

What do we currently know about this area?

What are the strengths of the current body of knowledge?

What are the gaps in the current body of knowledge?

How will this proposed study build on the strengths, correct the limitations and fill the gaps (or begin to do this)?

Design

Example 1

Epidemiology is the study of populations and the distribution of and determinants of health and diseases, morbidity and mortality in those populations.[1] Clinical epidemiology uses the concepts of epidemiology and applies them to clinical practice.[2] Clinical epidemiology is defined as "the science of making predictions about individual patients by counting clinical events in similar patients, using strong scientific methods for studies of groups of patients to ensure that the predictions are accurate."[2] Epidemiology has an important place in medicine, providing methods that help clinicians and researchers collect information to build foundational knowledge. Ultimately, this allows physicians to better diagnose and treat individual patients by learning from a population. Epidemiologic studies are greatly needed in the osteopathic profession to strengthen the foundation of knowledge that future studies must build upon. The diagnosis and treatment of somatic dysfunction, or impaired function of the body framework, is at the core of osteopathic medicine. However, more data is needed on the incidence and prevalence of somatic dysfunction in various populations. In addition, the relationship between somatic dysfunction and chronic disease in various populations needs to be studied. Older adults over the age of 65 years have not been the focus of many studies in the osteopathic literature.

The purpose of this study is to utilize epidemiologic methods to explore the relationship between somatic dysfunction and chronic conditions in a population of older adults.

Example 2

OMT and OMT plus stretching may favorably impact CTS by several mechanisms including 1) mechanical expansion of the carpal tunnel resulting in decreased mechanical compression of the Median nerve, 2) decreased tension within upper extremity fascial strain patterns that may be compressing low pressure lymphatic and venous vessels and contributing to fluid congestion and edema within the carpal tunnel, and 3) decreasing minor upper extremity neural impingements at the nerve roots, brachial plexus, Median or Ulnar nerves that may be causative of other upper extremity symptomatology and may be contributory to the "Double Crush Syndrome" of the Median nerve at the wrist.

Based on this understanding of the mechanisms of OMT and stretching exercise on CTS, six hypotheses will guide this study. Each hypothesis is built around an important outcome measure.

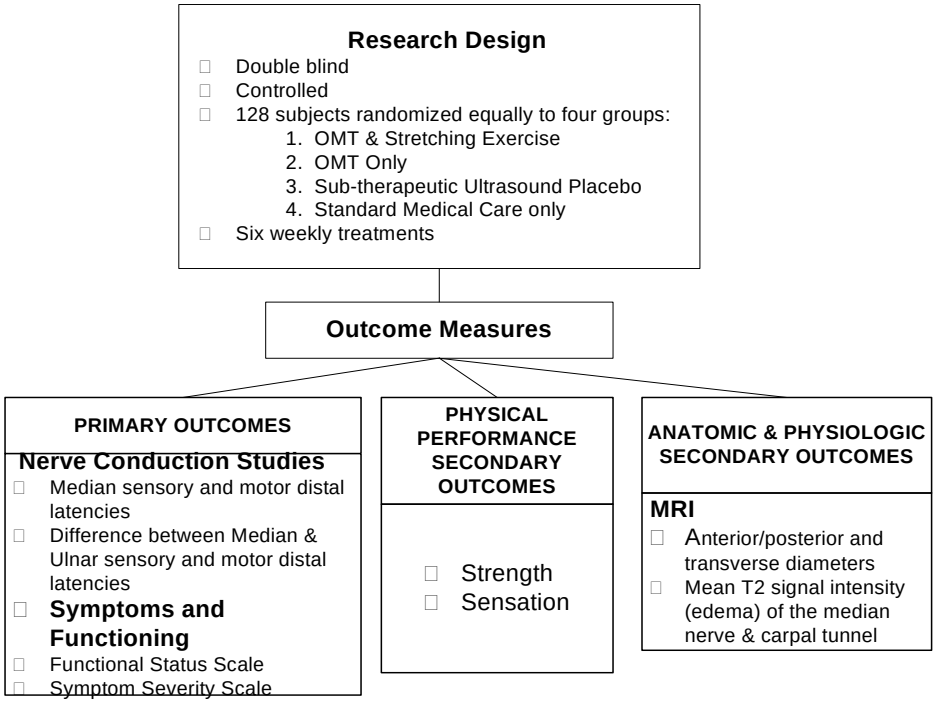


Figure 3. *Research Design*

Power Analysis

After the research question/s or hypotheses are final you should determine what result/s you anticipate. Typically, the smaller the sample size, the larger any difference between group scores will have be in order to achieve statistical significance.

Statistical power analysis is a set of procedures and formulas that allow us to determine how likely we would achieve statistical significance with a particular sample size (given an assumed true difference between groups).

If the likelihood is good (e.g. greater than or equal to an 80% chance), then the sample size is considered adequate.

Power Analysis Examples

You expect elderly patients with chronic illnesses to have a certain amount of somatic dysfunction that is greater by about 25% compared to those with no chronic illnesses.

You can use data from the literature or a theoretical clinical foundation to estimate the anticipated effect size for your study. For example, if the literature reports a 10% decrease in pain in carpal tunnel syndrome patients after two treatments of OMT, and the nerve conduction values are not statistically different in the subjects in that study at baseline, you can use this data to estimate your anticipated effect size.

You need to have as much valid and reliable data as possible for the biostatistician to estimate a sample size for your study at a specified power. This applies to ANY design. This applies if you are taking all your data from medical records. This applies if you are using existing data from another source.

Do not try to do this by yourself (at home). Use an expert. Do not lick your finger and hold it up to the wind or put your thumb up to the portrait or MRI and say “oh, about a hundred should do it” or “Seems like 25 would be okay”.

Regardless of whether you intend to adhere to the sample size indicated by a power analysis, it is <u>always</u> important to include a discussion of this effort.

If you are willing to accept a small change or difference you may not need so many subjects. If you are conducting an exploratory study or a case-control study you do not need a specific sample size but you always should guard against both type one and type two errors (see glossary of terms).

<u>Example</u>

In order to allow for subgroup analyses and for stratification on the multiple sites, a total of 360 subjects, 120 in each group, will be recruited into the study. Each site will recruit 90 subjects, 30 in each group, over the course of the study. Censored values (those dying in the hospital and those being placed on a ventilator) and drop-outs are estimated to account for no more than 20 percent of the total number of subjects. Based on preliminary data on length of hospital stay, when the sample size of non-censored values in each group is 96, an exponential maximum likelihood test of equality of survival curves with an α=0.05 two-sided significance level will have power of 0.80 to detect the difference between a median length of hospital stay of 6 days versus 9 days. Therefore, recruitment of 120 subjects per group should be sufficient to detect the hypothesized differences in hospital length of stay. Figure 7 on the following page illustrates the projected enrollment timeline for the 360 subjects.

<u>Example</u>

This is a cross-sectional study using available data from a sample of patients treated in a specialty geriatric clinic at the University of North Texas Health Science Center Department of Internal Medicine. Data collection took place in 2005 over a period of approximately 6 weeks in the medical records department of the University of North Texas Health Science Center patient clinic. A list of patients seen for OMT in the geriatrics assessment and planning (GAP) clinic between January 1, 2000-May 1, 2005 was created. Both genders were included. Because this study focuses on a geriatric population in a geriatrics clinic, only persons 65 years of age and older were included.

Every available consecutive chart was selected from the active files for persons seen for OMT between January 1, 2000-May 1, 2005 until 139 charts were obtained. Data were recorded on data collection forms.

Data elements collected from medical documents in the chart and the first (or oldest available) OMT

visit included the following:

Outcome Measures

OMM research has been criticized for its use of subjective measures in clinical trials. However, an interesting question is: “What are the possible outcome measures that can be examined to determine the effects of OMM in various conditions?” The Foundations textbook chapters are replete with opportunities to research this question. One public health student (Rebecca Whitesell, MPH, now a third year medical student at the University of Texas at San Antonio) examined the literature with respect to what outcome measures have been used in OMM research since 1994. Her thesis is available from Dr. Cruser and will be published in 2008.

A meta-analysis of osteopathic clinical trials is being conducted by Dr. John Licciardone and will be available in report form later in 2007.

This manual is not long enough to address issues in outcome measures for OMM research. However, the outcome measure drives the type of data analysis you will use for your research. An outcome may be found in existing data, and is not necessarily only prospectively collected.

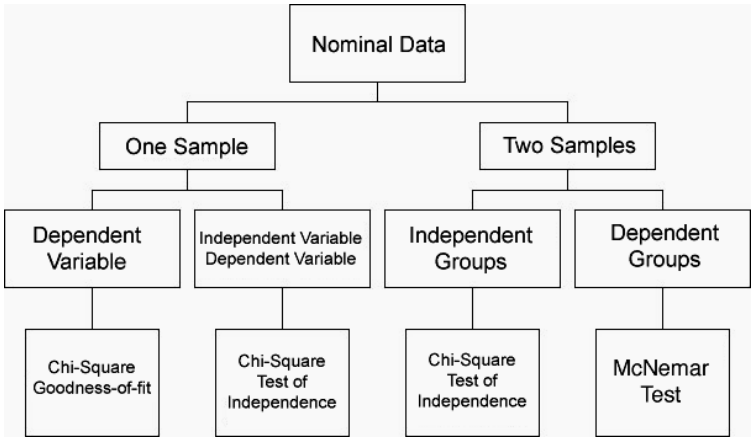
Data Management and Analysis

Reference the statistical software that will be used for the analysis.

Meet with a biostatistician who will help you THINK critically about your design and methods, not merely crunch numbers. This will require an investment of time on your and their part.

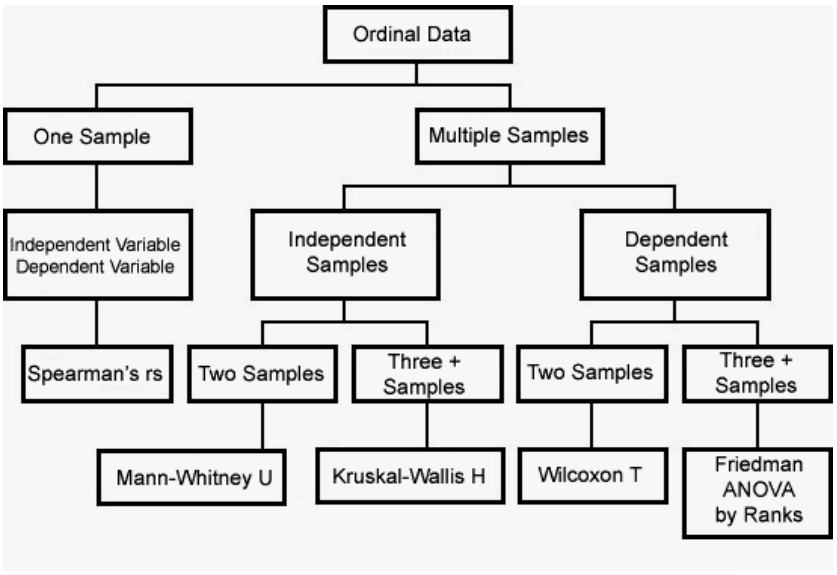
Correlational vs. Experimental Research

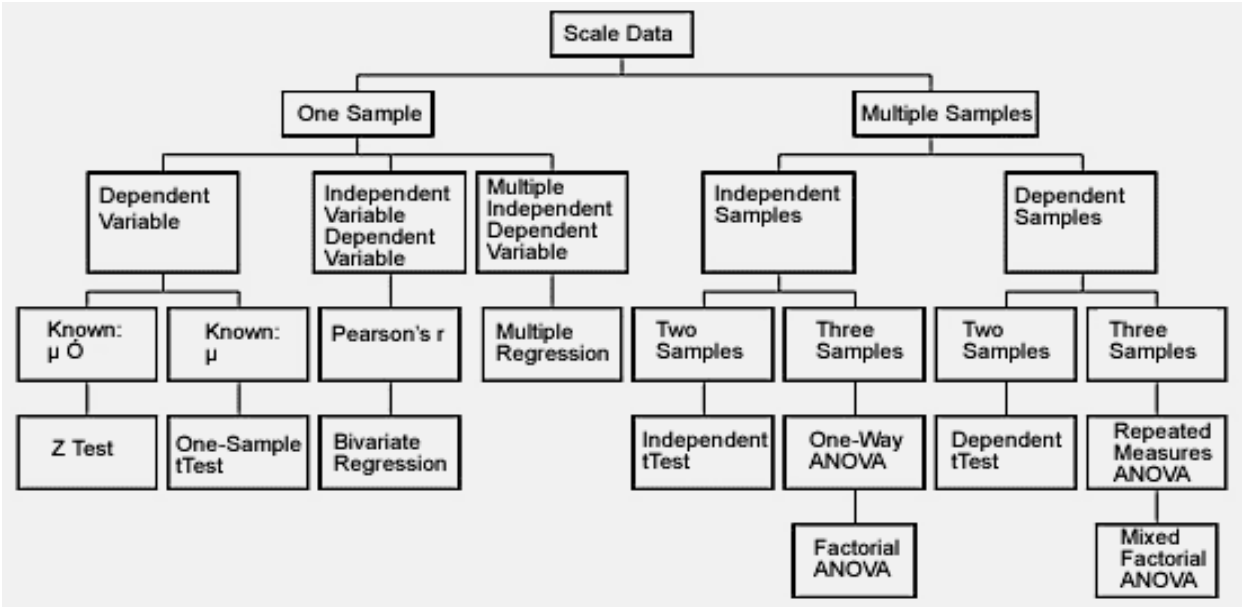
Most empirical research belongs clearly to one of those two general categories. In correlational research we do not (or at least try not to) influence any variables but only measure them and look for relations (correlations) between some set of variables, such as blood pressure and cholesterol level. In experimental research, we manipulate some variables and then measure the effects of this manipulation on other variables; for example, a researcher might artificially increase blood pressure and then record cholesterol level. Data analysis in experimental research also comes down to calculating "correlations" between variables, specifically, those manipulated and those affected by the manipulation. However, experimental data may potentially provide qualitatively better information: Only experimental data can conclusively demonstrate causal relations between variables. For example, if we found that whenever we change variable A, variable B changes, then we can conclude that "A influences B." Data from correlational research can only be "interpreted" in causal terms based on some theories that we have, but correlational data cannot conclusively prove causality.



DECISION TREE FOR NOMINAL DATA ANALYSIS

DECISION TREE FOR ORDINAL DATA ANALYSIS





DECISION TREE FOR CONTINUOUS-SCALED DATA ANALYSIS

SECTION 3
GLOSSARY OF COMMON RESEARCH DESIGN AND METHODS TERMINOLOGY

Elementary Concepts in Statistics:

http://www.statsoft.com/textbook/esc.html

http://www.bath.ac.uk/e-learning/gold/glossary.html#N1862

http://www.ecs.org/html/educationIssues/Research/primer/foreword.asp

A Policymaker's primer on education research

Mid-continent research for education and learning

How to determine that a result is "really" significant: There is no way to avoid arbitrariness in the final decision as to what level of significance will be treated as really "significant." That is, the selection of some level of significance, up to which the results will be rejected as invalid, is arbitrary. In practice, the final decision usually depends on whether the outcome was predicted a priori or only found post hoc in the course of many analyses and comparisons performed on the data set, on the total amount of consistent supportive evidence in the entire data set, and on "traditions" existing in the particular area of research. Typically, in many sciences, results that yield $p \leq .05$ are considered borderline statistically significant, but remember that this level of significance still involves a pretty high probability of error (5%). Results that are significant at the $p \leq .01$ level are commonly considered statistically significant, and $p \leq .005$ or $p \leq .001$ levels are often called "highly" significant. But remember that those classifications represent nothing but arbitrary conventions that are only informally based on general research experience.

Dependent vs. independent variables: Independent variables are those that are manipulated, whereas dependent variables are only measured or registered. This distinction appears terminologically confusing to many because, as some students say, "all variables depend on something." However, once you get used to this distinction, it becomes indispensable. The terms dependent and independent variable apply mostly to experimental research where some variables are manipulated, and in this sense, they are "independent" from the initial reaction patterns, features, intentions, etc. of the subjects. Some other variables are expected to be "dependent" on the manipulation or experimental conditions. That is to say, they depend on "what the subject will do" in response. Somewhat contrary to the nature of this distinction, these terms are also used in studies where we do not literally manipulate independent variables, but only assign subjects to "experimental groups" based on some pre-existing properties of the subjects. For example, if in an experiment, males are compared with females regarding their white cell count (WCC), Gender could be called the independent variable and WCC the dependent variable.

Measurement scales: Variables differ in "how well" they can be measured, i.e., in how much measurable information their measurement scale can provide. There is obviously some measurement error involved in every measurement, which determines the "amount of information" that we can obtain. Another factor that determines the amount of information that can be provided by a variable is its "type of measurement scale." Specifically variables are classified as (a) nominal, (b) ordinal, (c) interval or (d) ratio.

a. *Nominal variables* allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories. For example, all we can say is that two individuals are different in terms of variable A (e.g., they are of different race), but we cannot say which one "has more" of the quality represented by the variable. Typical examples of nominal variables are gender, race, color, city, etc.

b. *Ordinal variables* allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but they do not allow us to say "how much more." A typical example of an ordinal variable is the socioeconomic status of families. For example, we know that upper-middle is higher than middle but we cannot say that it is, for

example, 18% higher. Also this very distinction between nominal, ordinal, and interval scales itself represents a good example of an ordinal variable. For example, we can say that nominal measurement provides less information than ordinal measurement, but we cannot say "how much less" or how this difference compares to the difference between ordinal and interval scales.

c. *Interval variables* allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them. For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale. We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees.

d. *Ratio variables* are very similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus they allow for statements such as x is two times more than y. Typical examples of ratio scales are measures of time or space. For example, as the Kelvin temperature scale is a ratio scale, not only can we say that a temperature of 200 degrees is higher than one of 100 degrees, we can correctly state that it is twice as high. Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

Null hypothesis: A hypothesis test involves calculating the probability of seeing a result at least as extreme as the observed data, given some initial assumption about what the underlying probability distribution actually is. This initial assumption is typically based on some general concept that nothing noteworthy is happening, such as "the means for all groups are the same regardless of treatment," or "the survival rate remains the same for subjects given drug X or drug Y." This initial assumption is called the null hypothesis. You then compare the calculated probability with your pre-selected significance level. If the P value is smaller than the significance level, the null hypothesis is rejected, and the test result is termed significant. For the two-sample unpaired t test, the null hypothesis is that the two population means are equal, and the t test involves finding the probability of observing a t statistic at least as extreme as the one calculated from the data, assuming the null hypothesis is true.

P value: In a statistical hypothesis test, the P value is the probability of observing a test statistic at least as extreme as the value actually observed, assuming that the null hypothesis is true. This probability is then compared to the pre-selected significance level of the test. If the P value is smaller than the significance level, the null hypothesis is rejected, and the test result is termed significant. The P value depends on both the null hypothesis and the alternative hypothesis. In particular, a test with a one-sided alternative hypothesis will generally have a lower P value (and thus be more likely to be significant) than a test with a two-sided alternative hypothesis. However, one-sided tests require more stringent assumptions than two-sided tests. They should only be used when those assumptions apply.

Robust: Robust statistical tests are tests that operate well across a wide variety of distributions. A test can be robust for validity, meaning that it provides P values close to the true ones in the presence of (slight) departures from its assumptions. It may also be robust for efficiency, meaning that it maintains its statistical power (the probability that a true violation of the null hypothesis will be detected by the test) in the presence of those departures.

Scale: The generalized concept of the variability or dispersion of a distribution. Typical measures of scale are variance, standard deviation, range, and interquartile range. Scale and spread both refer to the same general concept of variability.

Sensitivity: The sensitivity of a test is the probability that the test will declare the condition of interest present when the condition is in fact present.

Specificity: The specificity of a test is the probability that the test will declare the condition of interest absent when the condition is in fact absent.

Skewness: Skewness is a lack of symmetry in a distribution. Data from a positively skewed (skewed to the right) distribution have values that are bunched together below the mean, but have a long tail above the mean. (Distributions that are forced to be positive, such as annual income, tend to be skewed to the right.) Data from a negatively skewed (skewed to the left) distribution have values that are bunched together above the mean, but have a long tail below the mean. Boxplots may be useful in detecting skewness to the right or to the left; normal probability plots may also be useful in detecting skewness to the right or to the left.

Type I and Type II error: When testing a null hypothesis, there are two ways to draw a mistaken conclusion from the test. The first, Type I error, is to incorrectly conclude that the null hypothesis is false when it is in fact true. The probability of this error is usually denoted by the Greek letter alpha. By selecting the significance level (alpha-level) for a hypothesis test, you specify the value of the Type I error you are willing to tolerate if the null hypothesis is true. A Type II error occurs when the test fails to reject the null hypothesis when it is in fact false. The probability of this error is usually denoted by the Greek letter beta, and is equal to 1 minus the power of the test. The probability of a Type II error depends on the significance level (alpha-level) of the test, the components of the calculation of the test statistic, and on the specific alternative hypothesis under consideration. In general, for a particular hypothesis test, significance level, and alternative hypothesis, the probability of a Type II error decreases as the amount of data collected increases. Because the probability of a Type II error depends on the alternative hypothesis, it can only be calculated with reference to a specific alternative hypothesis. A power curve shows (1 - beta) plotted for different possible alternatives, such as the possible differences between the two population means for a two-sample unpaired t test.

Violation of assumptions: Statistical hypothesis tests generally make assumptions about the population(s) from which the data were sampled. For example, many normal theory-based tests such as the t test and ANOVA assume that the data are sampled from one or more normal distributions, as well as that the variances of the different populations are the same (homoscedasticity). If test assumptions are violated, the test results may not be valid.

Confidence Intervals: Statisticians stress the importance of using confidence intervals (CIs). There is, however, debate over which type of CIs to use and how to best define and interpret them. In spite of this confusion, you should use CIs to express the results of statistical tests because they convey more information than P values alone. The confidence level sets the boundaries of a confidence interval, this is conventionally set at 95% to coincide with the 5% convention of statistical significance in hypothesis testing. In some studies wider (e.g. 90%) or narrower (e.g. 99%) confidence intervals will be required. This rather depends upon the nature of your study. You should consult a statistician before using CI's other than 95%.

You will hear the terms confidence interval and confidence limit used. The confidence interval is the range Q-X to Q+Y where Q is the value that is central to the study question, Q-X is the **lower confidence limit** and Q+Y is the **upper confidence limit**.

> Common **A 95% CI is the interval that you are 95% certain contains the true population value as it might be estimated from a much larger study**.
>
> The value in question can be a mean, the difference between two means, a proportion, etc. The CI is usually, but not necessarily, symmetrical about this value.
>
> Pure Bayesian The Bayesian concept of a **credible interval** is sometimes put forward as a more practical concept than the confidence interval. For a 95% credible interval, the value of interest (e.g. size of treatment effect) lies with a 95% probability in the interval. This interval is then open to subjective molding of interpretation. Furthermore, the credible interval can only correspond exactly to the confidence interval if prior probability is so called "uninformative".
>
> Pure frequentist Most pure frequentists say that it is not possible to make probability statements, such CI interpretation, about the study values of interest in hypothesis tests.

Normal probability plot: A normal probability plot, also known as a normal Q-Q plot or normal quantile-quantile plot, is the plot of the ordered data values (as Y) against the associated quantiles of the normal distribution (as X). For data from a normal distribution, the points of the plot should lie close to a straight line. Examples of these plots illustrate various situations.

Normal (Gaussian) distribution: The normal or Gaussian distribution is a continuous symmetric distribution that follows the familiar bell-shaped curve. The distribution is uniquely determined by its mean and variance. It has been noted empirically that many measurement variables have distributions that are at least approximately normal. Even when a distribution is non-normal, the distribution of the mean of many independent observations from the same distribution becomes arbitrarily close to a normal distribution as the number of observations grows large.

Boxplot: A boxplot is a graph summarizing the distribution of a set of data values. The upper and lower ends of the center box indicate the 75th and 25th percentiles of the data, the center box indicates the median, and the center + indicates the mean. Suspected outliers appear in a boxplot as individual points o or x outside the box. The o outlier values are known as outside values, and the x outlier values as far outside values. If the difference (distance) between the 75th and 25th percentiles of the data is H, then the outside values are those values that are more than 1.5H but no more than 3H above the upper quartile, and those values that are more than 1.5H but no more than 3H below the lower quartile. The far outside values are values that are at least 3H above the upper quartile or 3H below the lower quartile.

Confidence interval: Select a significance level alpha. Perform an experiment to estimate an unknown parameter (such as collecting a sample from population and estimating the population mean by the sample mean). The confidence interval for a unknown parameter is a range of values such that, if the same experiment were done many times and a confidence interval calculated each time, the proportion 1-alpha of the confidence intervals would contain the true value of the unknown parameter. A 95% confidence interval for the mean is sometimes loosely described as having a 95% probability of containing the true population mean. Obviously, once the data have been collected and a confidence interval calculated, the interval either contains the true population mean or it doesn't, so this description can't be strictly true. More correctly, the description is a type of "shorthand" for the statement that if the same experiment had been done many times, and a thus many confidence intervals calculated, then 95% of them would contain the true population mean, and that therefore, before we actually calculated our own confidence interval, it had an a priori 95% chance of containing the true population mean.

Chi-square test for goodness of fit: The chi-square test for goodness of fit tests the hypothesis that the distribution of the population from which nominal data are drawn agrees with a posited distribution. The chi-square goodness-of-fit test compares observed and expected frequencies (counts). The chi-square test statistic is basically the sum of the squares of the differences between the observed and expected frequencies, with each squared difference divided by the corresponding expected frequency.

Chi-square test for independence (Pearson's): Pearson's chi-square test for independence for a contingency table tests the null hypothesis that the row classification factor and the column classification factor are independent. Like the chi-square goodness-of-fit test, the chi-square test for independence compares observed and expected frequencies (counts). The expected frequencies are calculated by assuming the null hypothesis is true. The chi-square test statistic is basically the sum of the squares of the differences between the observed and expected frequencies, with each squared difference divided by the corresponding expected frequency. Note that the chi-square statistic is always calculated using the counted frequencies. It cannot be calculated using the observed proportions, unless the total number of subjects (and thus the frequencies) is also known.

Inappropriate use of chi-square test: Pearson's chi-square test for independence for a contingency table involves using a normal approximation to the actual distribution of the frequencies in the contingency table. This approximation becomes less reliable when the expected frequencies for the contingency table are very small. A standard (and conservative) rule of thumb (due to Cochran) is to avoid using the chi-square test for contingency tables with expected cell frequencies less than 1, or when more than 20% of the contingency table cells have expected cell frequencies less than 5. In such cases, an alternate test like Fisher's exact test for a 2x2 contingency table should be considered for a more accurate evaluation of the data.

Correlation: The most commonly used correlation statistic is the Pearson correlation coefficient. This statistic measures both the **strength** and **direction** of the linear relationship between two variables.

Correlation Example

Suppose we want to look at the relationship between age and height in children. We select a group of children for study, and for each child we record their age in years and their height in inches. We could plot these values on a graph so that the child's age would be on the horizontal axis and the child's height would be on the vertical axis. Each dot on the plot represents a single child's age and height. This is called a scatter plot.

Since older children are generally taller than younger children, we would expect the dots on the plot to roughly approximate a straight line (a linear relationship between the variables) and that the line will slope upward (since age and height tend to increase at the same time).

Correlation Coefficient: The Pearson correlation coefficient is a number between -1 and +1 that measures both the strength and direction of the linear relationship between two variables.

The magnitude of the number represents the strength of the correlation. The larger r is in absolute value, the stronger the linear association between X and Y. If r is 0, X and Y are said to be uncorrelated, with no linear association between X and Y. Independent variables are always uncorrelated, but uncorrelated variables need not be independent. A correlation coefficient of zero represents no linear relationship (the scatter plot does not resemble a straight line at all), while a correlation coefficient of -1 or +1 means that the relationship is perfectly linear (all of the dots fall exactly on a straight line).

The sign (+/-) of the correlation coefficient indicates the direction of the correlation. A positive (+) correlation coefficient means that as values on one variable increase, values on the other variable tend to also increase; a negative (-) correlation coefficient means that as values on one variable increase, values on the other tend to decrease, that is, they tend to go in opposite directions.

T-Test: There are several kinds of t-tests, but the most common is the "two-sample t-test" also known as the "Student's t-test" or the "independent samples t-test".

T-Test Example

The two sample t-test simply tests whether or not two independent populations have different mean values on some measure.

For example, we might have a research hypothesis that rich people have a different quality of life than poor people. We give a questionnaire that measures quality of life to a random sample of rich people and a random sample of poor people. The null hypothesis, which is assumed to be true until proven wrong, is that there is really no difference between these two populations.

We gather some sample data and observe that the two groups have different average scores. But does this represent a real difference between the two populations, or just a chance difference in our samples?

T-Test Statistic

The statistics t-test allows us to answer this question by using the t-test statistic to determine a p-value that indicates how likely we could have gotten these results by chance. By convention, if there is a less than 5% chance of getting the observed differences by chance, we reject the null hypothesis and say we found a statistically significant difference between the two groups. See Statistical Data Analysis for more information about hypothesis testing.

Odds Ratio: http://www.bmj.com/cgi/content/full/320/7247/1468?ck=nck

Nonparametric tests: Nonparametric tests are tests that do not make distributional assumptions, particularly the usual distributional assumptions of the normal-theory based tests. These include tests that do not involve population parameters at all (truly nonparametric tests such as the chi-square goodness of fit test), and distribution-free tests, whose validity does not depend on the population distribution(s) from which the data have been sampled. In particular, nonparametric tests usually drop the assumption that the data come from normally distributed populations. However, distribution-free tests generally do make some assumptions, such as equality of population variances.

When to Use Which Method

It is not easy to give simple advice concerning the use of nonparametric procedures. Each nonparametric procedure has its peculiar sensitivities and blind spots. For example, the Kolmogorov-Smirnov two-sample test is not only sensitive to differences in the location of distributions (for example, differences in means) but is also greatly affected by differences in their shapes. The Wilcoxon matched pairs test assumes that one can rank order the magnitude of differences in matched observations in a meaningful manner. If this is not the case, one should rather use the Sign test. In general, if the result of a study is important (e.g., does a very expensive and painful drug therapy help people get better?), then it is always advisable to run different nonparametric tests. Should discrepancies in the results occur contingent upon which test is used, one should try to understand why some tests give different results. On the other hand, nonparametric statistics are less statistically powerful (sensitive) than their parametric counterparts, and if it is important to detect even small effects (e.g., is this food additive harmful to people?) one should be very careful in the choice of a test statistic.

Large data sets and nonparametric methods: Nonparametric methods are most appropriate when the sample sizes are small. When the data set is large (e.g., *n* > 100) it often makes little sense to use nonparametric statistics at all. When the samples become very large, then the sample means will follow the normal distribution even if the respective variable is not normally distributed in the population, or is not measured very well. Thus, parametric methods, which are usually much more sensitive (i.e., have more *statistical power*) are in most cases appropriate for large samples. However, the tests of significance of many of the nonparametric statistics described here are based on asymptotic (large sample) theory; therefore, meaningful tests can often not be performed if the sample sizes become too small. Please refer to the descriptions of the specific tests to learn more about their power and efficiency.

Nonparametric Correlations

The following are three types of commonly used nonparametric correlation coefficients

Spearman R. Spearman R (Siegel & Castellan, 1988) assumes that the variables under consideration were measured on at least an ordinal (rank order) scale, that is, that the individual observations can be ranked into two ordered series. Spearman R can be thought of as the regular Pearson product moment correlation coefficient, that is, in terms of proportion of variability accounted for, except that Spearman R is computed from ranks.

Kendall tau. Kendall tau is equivalent to Spearman R with regard to the underlying assumptions. It is also comparable in terms of its statistical power. However, Spearman R and Kendall tau are usually not identical in magnitude because their underlying logic as well as their computational formulas are very different. Siegel and Castellan (1988) express the relationship of the two measures in terms of the inequality:

-1 £ 3 * Kendall tau - 2 * Spearman R £ 1

More importantly, Kendall tau and Spearman R imply different interpretations: Spearman R can be thought of as the regular Pearson product moment correlation coefficient, that is, in terms of proportion of variability accounted for, except that Spearman R is computed from ranks. Kendall tau, on the other hand, represents a probability, that is, it is the difference between the probability that in the observed data the two variables are in the same order versus the probability that the two variables are in different orders.

Gamma. The Gamma statistic (Siegel & Castellan, 1988) is preferable to Spearman R or Kendall tau when the data contain many tied observations. In terms of the underlying assumptions, Gamma is equivalent to Spearman R or Kendall tau; in terms of its interpretation and computation it is more similar to Kendall tau than Spearman R. In short, Gamma is also a probability; specifically, it is computed as the difference between the probability that the rank ordering of the two variables agree minus the probability that they disagree, divided by 1 minus the probability of ties. Thus, Gamma is basically equivalent to Kendall tau, except that ties are explicitly taken into account.

The Mann Whitney U statistic This is a method for the comparison of two independent random samples (x and y) defined as:

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=n_2+1}^{n2} R_i$$

- where samples of size n1 and n2 are pooled and Ri are the ranks. U can be resolved as the number of times observations in one sample precede observations in the other sample in the ranking.

Wilcoxon rank sum, Kendall's S and the Mann-Whitney U test are exactly equivalent tests. In the presence of ties the Mann-Whitney test is also equivalent to a chi-square test for trend.

In most circumstances a two sided test is required; here the alternative hypothesis is that x values tend to be distributed differently to y values. For a lower side test the alternative hypothesis is that x values tend to be smaller than y values. For an upper side test the alternative hypothesis is that x values tend to be larger than y values.

Assumptions of the Mann-Whitney test include random samples from populations; independence within samples and mutual independence between samples; measurement scale is at least ordinal

A confidence interval for the difference between two measures of location is provided with the sample medians. The assumptions of this method are slightly different from the assumptions of the Mann-Whitney test: random samples from populations; independence within samples and mutual independence between samples; two population distribution functions are identical apart from a possible difference in location parameters

Kappa Statistics: an index that compares the agreement against that which might be expected by chance. Kappa can be thought of as the chance-corrected proportional agreement, and possible values range from +1 (perfect agreement) via 0 (no agreement above that expected by chance) to -1 (complete disagreement).

To assess the accuracy of any particular measuring 'instrument', it is usual to distinguish between the *reliability* of the data collected and their *validity*. Reliability is essentially the extent of the agreement between repeated measurements, and validity is the extent to which a method of measurement provides a true assessment of that which it purports to measure. When studying the variability of observer *categorical* ratings, two components of possible lack of accuracy must be distinguished. The first is inter-observer bias, which is reflected in differences in the marginal distributions of the response variable for each of the observers (Cochran's Q-test is the appropriate test for the hypothesis of no inter-observer bias). The second is observer disagreement, which is indicated by how observers classify individual subjects into the same category on the measurement scale (Kappa coefficient is one of the most common approaches). In this part, we will focus on the Kappa coefficient (or Kappa statistics).

Hypothetical Example using Kappa: 29 patients are examined by two independent doctors (see Table). 'Yes' denotes the patient is diagnosed with disease X by a doctor. 'No' denotes the patient is classified as no disease X by a doctor.

		Doctor A		
		No	Yes	Total
Doctor B	No	10 (34.5%)	7 (24.1%)	17 (58.6%)
	Yes	0 (0.0%)	12 (41.4%)	12 (41.4%)
Total		10 (34.5%)	19 (65.5%)	29

Kappa = (Observed agreement - Chance agreement)/(1 - Chance agreement)
Observed agreement = (10 + 12)/29 = 0.76
Chance agreement = 0.586 * 0.345 + 0.655 * 0.414 = 0.474
Kappa = (0.76 - 0.474)/(1 - 0.474) = 0.54

Parametric statistics are statistics that estimate population parameters.

Parametric inferential statistical methods are mathematical procedures for statistical hypothesis testing which assume that the distributions of the variables being assessed belong to known parametrized families of probability distributions. In that case we speak of parametric model.

All parametric tests involve (a) estimating at least one population parameter, (b) assumptions about the distribution of the population from which the data were randomly sampled, and (c) assumptions about the measurement of the dependent variable.

For example, analysis of variance (ANOVA) assumes that the underlying distributions are normally distributed and that the variances of the distributions being compared are similar. The Pearson product-moment correlation coefficient also assumes normality.

While parametric techniques are robust – that is, they often retain considerable power to detect differences or similarities even when these assumptions are violated – some distributions violate the assumptions so markedly that a non-parametric alternative is more likely to detect a difference or similarity.

ANOVA: While the t-test is used to compare the means between two groups, ANOVA is used to compare means between 3 or more groups.

There are several varieties of ANOVA, such as one-factor (or one-way) ANOVA, two-factor (or two-way) ANOVA, and so on, and also repeated measures ANOVA. The factors are the independent variables, each of which must be measured on a categorical scale - that is, levels of the independent variable must define separate groups.

One-Way ANOVA Example

One-factor ANOVA, also called one-way ANOVA is used when the study involves 3 or more levels of a single independent variable. For example we might look at average test scores for students exposed to one of three different teaching techniques (three levels of a single independent variable).

ANOVA Statistics

The null hypothesis for ANOVA is that the mean (average value of the dependent variable) is the same for all groups. The alternative or research hypothesis is that the average is not the same for all groups.

The ANOVA test procedure produces an F-statistic, which is used to calculate the p-value. In the most rigid interpretation, $p < .05$ indicates that the null hypothesis if false. However, there may be clinically important information hidden with p values under .10.

With ANOVA, if the null hypothesis is rejected, then all we know is that at least 2 groups are different from each other. In order to determine which groups are different from which, post-hoc t-tests are performed using some form of correction (such as the Bonferroni correction) to adjust for an inflated probability of a Type I error.

Within effects: In a repeated measures ANOVA, there will be at least one factor that is measured at each level for every subject. This is a within (repeated measures) factor. For example, in an experiment in which each subject performs the same task twice, trial number is a within factor. There may also be one or more factors that are measured at only one level for each subject, such as gender. This type of factor is a between or grouping factor.

MANOVA Interaction: In multi-factor analysis of variance, factors A and B interact if the effect of factor A is not independent of the level of factor B. For example, in a drug experiment involving rats, there would be an interaction between the factors sex and treatment if the effect of treatment was not the same for males and females.

Repeated measures ANOVA: In a repeated measures ANOVA, there will be at least one factor that is measured at each level for every subject in the experiment. This is a within (repeated measures) factor. For example, in an experiment in which each subject performs the same task twice is a repeated measures design, with trial (or trial number) as the within factor. If every subject performed the same task twice under each of two conditions, for a total of 4 observations for each subject, then both trial and condition would be within factors. In a repeated measures design, there may also be one or more factors that are measured at only one level for each subject, such as gender. This type of factor is a between or grouping factor.

Linear logistic model: A linear logistic model assumes that for each possible set of values for the independent (X) variables, there is a probability p that an event (success) occurs. Then the model is that Y is a linear combination of the values of the X variables**:**

Linear regression: In a linear regression, the fitted (predicted) value of the response variable Y is a linear combination of the values of one or more predictor (X) variables**:**

fitted $Y = b0 + b1*X1 + b2*X2 + \ldots + bk*Xk$.

An X variable in the model equation could be a nonlinear function of an observed variable (e.g., one might observe distance, but use distance squared as an X variable in the model, or X2 might be the square of X1), as long as the fitted Y remains a sum of terms that are each an X variable multiplied by a coefficient. The most basic linear regression model is simple linear regression, which involves one X variable**:** fitted $Y = b0 + b1*X$.

Multiple linear regression refers to a linear regression with more than one X variable.

It is used to determine the extent to which there is a linear relationship between a dependent variable and one or more independent variables. There are two types of linear regression, simple linear regression and multiple linear regression.

In **simple linear regression** a single independent variable is used to predict the value of a dependent variable. In **multiple linear regression** two or more independent variables are used to

predict the value of a dependent variable. The difference between the two is the number of independent variables. In both cases there is only a single dependent variable.

Linear Regression - Data Considerations

The dependent variable must be measured on a continuous measurement scale (e.g. 0-100 test score) and the independent variable(s) can be measured on either a categorical (e.g. male versus female) or continuous measurement scale. There are several other assumptions that the data must satisfy in order to qualify for linear regression.

Correlation and Regression

Simple linear regression is similar to correlation in that the purpose is to measure to what extent there is a linear relationship between two variables. The major difference between the two is that correlation makes no distinction between independent and dependent variables while linear regression does. In particular, the purpose of linear regression is to "predict" the value of the dependent variable based upon the values of one or more independent variables.

Path Analysis is a sophisticated statistical mapping procedure for use with complex multi-factorial designs. This is often used in studies of the etiology or causes of behaviors. It could be used in a study of OMM outcomes if multiple factors are hypothesized to contribute to an outcome.

Survival Analysis This is also a sophisticated statistical tool. It is used to predict the end point for subjects in a study for whom the data is not collected. It may predict, for example what outcomes might occur in a sample of patients with pulmonary dysfunctions over time, given a set of initial parameters. Often it is referred to the prediction of when a subject will die.

Factor Analysis: A factor is a single discrete classification scheme for data, such that each item classified belongs to exactly one class (level) for that classification scheme. For example, in a drug experiment involving rats, sex (with levels male and female) or drug received could be factors. A one-way analysis of variance involves a single factor classifying the subjects (e.g., drug received); multi-factor analysis of variance involves multiple factors classifying the subjects (e.g., sex and drug received).

SECTION 4
EXAMPLES FOR METHODS SUB-SECTION

Study Protocol

Example

Figure 4 is the study protocol schedule. All subjects will be enrolled in the study via the level one screening and consent process by a clinical research coordinator (CRC). Level two screening data will be collected by the CRC with consultation and review by the PI or Co-I. All treatment and outcome visits will occur in the Osteopathic Manipulative Medicine (OMM) clinic at UNTHSC/TCOM. MRI tests will be done at the Monticello Diagnostic Clinic, described in the Resources Section of the proposal. Subjects will receive $10 for their travel time and expenses for each study visit. All forms (Demographic Data Sheet, SOAP Note Form, Symptoms Severity Scale, Functional Status Scale, Data Collection Form) are contained in Appendix A.

Experimental Groups and Interventions

Example

The experimental design for this study is a two-factor design, with repeated measures on one factor, and is illustrated in Figure 6. The independent factor is the treatment group that consists of three levels: OMT group, sham treatment group, and conventional care only group. The repeated factor is time which consists of a variable number of levels, dependent on the individual subject and the time until s/he reaches one of the conditions for stopping the treatments. A total of 360 subjects, 120 in each group, will be recruited into the study. After informed consent is obtained, the study participants will be stratified (as described in Section D.4) within each site and then randomized into one of three treatment groups. Measurements will be taken on all subjects preceding the initial treatment (as described in Section D.5). Following the pre-treatment measurements, subjects will receive …

Outcome Measures

Example (adequate but could be stronger by eliminating the rationale statements that should be elsewhere in the proposal)

Hypothesis 1: *Persistent hyperglycemia increases myocardial stiffness and decreases left ventricular compliance. Rationale. Hyperglycemia causes oxyradical-mediated crosslinking of myocardial interstitial collagen.[4] As these crosslinks, termed advanced glycation endproducts (AGEs) accumulate, the ventricles become … Although myocardial stiffening has been well characterized in diabetic rats[24] it has never been examined in large diabetic mammals including dogs and patients, and the impact of antioxidants on myocardial stiffness has not been tested in any species.*

Protocol and interpretation. Left ventricular chamber stiffness will be assessed from end-diastolic pressure:volume relationships obtained in conscious dogs. Left ventricular pressure and volume will be measured … These studies will be conducted weekly to monitor the …..and compared with pre-alloxan baseline and values obtained in non-diabetic dogs.

Data Analysis Plan

Example 1

Secondary aim one is to evaluate the effects of OMT with and without stretching exercise on strength and nerve sensation in persons with CTS.

Hypothesis 3*. OMT plus stretching exercise will most effectively improve strength.*

Grip strength will be measured with the Jaymar Dynomometer™. A pinch gauge by Baseline® will be used for pinch strength. Adjustments will be made as necessary for males and females, and for age. To test this hypothesis grip and pinch strength will be analyzed using repeated measures ANOVA between groups, and t-tests for gender and other relevant classification variables.

Hypothesis 4*. OMT plus stretching exercise will most effectively improve sensation.*

Sensation data will be in the form of microampere sensation thresholds for Median and Ulnar nerves at the three frequencies tested for each nerve (2000, 250 and 5 Hz). Median and Ulnar sensation thresholds will be compared using repeated measures analysis of variance (ANOVA) between treatment groups to determine which groups had the greatest improvement in sensation as measured by CPT.

Example 2

Statistical Analysis

To test the hypotheses regarding group differences on health care resource utilization and health outcomes, four methods of analysis will be used. Statistical methods that stratify by site will be used for analyses of data from more than one site. To compare the three groups on length of hospital stay, duration of IV antibiotic use, duration of oral antibiotic use in the hospital, and time to clinical stability, stratified Cox proportional hazards models will be employed. [65] The differences in the effects of the three groups on continuous outcome measures where repeated measurements are made (e.g., leukocyte counts, rate of symptomatic and functional recovery, somatic dysfunction burden) will be examined using general linear mixed-effects models (marginal models) with the subjects and the sites treated as random effects and the group as a fixed effect. [66] For repeated categorical outcome measures (e.g., in-hospital complications), generalized estimating equations (GEE) will be utilized. [67, 68] Analysis of categorical outcome measures that are measured only once (for example, reason for ending study treatments, re-admission rates) will be performed using logistic regression models. [68]

Study Timetable and Management Plan

A research proposal should always contain a timetable and management plan, however brief it may be. It lets the reader know that you have considered the logistics of carrying out your research plan. The shape of this depends on your research, particularly if it is sequential, or multi-center, or involves a data safety and management board (DSMB) or other advisory groups. It lists the key steps involved in the project from start-up to final report and marks the weeks or months during which these will occur. It may be similar in format to the protocol time schedule and assures the reviewer that you have a plan in mind even though it may be modified at a later time. SEE THE EXAMPLE IN CHAPTER I.

SECTION 5
GOOD HABITS FOR SCIENTIFIC WRITING
IN THE RESEARCH DESIGN AND METHODS SECTION

You should be able to write a Research Design and Methods section using a non-experimental design in two to three pages. Flow charts are great to use.

The guiding principle for writing the Method section is that it should contain sufficient information for the reader to determine whether the plan you have to collect information from specific sources is logical, well considered, free of gaps (or you can address any inherent gaps), free of leaps of logic, doable, appropriate to the question.

A good design and methods section should contain sufficient details for another qualified researcher to implement the study.

THE LIST BELOW is for you to check these aspects of your proposal or manuscript. These are the 10 most often cited reasons for proposals or manuscripts to be rejected. These reasons are taken from many different sources that repeat these reasons, thus they are presented here as the most often cited

TOP TEN REASONS FOR REJECTION OF MANUSCRIPTS AND PROPOSALS.

- ☐ Failure to provide the proper context to frame the research question
- ☐ Failure to cite landmark studies
- ☐ Failure to accurately present the theoretical and empirical contributions by other researchers
- ☐ Failure to stay focused on the research question
- ☐ Failure to develop a coherent and persuasive argument for the proposed research
- ☐ Too much detail on minor issues, but not enough detail on major issues
- ☐ Too much rambling -- going "all over the map" without a clear sense of direction.(The best proposals move forward with ease and grace like a seamless river)
- ☐ Too many citation lapses and incorrect references
- ☐ Too long or too short
- ☐ Sloppy writing

Can you decide which is number one and which is number 1 and which is number 10?

SECTION 6
FINAL CHECKLIST

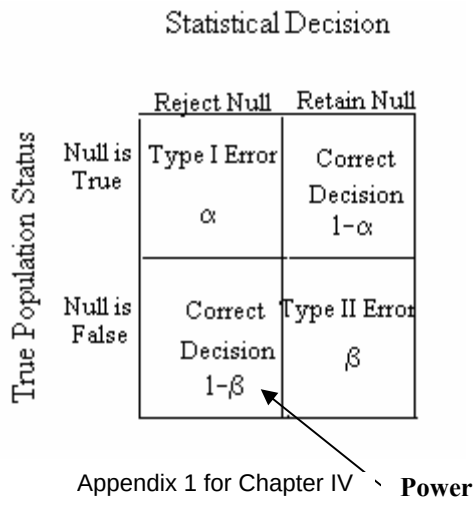
The Research Design and Methods section is the largest section of the proposal. It can be organized in many different ways according to the nature of your research. The Research Design and Methods section must describe the below listed aspects of the study. This list is presented here as a checklist for you to use as you develop this section of your proposal or scientific report.

- ☐ How you will achieve you specific aims
- ☐ What activities you will undertake to test the hypotheses for your study
- ☐ What you will do in your experiment, or what you will do to the human or animal subjects
- ☐ How you will do it
- ☐ When
- ☐ Where
- ☐ How
- ☐ How often
- ☐ Why
- ☐ What you will use to determine if the action you propose will generate the reaction or the result/outcome you hypothesize that it will generate – your measurements
- ☐ Subjects: Who will be in the study? How will you select them by inclusion and exclusion criteria, and how will you recruitment them? (NOTE: Human subjects (risk and exclusion justification) issues are covered in a separate section E and do not go here)
- ☐ The independent variables: what you are observing or what you expect to change – the outcomes or measures of change
- ☐ The 'Power' of the study: see guidelines on when and how to obtain a power analysis to estimate sample size
- ☐ The dependent variables: what you are using as an intervention or treatment
- ☐ Other data such as categorical variables
- ☐ The hypotheses to be tested, question to be explored, or the product to be developed
- ☐ The design of the study: randomized block, repeated measures, longitudinal, retrospective, prospective, cross-sectional, existing data, controlled, blinded, and the like
- ☐ The instrumentation to collect data, and to measure the experimental change, the clinical outcomes, the validity of the theory, or the utility/validity of the product
- ☐ The statistical tests you plan to use to test the hypotheses or question

Power Analysis and Sample Size Estimation

Daisha J. Cipher, Ph.D.
Director of Biostatistics, Veterans Affairs Rheumatoid Arthritis Registry

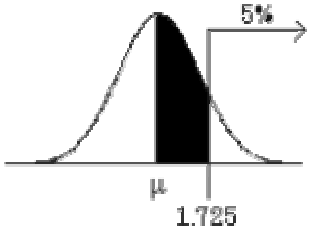
The *power* of a statistical test is the probability of correctly rejecting a false null hypothesis. This probability is inversely related to the probability of making a Type II error. Recall also that we choose the probability of making a Type I error when we set Alpha and that if we decrease the probability of making a Type I error we increase the probability of making a Type II error. The relationships are defined in the table below:



Power and Alpha

Thus, the probability of correctly retaining a true null has the same relationship to Type I errors as the probability of correctly rejecting an untrue null does to Type II error. Yet, as I mentioned above, if we decrease the odds of making one type of error, we increase the odds of making the other type of error. What is the relationship between Type I and Type II errors? The following demonstration attempts to illustrate this concept. In this demonstration a one-tailed one-sample t-test with 20 degrees of freedom is conducted at Alpha=.05.

The critical value for this test at α=.05 is 1.725



references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)
Cipher 2006

As you can see, the probability of making a Type II error (and thus having low power) varies as a function of Alpha. The lower our Alpha the less likely we are to make a Type I error, but the more likely we are to make a Type II error. What other factors affect the power of a test?

Anytime we test whether a sample differs from a population or whether two sample come from 2 separate populations, there is the assumption that each of the populations we are comparing has its own mean and standard deviation (even if we do not know it). The distance between the two population means will affect the power of our test. In the following demonstration an increase in the variance (the spread of the distribution) shows a corresponding overlap in the two distributions and an increase in Beta.

Power as a Function of Sample Size, Variance, and Effect Size

You should notice in the last demonstration that what really made the difference in the size of Beta was how much overlap there was in the two distributions. When the means were close together the two distributions overlapped a great deal compared to when the means were farther apart. Thus, anything that affects the extent the two distributions share common values will increase Beta (the likelihood of making a Type II error).

Sample size has an indirect effect on power because it affects the measure of variance we use to calculate the t-test statistic. Since we are calculating the power of a test that involves the comparison of sample means, we will be more interested in the standard error (the average difference in sample values) than standard deviation or variance by itself. Thus, sample size is of interest because it modifies our estimate of the standard deviation. When N is large we will have a lower standard error than when N is small. In turn, when N is large well have a smaller Beta region than when N is small.

references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)

Effect size measures, such as Cohen's d, R^2, or η^2 are indications of the magnitude of the difference, variance shared, or variance accounted for (respectively) in your variables. The most common effect size measure for a two-sample design is the Cohen's *d*. The d is calculated as:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{SD}$$

where the SD would be that of either group, considering that both groups are assumed to have approximately equal SDs.

The resulting *d* value represents the difference between the means of Group 1 and 2, in standard deviation units. Thus, a value of 1.0 means that the two groups differed exactly one standard deviation unit from one another. Cohen (1965) defined effect size as "the degree to which a phenomenon exists." He described effect sizes (*d* values) as having the following qualities:

d	
.25	small
. 50	medium
1.0	large

One can still use Cohen's *d*, even if the design has more than two groups. The *d* is calculated one pair at a time.

Estimating Sample Size and Power for Experimental Designs

How do you ensure that your experiment's sample size is adequate? Moreover, how can you know the extent of the power of your statistical test? The procedure called *power analysis* can answer both of these questions. The four factors involved in a power analysis are:

1. Level of significance (α, or alpha level) set by you, the researcher.
2. Probability of obtaining a significant result (power desired, or 1 - β).
3. The population effect size, or the hypothesized effect (or difference) between your groups.
4. Sample size.

Knowing any three of the above allows you to compute the fourth.

references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)
Cipher 2006

The most common power analysis actually entails estimating a sample size before the study begins. The following formula estimates adequate sample size for a two-sample design, two-sided test:

$$n = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2}$$

where:

$z_{1-\alpha/2}$ refers the z value that cuts off the middle 95% of a normal distribution

$z_{1-\beta}$ refers the desired power level

d is the Cohen's d value (see formula above)

Typically, $z_{1-\alpha/2}$ is 1.96 and $z_{1-\beta}$ is .80 (which means that you have a 20% chance of making a Type II error).

If you don't want to do power analyses by hand, you can refer to the many tables in the literature that will estimate sample size, power, or both. Otherwise, you can look on the Internet for online power calculators. Some of my favorites are:

http://calculators.stat.ucla.edu/powercalc/
http://www.math.yorku.ca/SCS/Demos/power/

When you finally compute the sample size that you need, you can interpret your findings as follows:

"I must have X amount of participants **in each group** to achieve a statistical power of .80, using a significance level of .05."

If you do NOT have X sample size per group above, if you run your analyses anyway and do not achieve significance, **you cannot differentiate between the possibilities that:**

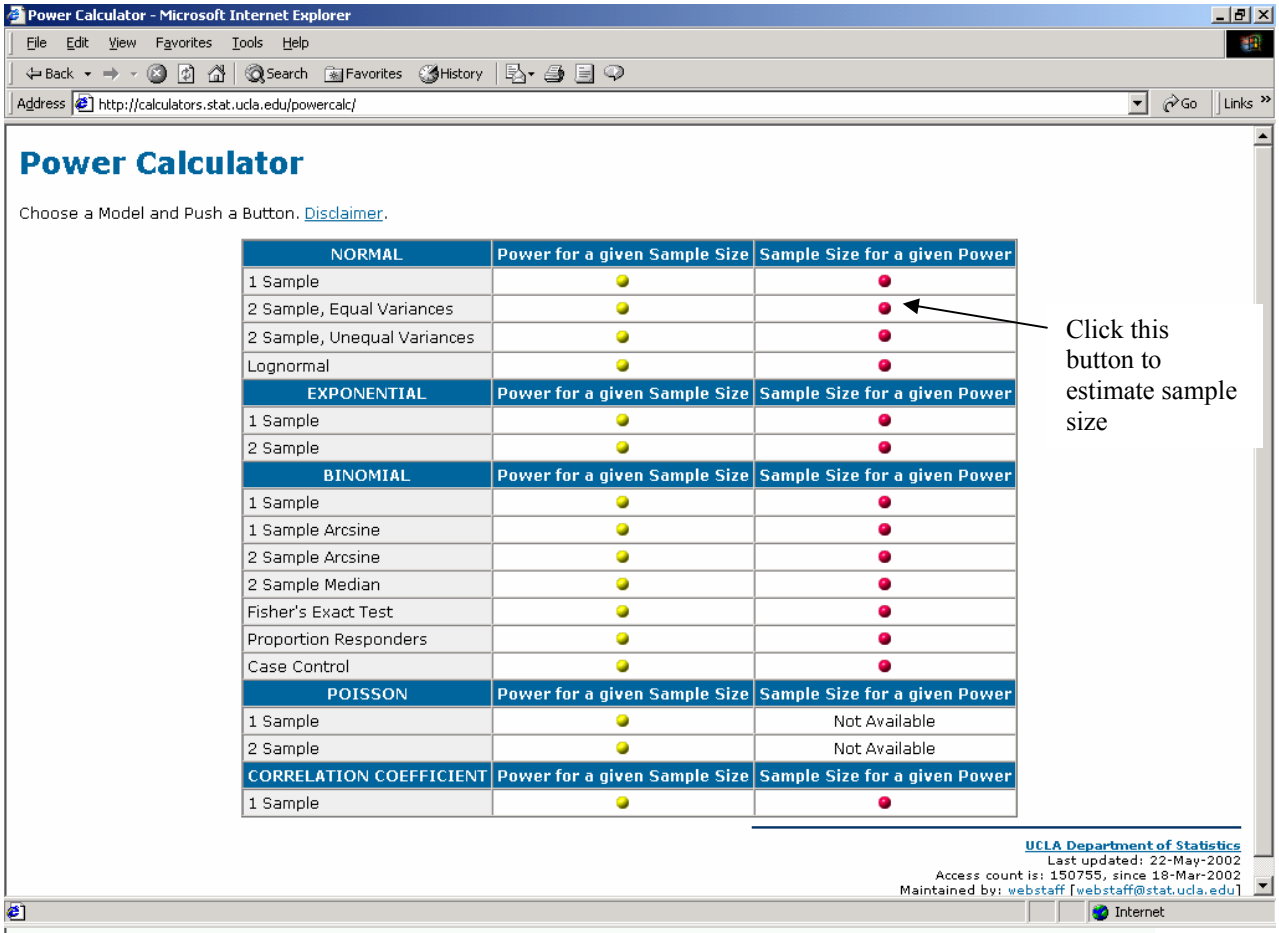
1. There are no differences in the population; or
2. There ARE differences but you don't have the adequate sample size to indicate that.

Example of Sample Size Calculation Using Online Power Calculator

Using the UCLA Department of Statistics website (calculators.stat.ucla.edu/powercalc), we can estimate our sample size for a two-sample design. Using the following hypothetical data reflecting IQ scores from two groups, we'll compare the calculator's results to our hand calculations.

Group 1 $\overline{X}$ = 100 SD = 15 α= .05 β= .80
Group 2 $\overline{X}$ = 107.5
(this means you're estimating d to equal .50)

references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)
Cipher 2006

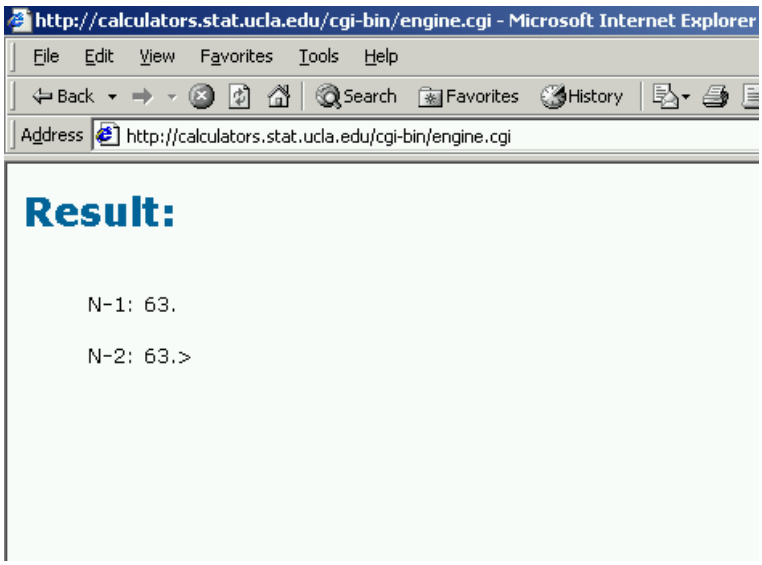


Normal Power Calculations

Normal Distribution 2-Sample Equal Variances

μ_1 The Mean of Population 1	100
μ_2 The Mean of Population 2	107.5
Sigma Common Standard Deviations for both Populations	15
Number of Sides Specifies Alternative Hypothesis. One sided and $\mu_1 > \mu_2$ => $H_1: \mu_1 > \mu_2$ One sided and $\mu_1 < \mu_2$ => $H_1: \mu_1 < \mu_2$ Two sided => $H_1: \mu_1$ not equal μ_2	○ 1 Side ◉ 2 Sides
Significance Level The Significance Level of the test or Prob(reject null hypothesis ($H_0: \mu_1 = \mu_2$) given it is true)	.05
Power The Power desired for the test or Prob(reject H_0 given that H_a is true)	.80
Submit Query	

references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)



The online calculator estimated that we need 63 observations in EACH group to achieve a power of .8 at alpha=.05, estimated effect size, .50.

Now, using our equation, we can plug in our numbers to compare results:

$$n = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2}$$

$$n = \frac{2(.80 + 1.96)^2}{.50^2}$$

$$n = \frac{2(.80 + 1.96)^2}{.50^2}$$

$$n = \frac{15.24}{.25}$$

$$n = 60.96$$

Note that our estimated sample size, 61, is a little smaller than the estimated sample size from the online calculator. The online calculator probably uses 2.0 to represent $z_{1-\beta}$, instead of 1.96. When we use 2.0 instead of 1.96, our formula yields 62.72.

Transformations

Nonnormal data can reduce the power of a parametric statistical procedure such as ANOVA and regression analysis. A major assumption of ANOVA and regression analysis is that your data are approximately normal (eg. bell-shaped). However, there are situations where you can **transform** your data to approximate normality. A transformation is any systematic alteration in a set of scores whereby certain characteristics of the set are changed and other characteristics remain unchanged. Reasons for performing transformations in analysis of variance often involve changing the distribution of the dataset, making it more normal.

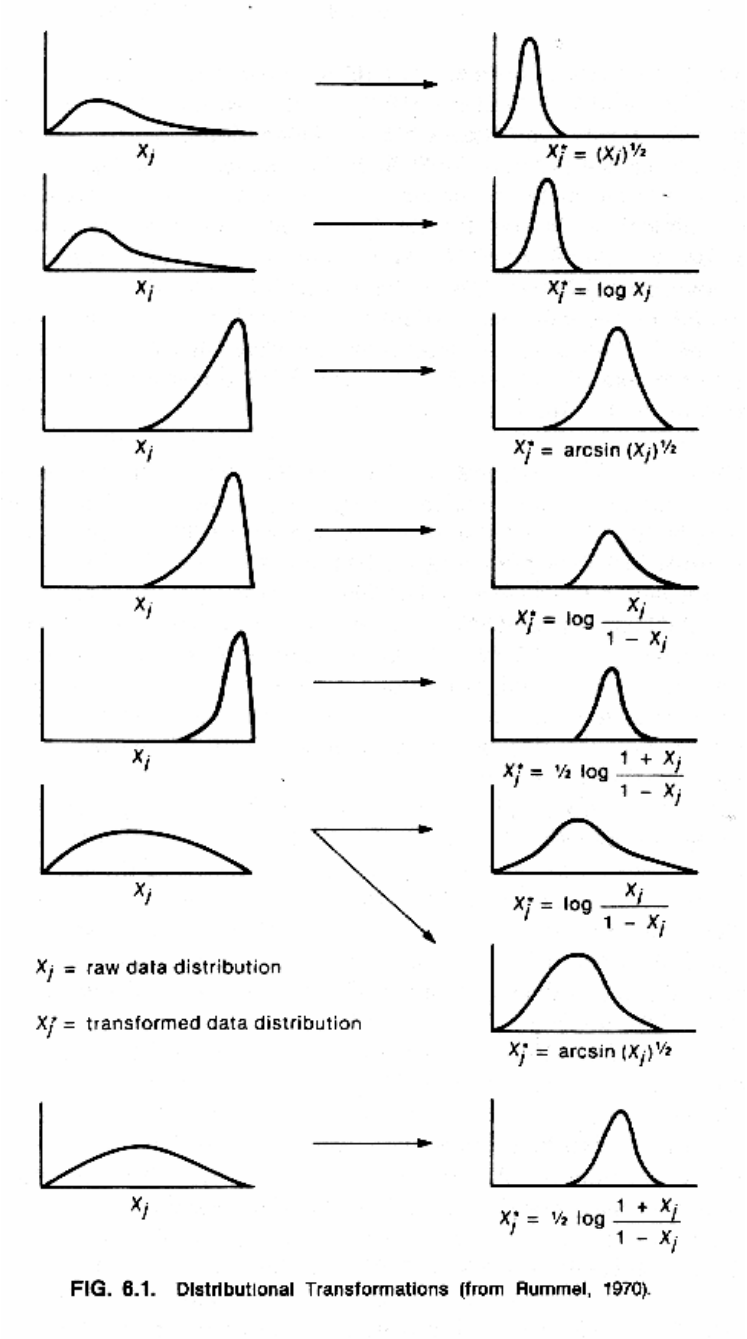


FIG. 6.1. Distributional Transformations (from Rummel, 1970).

references: Kirk (1988), Maxwell & Delaney (1990), Stevens (1996), Winer (1991)
Cipher 2006

This page intentionally left blank and not numbered

Chapter V
THE RESEARCH TEAM COLLABORATION, AND ETHICAL RESEARCH

SECTION 1 – Chapter V
INTRODUCTION

Chapter V, "The Research Team: Collaboration and Ethical Research" provides you with descriptions of the roles and responsibilities for each member of a team. Some teams may have more than one of these "types" of individuals. Principal Investigators vary in their management style. Unfortunately the errors of the past have inspired rules and regulations that may seem unnecessary and overly detailed for some of us. Some ethical dilemmas that researchers face may seem minor but there is ample information on the world-wide-web about ethical errors that have brought major universities to a halt in their research and caused millions of dollars to be repaid to the NIH due to irresponsible research management.

No matter what type of research you chose to conduct you will need to work with a Principal Investigator (PI). Remember that the word principal means "highest in rank or importance, chief, the head, someone who authorizes another as an agent to represent her or him." The word principle means "a fundamental doctrine or tenet, specific basis of conduct or management, or an accepted or professed rule of action or conduct." The PI is your pal.

Coordination means "harmonious or skillful combination or interaction as of functions or parts."

The main sections always include at least the Specific Aims, the Background and Significance, (your own, if any, Preliminary Studies), and Methods.

Multiple sources of information on the role and function of research teams and on collaboration and ethics in research are available. Some are provided at the end of this chapter. As a student you should learn to seek out these sources and use them to their maximum benefit.

SECTION 2 – Chapter V
THE RESEARCH TEAM

For a student or resident project the research team will normally include:

1. A Principal Investigator (not necessarily the student's mentor)
2. A student (or more than one) researcher (not a co-investigator)
3. A research coordinator or a staff person such as a nurse or medical assistant, or medical records clerk

Principal Investigator

The PI is the head of the team. She may delegate to other members of the team, the conduct of certain tasks, but may not delegate overall responsibility for the conduct of a study. The PI may have a Co-PI or a Co-Investigator. The PI may delegate certain tasks to the Co-I, but a Co-PI has pre-established leadership responsibilities that clearly do not duplicate or overlap with the PI. In other words, if there are multiple PIs for a project, they must write and sign documents that define their roles, scope, and functions for the study.

Responsibilities of the PI include at least the following ten functions.

1. Development, or supervision of the development of the project proposal
2. Final approval of the project proposal
3. Submission to the Institutional Review Board or Institutional Animal Care and Use Committee for approval
4. Compliance
5. Management of funds
6. Supervision of the team

7. Supervision of the data collection and management – analysis procedures
8. Establishment and performance of the Data and Safety Monitoring Board
9. Convening, chairing and documenting regular meetings of the full research team
10. Final review and submission of all reports

If there is a co-investigator the PI may delegate day to day supervision of the research coordinator, oversight of data collection and management, preparation and review of reports, and the leadership of the research team meetings. The PI may delegate spending authority but retains the responsibility for appropriate use of funds.

Student Researcher

1. Development and final draft review of the proposal (see chapters I, II and IV)
2. Development of IRB materials in consultation with a research coordinator or designated IRB liaison (see chapter II)
3. Coordination with site managers who govern access to information to be used in the study
4. Data collection forms (DCF) development in consultation with DCF design faculty
5. Data collection
6. Data entry if not blinded
7. Data analysis under the direction of the PI and a Ph.D. with appropriate experience in data analysis
8. Draft final reports for review and approval of the PI

Research Coordinator

1. Development of human subjects' materials including recruitment materials, risk precautions in consultation with the PI, and study files
2. Data collection
3. Compliance with IRB rules, regulations, and human subjects' research laws
4. File maintenance
5. Report development
6. Data entry if not blinded
7. Coordination with site managers who govern access to information for the research project

For studies that do not have a research coordinator, and the study involves touching or interacting with humans, or using information about humans, the student acts as the research coordinator. Even in the case of using existing data the IRB must be assured that the subjects have given permission for its use in research either by using a waiver procedure, or by demonstrating that the subject has previously given approval for its use.

Other team members usually include a person who creates the data entry forms or structures in the software to be used for the analysis, a person who enters data blinded or unblended, and a person who performs the technical functions of manipulating the data to answer the research question or test the research hypotheses.

SECTION 3 – Chapter V
COLLABORATION

The word collaborate means to work together in a joint effort or to collude, assist or abet another individual. Although the most productive research often emanates from a team rather than from one independent individual, collaborative projects students must make specific efforts to clarify each other's roles in the project. When students collaborate it is important that faculty understand that there is a distinction between the work each student will perform. Recognizing and leveraging strengths of each student is important. Equally important is that persons in authority clearly understand and approve the collaborative relationship. If two students share a project, each addressing a different perspective, there may be only one specific aims page, and only one background and significance and research design and methods sections, but different components of the data set will be analyzed by each student. In fact, while some data is shared, each student may collect different outcome variables.

Collaboration may be an effective use of resources if students or residents work together to maximize the human resources involved in collecting data.

An example would be three students collaborating on three different research questions about the same population. The population might be all patients at an OMM clinic whose physicians treat persons from infancy to 80 years old. One student might be interested in describing the health status and presenting chief complaint of a sample of all patients seen in the past five years. One student might ask what is the duration and frequency of utilization of the services by different age groups over the course of the past five years. The third student might explore differences in somatic dysfunction documented by region of the body by age groups in a sample of patients from any one year. In this case a sample from each year can be drawn stratified by clinically relevant age groups. All the data for each student's question can be collected at the same time, thus conserving time and resources and using consistent samples across several studies.

Another example would be three students from different fields of interest collaborating in a similar way. Suppose that one student is interested in geriatric patients, another in children, while another is interested in persons with disabilities. In a family medicine clinic that sees all these types of patients the students can agree on the data they wish to measure, such as quality of life using a questionnaire such as the SF36 or a shorter version. They might choose to examine services utilization by billing codes thus pulling charts using a stratified random method and the same DCF.

One additional example would be a situation in which a resident works with a large number of OMM procedures while two students do not have access to these types of patients. The questions might be: 1) Which conditions recorded on the osteopathic soap note are resolved fastest, 2) which OMM procedures are used the most frequently in the clinic in patients who also have COPD, and 3) is body mass index related to the somatic dysfunction findings, treatment frequency, or outcomes (BMI would have to be available in the record or this becomes an additional question to ask and the study becomes prospective). The resident can assist the students in identifying the size of the populations of interest, each study is an independent question, but enough of the data is shared across projects that one DCF could be used.

Students and residents can also collaborate among different sites, asking the same questions and increasing the number of subjects included in the data base. Data can be analyzed by site for between group differences in many different types of questions about treatment and treatment outcomes or service delivery conditions.

SECTION 4 – Chapter V
ETHICAL RESEARCH PRINCIPLES

Ethical means "being in accordance with the rules or standards for tight conduct or practice, or pertaining to a system of moral principles, or a branch of philosophy dealing with values related to human conduct with respect to the rightness and wrongness of certain actions, and the goodness and badness of the motives and ends of such actions." Ethical behavior may not be the same for all in the same situations because situations are inherently different for different individuals. Unethical is not wrong or illegal per se. Unethical is not adhering to moral principles commonly accepted or ruled-upon by society or a designated responsible group."

Is it illegal to tell a research participant that you do not believe that a particular OMT is effective? Could it be unethical? Is it illegal to screen charts for study subject criteria before you have IRB approval? Could it be unethical?

There is a difference between the Research Proposal and an IRB proposal. The IRB functions to protect the subjects in a study, and to protect the organization (the university, the hospital, and such) by ensuring ethical and legal practices in all research conducted by its faculty. The research proposal usually accompanies the IRB proposal that focuses on inclusion/exclusion of protected groups of persons, risks, and protections for subjects in a study. An IRB outline and sample clinical trial proposal is included in Chapter II.

Questions dealing with ethical dilemmas usually need discussion. What may seem obviously unethical to one person may not seem so to another. Critical thinking skills help you determine what behaviors or actions to use in various situations you may encounter in research or in other human interactions.

There are three main principles for research.

1. Respect for persons
2. Beneficence
3. Justice

Respect for persons incorporates at least two ethical convictions: 1) individuals should be treated as autonomous agents; 2) persons with diminished autonomy are entitled to protection. The principle of respect for persons thus divides into two separate moral requirements: the requirement to acknowledge autonomy and the requirement to protect those with diminished autonomy.

Beneficence: Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being. Such treatment falls under the principle of beneficence. The term "beneficence" is often understood to cover acts of kindness or charity that go beyond strict obligation. In this document, beneficence is understood in a stronger sense, as an obligation. Two general rules have been formulated as complementary expressions of beneficent actions in this sense: (1) do not harm and (2) maximize possible benefits and minimize possible harms.

Justice: Who ought to receive the benefits of research and bear its burdens? This is a question of justice, in the sense of "fairness in distribution" or "what is deserved." An injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly. Another way of conceiving the principle of justice is that equals ought to be treated equally. However, this statement requires explication. Who is equal and who is unequal? What considerations justify departure from equal distribution? Almost all commentators allow that distinctions based on experience, age, deprivation, competence, merit and position do sometimes constitute criteria justifying differential treatment for certain purposes. It is necessary, then, to explain in what respects people should be treated equally. There are several widely accepted formulations of just ways to distribute burdens and benefits. Each formulation mentions some relevant property on the basis of which burdens and benefits should be distributed. These formulations are (1) to each person an equal share, (2) to each person according to individual need, (3) to each person according to individual effort, (4) to each person according to societal contribution, and (5) to each person according to merit.

(resource: http://ohsr.od.nih.gov/guidelines/belmont.html#gob)

Examples of Ethical Dilemmas

Example 1

The first annual continuing review report to the IRB is late and as a result the board does not review it timely such that the approval is no longer valid at one of the facilities that govern access to data used in the study or access to patients who were being screened for the study. Your study is approved for three years.

You:

a. continue to screen anyway because you can do it by remote access to the hospital computer
b. call the IRB and ask for an exception to the rule
c. ask someone on the staff of the hospital to screen patients for you
d. stop the study until the board can review the continuing review report

(answer is: "d. stop the study until the board can review the continuing review report")

Example 2

A former student at your school collected a lot of information from medical records at one of the clinics when she did her research project. All the information is in an SPSS data base and you have it because she left it on the hard drive of the department's notebook computer that you are using for your project. You know she did not use all the data in her research project analysis and report.

You:

a. Don't tell anyone and use it as existing data and do not submit any request to the IRB because it was already approved for collection when the other student collected it
b. Immediately contact the person who was the PI for that study and let him know about the data base being on the computer
c. Erase the data base from the computer by putting it in the "recycle bin" and emptying it from that file
d. Call the IRB office director and report the data base find
e. E-mail the former student and ask for permission to use her data in your study.

(answer is: "b. immediately contact the person who was the PI for that study and let him know about the data base being on the computer." This is so he can first decide what action needs to be taken per the protocol for confidentiality of research data; second, consult with IRB officials; third, possibly audit the files from that study, or if it is an open and active study, he could add you to the IRB protocol as approved key personnel; and fourth as a courtesy, inform the former student of the future use of the data particularly if she has published using that data)

Example 3

Your study, approved by your supervisor, will examine the attitudes of first and fourth year medical students toward the clinical effectiveness of a particular OMM treatment technique. You have created a survey of five questions on a likert-type scale and you have included space for the student to enter age, gender, year of graduation, and race and ethnic group on the form. You also want to use rank in class as a classification variable.

Your next step is to

a. contact the student association to distribute and collect the survey for you
b. contact the office that keeps student grades to get a list of students and their class ranking
c. add "class rank" to the survey form and let the student give it to you
d. create a consent section for the survey form in consultation with an IRB liaison or staffer

(answer is: "d create a consent section for the survey form in consultation with an IRB liaison or staffer")

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter VI
OMM RELIABILITY AND VALIDITY STUDIES

INTRODUCTION

The most pragmatic research to engage in during osteopathic medical education and training is ongoing research that is of interest to you. However, if you are interested in engaging in OMM research that you have most control over the time put in, the data, its analysis, and want to make the most of your time on the project, it might be most prudent to engage in a quality assessment and improvement study, such as the Clinical Assessment Program. Another option is to do a diagnostic reliability study or a validity study.

The International Federation of Manual Medicine (FIMM) Scientific Committee developed a guideline for how to do reliability or validity studies that is very useful.

REPRODUCIBILITY AND VALIDITY STUDIES
of
Diagnostic Procedures in Manual/Musculoskeletal Medicine
Protocol formats
THIRD EDITION
FIMM SCIENTIFIC COMMITTEE
Editor: J. Patijn, MD, PhD

This document in its entirety has been reproduced as Chapter VI in this manual. The entire text is included in this research manual to help researchers and mentors who wish to embark along this path.

Indeed, before clinical trials are performed in manual and manipulative medicine, reliability and validity research is important to ensure reproducibility and accuracy of diagnostic findings, and post-OMT re-evaluation palpatory assessments.

(Chapter created by M. Kuchera, D.O., March 2007)

FIMM

REPRODUCIBILITY AND VALIDITY STUDIES

of
Diagnostic Procedures in Manual/Musculoskeletal Medicine

Protocol formats

THIRD EDITION

FIMM SCIENTIFIC COMMITTEE
Editor: J. Patijn, MD, PhD

Preface to the third Reproducibility and Validity Protocol

Based on an internal discussion within the Scientific Committee (SC) of the International Federation for Manual/Musculoskeletal Medicine (FIMM), a third protocol became necessary. It became clear that the second protocol showed shortcomings with respect to other aspects of kappa statistics, such as the weighted kappa, the value of the significance and confidence intervals of the kappa. New research results such as the effect of education on the kappa value and the proof of the 50% method to influence the prevalence in advance are incorporated in the present protocol. It became clear that there was a need to describe inappropriate statistics used in reproducibility studies in Manual/Musculoskeletal Medicine (M/M Medicine). More attention is paid to the different kind of characteristics of data collected in reproducibility studies, such as ordinal, nominal and interval or continuous data.

Based on experience with recent kappa studies, the format of the reproducibility protocol is adapted in several aspects of the different phases of a study.

This SC protocol in particular emphasises the kappa method for reproducibility studies of diagnostic procedures in M/M Medicine. For these kind of studies a "Cook Book" format is presented in a very practical way to make it available for both clinics with two or more physicians in M/M Medicine and Educational Committees of National Societies to perform these kind of studies.
For university departments more in-depth information about statistics in every kind of reliability studies is provided in this SC protocol.

The Scientific Committee of the FIMM is aware that developing this kind of protocols is a continuous process.
By publishing the third protocol on the website of the FIMM, the Scientific Committee hopes that those scientists who use this protocol will send their comments to the Chairman of the Scientific Committee. In this way, we hope to improve the present protocol.

The SC asks those scientists who receive this protocol to distribute this protocol to their fellow scientists. In this way, the protocol

becomes accessible for all practitioners in the field of M/M Medicine.
This protocol is the end product of the energy of all members of the SC.

Dr. Jacob Patijn, MD, PhD, Neurologist,
Physician for Manual/Musculoskeletal Medicine
Chairman of the Scientific Committee of the FIMM
Responsible member for the Reliability Group of this Committee

SCIENTIFIC COMMITTEE FIMM

Address for reprints and comments:
Dr. Jacob Patijn, MD, PhD, Neurologist
University Hospital Maastricht, Pain Management and Research Centre, Dept. Anaesthesiology, Maastricht, The Netherlands, Fax: 31 43 3875457, E-mail jpat@sane.azm.-nl, jacobpatijn@hccnet.nl

I. INTRODUCTION CHAIRMAN SCIENTIFIC COMMITTEE

I.1 Background

This is the fifth of the protocols published by the Scientific Committee (SC) of FIMM.
Its concerns a standardised format for validity, sensitivity and specificity studies. Besides, it provides the scientist as well as the daily practitioners in our field in more or less cook book form with a format for reproducibility studies in Manual/Musculoskeletal Medicine.
In the future, continuously improved scientific committee protocols will be developed.
The reason of the SC to develop these kind of protocols has been extensively discussed in previous reports of the SC for the General Assembly and has been published in FIMM NEWS. In different countries, the previously published protocols gradually have led to reproducibility studies in M/M Medicine.
The primary reason to develop this kind of protocols by the SC is still actual. Therefore, as we did in previous protocols, a short background is provided of these protocols and a brief overview of the past SC activities is included.
The Scientific Committee of FIMM (SC) formulated the problem with respect to diagnostic procedures in Manual/Musculoskeletal Medicine (M/M Medicine), and it is summarised in the statement:

There are too many different schools in Manual/ Musculoskeletal Medicine in many different countries of the world, with too many different diagnostic procedures and too many different therapeutic approaches.

The consequences of this statement are five-fold:

I.1 Most schools within M/M Medicine have not validated yet their own characteristic diagnostic procedures in the different regions of the locomotion system. Therefore reproducibility, validity, sensitivity and specificity of these diagnostic procedures are still lacking.

I.2 All the different schools within M/M Medicine still coexist. Because of lack of good reproducibility, validity, sensitivity and specificity studies, mutual comparison of diagnostic procedures is impossible. Scientific information exchange and fundamental discussions between these different schools, based on solid scientific methods, are hardly possible in the present situation.

I.3 Absence of validated diagnostic procedures in M/M Medicine leads to heterogeneously defined populations in efficacy trials. Therefore, comparison of efficacy trials, with the same therapeutic approach (for instance manipulation), is impossible.

I.4 If the present situation is allowed to continue, it will lead to a slowing down of the badly needed process of professionalisation of M/M Medicine.

I.5 Non-validated diagnostic procedures of different schools, ill-defined therapeutic approaches and low quality study designs are the main causes for the weak evidence of a proven therapeutic effect of M/M Medicine.

It is still the opinion of the SC that the committee should create conditions for exchange of scientific information between the various schools in M/M Medicine. This information exchange must be based on results of solid scientific work. By comparing the results of good reproducibility, validity, sensitivity and specificity studies, performed by different schools, a fundamental discussion will arise. The main aim of this discussion is not to conclude which school has the best diagnostic procedure in a particular area of the locomotion system, but to define a set of validated diagnostic procedures which can be adopted by the different schools and become transferable to regular medicine.
The SC wants to provide the National Societies of FIMM with standardised scientific protocols for future studies.

The SC thought that the best forum for creating a discussion platform would be to organise every other year a SC Conference in cooperation with a particular National Society. The SC Conference was organised in Odense, Denmark, 2003, in cooperation with the Danish Society for Manual Medicine. Many researchers presented their preliminary results, proposals for protocol formats and therapeutic algorithms. In a fruitful discussion between audience and presenters many ideas were exchanged based on solid scientific work, without interference of “school politics”.

As Chairman of the SC, I want to emphasise that good reproducibility, validity, sensitivity and specificity studies still have the first priority. This kind of studies is easy and cheap to perform, and they form the best base for mutual discussion between schools in M/M Medicine.
Co-operation and active involvement of the National Societies of FIMM is indispensable and crucial for the future work of the SC.

In providing this third protocol to the National Societies of FIMM, the SC hopes to add a substantial contribution to the professionalisation of M/M Medicine.

Dr. Jacob Patijn, MD, PhD, Neurologist

II. REPRODUCIBILITY AND VALIDITY

Nomenclature

One of the major problems in medicine and in research is the fact that different names are used for the same definition Therefore we thought it important first to provide the reader of this protocol with an overview of the definitions used in this protocol. In clarifying the definitions in advance we hope to make reading easier.

II.1 **Reliability** can be divided in **Precision** and **Accuracy.**

II.1.1 **Precision,** also called **Reproducibility**

In the case of reproducibility of an observation made by one observer on two separate occasions, we call it the *intra-observer variability* or the *intra-observer agreement.*

In the case of reproducibility of an observation by two observers on one occasion, we call it the *inter-observer variability* or the *inter-observer agreement.*

In this protocol, we use the terms **reproducibility, intra-observer agreement** and **inter-observer agreement.**

Reproducibility of diagnostic procedures in M/M Medicine evaluates whether two observers find the same result of a diagnostic procedure in the same patient population, or whether a single observer finds the same result of a diagnostic procedure in the same patient population on two separate moments in time.

II.1.2 **Accuracy,** also called **Validity**

In this protocol, we use the term **validity.**

Validity measures the extent to which the diagnostic test actually does what it is supposed to do. More precisely, validity is determined by measuring how well a test performs against the gold or criterion standard.
When a diagnostic test has to be evaluated with respect to what it is supposed to do (validity), a gold standard as reference is needed. This is a major problem not only in M/M Medicine but in the whole medical profession. Sometimes, radiological findings, post-mortem findings or findings during an operation can act as gold standard. In the case of subjective quantification of range of motion, the gold standard can be the result of a quantitative method performed in a normal population. Frequently, it is only possible to define a gold standard by consensus of experts in a particular field of medicine.

Gold standards are needed for estimation of the sensitivity and specificity of a test (see V.1).

II.2 **Index Condition and its Prevalence**

II.2.1 The **index condition** is synonymous with the diagnosis of a patient. This diagnosis must be based on reproducible diagnostic procedures with a proven validity.
In case of reproducibility studies of diagnostic procedures, a positive judged test by observers is called the index condition.

II.2.2 The **prevalence of the index condition** is the frequency of the index condition in a particular population at a particular moment. In reproducibility studies of tests, the prevalence of the index condition is only related to the study population.
It is essential to realise that the prevalence of an index condition can vary in different institutes, countries and can change in time.

In the reproducibility section of this protocol, we will use the terms **index condition** and **prevalence of the index condition in relation to positive found test procedures.**

In the 2 x 2 contingency table hereunder, a theoretical example of the results of a reproducibility study of two observers A and B is shown.

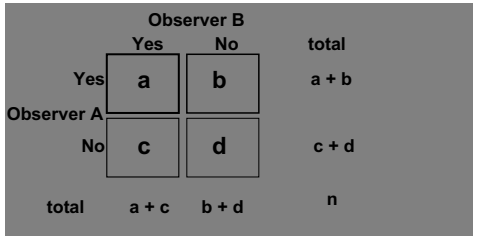


Figure 1. 2 x 2 contingency table

The squares with a and b represent the number of patients with positive tests as judged positive by observer A. The squares with a and c represent the number of patients with positive tests as judged by observer B. The squares a, b and c represents the number of patients with positive tests as judged by either one or both observers among the total patients n.
The prevalence is calculated by the formula for the prevalence (P):

$$P = \frac{[a + (b + c)/2]}{n} \qquad \textbf{(formula 1)}$$

II.3 **Overall Agreement**

The overall agreement reflects the percentage of the patients in which both observers A and B agree about the judgement of the

test. Based on figure 1, both observers agree in a and d (respectively positive and negative). In the squares with b and c, the observers disagree.
Overall agreement P_o is calculated by the formula:

$$P_o = \frac{[a + d]}{n} \qquad \textbf{(formula 2)}$$

II.4 Sensitivity and Specificity

II.4.1 The **sensitivity** of a test is defined as the proportion of the cases that have the index condition that the test correctly detects.

II.4.2 The **specificity** of a test is defined as the proportion of the cases that do not have the index condition that the test correctly detects.

In this protocol the so-called "Nosographic Sensitivity and Specificity" is identical with the terms "Sensitivity and Specificity".

II.4.3 To translate the statistics of sensitivity and specificity figures into daily practice, the physician has to know whether a positive test in the individual patient is truly positive as opposed to false-positive. This is expressed respectively as the so-called "*positive predictive value* of a test" and "*negative predictive value* of a test".

In contrast to the "Nosographic Sensitivity and Specificity", the positive predictive value of a test and negative predictive value of a test are also called the "Diagnostic Sensitivity and Specificity".

In this protocol the so-called "Diagnostic Sensitivity and Specificity" is identical with the terms "**positive and negative predictive value** of a test".

II.5 Kappa Value: Interpretation

In this protocol, kappa statistics will be the method of choice for reproducibility studies (see below).
Kappa value is a statistical measurement for the intra-observer and inter-observer agreement corrected for chance. The kappa value can be either negative or positive and ranges between –1 and +1.
Several schemes are available (see Haas 5,6,12) to draw the line on good agreement. The most widely used scheme is that of Landis and Koch. They stated that kappa values above 0.60 represent good to excellent agreement beyond chance between two raters. In contrast, kappa values of 0.40 or less represent poor agreement beyond chance. Kappa values between 0.40 and 0.60 reflect a fair to good agreement beyond chance.

Bogduk uses a kappa value of 0.4 as cut off level of good agreement.
In this protocol we use a conservative kappa value cut off level 0.6, reflecting a good to excellent agreement.

III. STARTING POINTS IN REPRODUCIBILITY PROTOCOL OF DIAGNOSTICS IN M/M MEDICINE

To perform reproducibility studies for diagnostics in M/M Medicine, several points are important to consider to start with.

III.1 Character of the Diagnostic Procedure and Statistical Methods

Before starting a reproducibility study in M/M Medicine, it is important to be clear about what kind of diagnostic procedure we are dealing with and what kind of statistics are appropriate.
In general we have two kinds of diagnostic procedures: a. Qualitative Diagnostic Procedures, b. Quantitative Diagnostics Procedures.

III.1.1 Qualitative and Semi-Quantitative Diagnostic Procedures (Nominal and Ordinal Data)

Qualitative diagnostic procedures in M/M Medicine are characterised by subjective outcomes of observer and/or patient. These kinds of procedures can have both a nominal or an ordinal character. Typical examples of this kind of procedure in M/M Medicine are end feeling and pain provocation under different conditions (provoked by observer, provoked by movements of the patient). In case of existence or absence of a finding (Yes/No), for example pain provoking tests, we are dealing with nominal data and kappa statistics indicated. If different categories (with a natural order) of a test procedure can be distinguished, for example: no end feel, soft end feel and hard end feel and very hard end feel, we are dealing with ordinal data, and weighted kappa statistics are indicated. Also, semi-quantitative diagnostic procedures in M/M Medicine are in essence a qualitative diagnostic procedure with a dichotomous character. Typical examples of these kinds of semi-quantitative diagnostic procedures in M/M Medicine are measurement of left/right difference in subjective range of motion of the examiner (difference in range of motion, Yes or No or restricted motion, Yes or No).

III.1.2 Quantitative Diagnostic Procedures

In quantitative diagnostic procedures, mostly measured with a certain kind of device, findings are quantified in degrees, millimetres, kg etc. and are mentioned interval or continuous data.
For these kind of quantitative procedures normative values are needed. First a study of the procedure in normal subjects is

needed in which the reproducibility of the procedure has to be estimated in the same population on two different occasions. In this test/retest study, the systematic measurement failure can be estimated based on the distribution of the data values. Besides, factors such as age and gender, which can influence the data, have to be studied. Quantitative diagnostic procedures can serve as gold standard for semi-quantitative diagnostic procedures.

In reproducibility studies of any kind, the nature of the collected data (nominal, ordinal, interval or continuous) is decisive for the applied statistical method.

III.1.3 Inappropriate Statistics in Qualitative Data Reproducibility Studies

Frequently, inappropriate statistics are applied to measure the reproducibility. The main flaw is that agreement is often confused with trend or association, which is the assessment of the predictability of one variable from another. Hereunder the flaws of several statistical methods in reproducibility studies are listed.

III.1.3.1 Percent Agreement

Reproducibility studies, just mentioning the *percent agreement,* give no real information about the reproducibility. *Percent agreement* is the ratio of the number of subjects in which the observers agree to the total number of observations. The main problem is that the *percent agreement* does not take into account the agreement that is expected to occur by chance alone.

III.1.3.2 Correlation Coefficients

In many reproducibility studies correlation and association measures are used to evaluate the reproducibility of clinical data. The problem is that some do not have the ability to distinguish trend toward agreement from disagreement (Chi-Square [^{2}] and Phi) or do not account for systematic observer bias (Pearson's product moment correlation, Rank order correlation).

III.1.4 Appropriate Statistics in Qualitative and Semi-Quantitative Data Reproducibility Studies

III.1.4.1 *Normal Kappa* is the statistics of choice for evaluating reproducibility between two observers for nominal (dichotomous) data.

III.1.4.2 In case of many observers (>2) the *overall kappa* can be used to generalise the results to broader populations of observers. For example, evaluating the existence of segmental dysfunctions in a particular area as indication for therapy, the overall kappa would

give an estimate of the overall reproducibility to detect segmental dysfunctions by observers in that particular area. For details see textbooks or ask your statistical expert.

III.1.4.3 In M/M Medicine the judgement of a diagnostic procedure can be subdivided into different grades, such as end feel (normal, elastic, hard). These ordinal data must have a natural order. In reproducibility studies with ordinal data the statistics of *weighted kappa* is indicated. For details see textbooks or ask your statistical expert.

III.1.4.4 Significance of the found kappa value, together with confidence intervals, can be calculated, in case of kappa values between 0.40 and 0.60. It provides you with the information whether the found kappa value differs from chance. In case of kappa values over 0.60, this procedure is not necessary.

The ins and outs of normal kappa statistics is elaborated more in detail below (see III).

III.1.5 Appropriate Statistics in Quantitative Data Reproducibility Studies

To evaluate the reproducibility of measurements with quantitative data (interval or continuous data) in repeated measures, the paired t-test is indicated.
One-way analysis of variance intraclass coefficient (ANOVA ICC) is the statistical method of choice for the reproducibility of observers for interval data (cm, mm, etc). The calculated factor R in this statistical procedure is 1 if there are identical ratings, less than 0 in absence of reproducibility. A limitation of the ICC is that it provides no information about the magnitude of disagreement between observers.

In reproducibility studies, the choice of statistics should depend not only on the character of the collected data (nominal, ordinal, interval), but also on the related type of clinical decision concluded from the findings of the study.
For instance, if one needs the findings of the study to decide whether or not a heel lift is indicated to correct leg length inequality, ANOVA ICC statistics for interval data are indicated. In contrast, if leg length differences are measured to adjust pelvic adjustment, the data characteristics are right, left and equal and therefore kappa statistics are indicated for the nominal data. The same is true for semi-quantitative data such as the side of restricted range of motion Yes or No.

III.2 Aim of the Diagnostic Procedure

In studying the reproducibility of diagnostic procedures in M/M Medicine, one has to be clear about the aim of the test(s). It is essential to realise the difference between a diagnosis, a syn-

drome and a diagnostic test used in daily practice. In a genuine diagnosis, the aetiology and prognosis is known. In syndromes, a combination of signs and symptoms that appear together in a high frequency in a certain population, the aetiology is unknown.
In both diagnosis and syndromes, diagnostic tests are needed. A diagnostic test is a procedure, performed by a clinician, to objectify in a qualitative way a clinical finding which is frequently not mandatory with a genuine diagnosis. For example the combination of sensory deficit, motor deficit and a positive Lasègue can be characteristic for a radicular syndrome. The aetiology can be as well an intervertebral disc protrusion as a tumour in the intervertebral foramen, both with root compression.
In M/M Medicine educational systems, many tests are taught to the student as a procedure, for instance passive cervical rotation. The student just learns how to perform the whole procedure of passive cervical rotation (setting of the hand, applied force etc.). The explanation for such a restriction can have many reasons and therefore gives no information about a diagnosis.
Therefore, the first priority is to make the procedures with their judgements of all kinds of tests in M/M Medicine reproducible. In second instance find gold standard to validate these procedures. For example: the finding of a restricted cervical rotation (Yes or No) must be validated by a quantitative method with a specially designed device that measures the rotation in degrees in different age and gender groups.
Subsequently, reliable tests (reproducible and validated) can be use to define syndromes in M/M Medicine.
Finally, and often very difficult, gold standards have to be found for validation.

III.2.1 Evaluating a single diagnostic test only gives information about the reproducibility of the whole *test procedure*.
In the vast majority of single diagnostic tests, no information is obtained about a specific diagnosis based on that single diagnostic test and consequently no indication for a specific therapy is provided.
Therefore, a single diagnostic test seldom differentiates between normal subjects and patients. In general, in the absence of a gold standard, sensitivity and specificity studies are useless if they are based on a single reproducible diagnostic test.

III.2.2 Evaluating a combination of test procedures gives information no more than the reproducibility of these combinations of the tests. Without a gold standard, reproducible combinations of tests have no diagnostic value and can be seen in specific diagnosis and non-specific pain syndromes, but also in normal subjects. The disadvantage of testing many tests at the same time in reproducibility studies is the potential mutual dependency of the tests.

III.2.3 Reproduction of a test in time (perform the same diagnostics in the same patient after a time interval) can be used to estimate the sensitivity and specificity of a test. Such tests, when combined with other clinical data, can increase the ability to differentiate between patients and normal subjects. However, in the vast majority of cases, no information is obtained regarding a specific diagnosis based on this combination. In general, it is only in the presence of a gold standard that it will be useful to perform sensitivity and specificity studies, based on a combination of valid test procedures.

III.3 Number of Tests to be Evaluated

Reproducibility studies in non-specific, for instance in low back pain, sometimes show evaluation of reproducibility of a large number of tests at the same time. In this kind of studies, many of the tests show low kappa values and therefore are judged of no clinical importance by the authors. Since prevalence and overall agreement figures are frequently lacking, a definite conclusion about the reproducibility of the tests with low kappa values cannot be drawn. Heterogeneous study populations consist most probably out of different subgroups each with different prevalences of the tests to be evaluated. This can result in the risk that some positive tests have a low prevalence in the study, because of a small size of that particular unknown subgroup.
The tests to be evaluated must have a relation with the characteristics of the study population. For example, evaluating the reproducibility of several radicular provocation tests in LPB patients without any signs of sciatica has no sense, because it is to be expected that positive radicular tests are rare in such a population.

In case of a population with sciatica, evaluating the reproducibility of several radicular provocation tests at the same time, one can decide on a minimal number of positive tests which is needed to make the diagnosis of a lumbar radicular syndrome. The disadvantage of evaluating a combination of tests for a particular diag-nosis (for example radicular syndrome, SI-dysfunction) is that there is a chance for mutual dependency.

For example, many SI-tests in M/M Medicine are supposed to test a SI-dysfunction or hypomobility of the SI-joint. This mutual dependency was shown in a reproducibility study of six SI-tests at the same time (Deursen van, Patijn). In this study, three observers (A, B, C) were supposed to use six different SI-tests (I to VI) for the final SI-diagnosis (see figure 2 and 3).
To evaluate the mutual dependency of the tests, for each observer, the kappa values were calculated of the fifteen possible combinations of pairs of their six SI-tests.

Test	Obsv.	I	II	III	IV	V	VI
I	A						
	B						
	C					Kappa values	
II	A	-0.09					
	B	+0.02					
	C	+0.36					
III	A	+0.25	-0.01				
	B	+0.34	+0.17				
	C	+0.36	+0.22				
IV	A	+0.34	-0.29	+0.25			
	B	+0.06	-0.05	+0.15			
	C	+0.22	-0.01	+0.36			
V	A	**+0.61**	-0.12	+0.28	**+0.43**		
	B	+0.33	+0.39	+0.34	+0.01		
	C	+0.10	+0.21	+0.21	+0.32		
VI	A	**+0.61**	-0.22	+0.18	**+0.43**	**+0.89**	
	B	+0.23	+0.19	+0.21	-0.15	**+0.52**	
	C	+0.21	+0.32	+0.24	+0.27	**+0.84**	

Figure 2. Mutual dependency of six SI-tests (I till VI) in three ob-servers A, B and C. The bold kappa values >0.40 reflect a mutual dependency.

Using a kappa of 0.40 as lowest level, figure 2 shows in different pairs of tests in all three observers a kappa value larger than 0.40. In particular between test V and VI, all observers showed high kappa values (+0.89, +0.52 and +0.84), reflecting a mutual dependency between test V and VI.
This means that all three observers unconsciously judged SI-test VI positive after they had judged SI-test V as positive. In this study, SI-tests II, III versus I, IV and VI show mutual independency (second and third column).

This aspect of mutual dependency is also very important in reproducibility studies when selecting tests for the same clinical feature/diagnosis. In kappa studies, besides evaluating the reproducibility of the tests themselves, the interobserver agreement of the final diagnosis, based on these tests, can be evaluated.
From the same study, as mentioned above, it became clear that with too many tests, observers use only a few tests for their final SI-diagnosis. By calculating the mutual kappa value of the single tests (I to VI) and the final diagnosis in all three observers A, B, and C this phenomenon is illustrated (see Figure 3).

SI-Tests	Obsv.	I	II	III	IV	V	VI	SI-Diagnosis Kappa
I	A							-0.61
	B							+0.23
	C							+0.21
II	A	-0.09						-0.22
	B	+0.02						+0.19
	C	+0.36						+0.32
III	A	+0.25	-0.01					+0.18
	B	+0.34	+0.17					+0.21
	C	+0.36	+0.22					+0.24
IV	A	+0.34	-0.29	+0.25				**+0.43**
	B	+0.06	-0.05	+0.15				-0.15
	C	+0.22	-0.01	+0.36				+0.22
V	A	+0.61	-0.12	+0.28	+0.43			**+0.89**
	B	+0.33	+0.39	+0.34	+0.01			**+0.52**
	C	+0.10	+0.21	+0.21	+0.32			**+0.84**
VI	A	+0.61	-0.22	+0.18	+0.43	+0.89		**+1.00**
	B	+0.23	+0.19	+0.21	-0.15	+0.52		**+1.00**
	C	+0.21	+0.32	+0.24	+0.27	+0.84		**+1.00**

Figure 3. Mutual dependency of six SI-tests (I till VI) with the final SI-diagnosis in three observers A, B and C. The bold kappa values > 0.40 reflect a mutual dependency.

Note that in the far right column “SI-Diagnosis”, all three observers only use SI-test V and VI for their final judgement of the SI-diagnosis. In all three observers A, B and C SI-tests I to IV contributed not at all to their final SI-diagnosis.
In general it is advisable to evaluate a maximum of three tests for the same clinical feature. It is advisable to choose tests each with a completely different procedure and not related to a single joint.

III.4 Number of Observers

There is no real statistical reason for performing a reproducibility study with more than two observers. In some studies, more observers are involved to evaluate the effect of the observers’ experience on the interobserver agreement. The problem with experienced observers is that they probably have developed a personal performance and interpretation of the test. Most of these studies lack a proper training period for standardisation of the performance of the test procedure and its interpretation. The results of these kinds of studies inform us more about the skills and/or the quality of the educational systems of the observers, rather than about the reproducibility of the evaluated tests. The same is true for reproducibility studies which estimate kappa values of tests done in the so-called “in-vivo condition”, in which no standardisation of the test procedures was carried out (to mimic the daily practice of a test).
The only case in which more observers can participate in kappa studies is to evaluate the effect of regular training on the kappa value. The same observers are repeatedly trained in a diagnostic

procedure and after each training period a new kappa is estimated to see whether a rise in kappa value in observers is seen.
In principle reproducibility studies, using the proposed format as discussed below, provide us with the potential reproducibility of a test procedure. If the reproducibility of a test procedure is established, a second study can be performed to evaluate the effect of observers' characteristics on the reproducibility.

A second flaw of using too many observers in a reproducibility study is the possibility of a therapeutic effect of the test procedure. If in a single patient, a passively performed procedure (passive cervical rotation) is performed too many times by different observers in a row, a therapeutic effect of the procedure may influence the range of motion and therefore the results of the last observer.

In general, using the proposed format in this protocol, two observers are sufficient to estimate the potential reproducibility of a test.

III.5 Hypothesis of a Test

It is very important for a reproducibility study of a test to discuss and analyse what the test is supposed to test. For range of motion there is no problem. For mobility, for instance hypomobility of the SI-joint, there is a problem. In many reproducibility studies of the SI-joint, the hypothesis for the various tests was that they were supposed to test the mobility of the SI-joint. Although SI-mobility is proven, based on cadaver studies, it is impossible, even for the most experienced observer, to test manually the mobility of the SI-joint. This incorrect belief is probably the reason for the low kappa values of SI-tests in the literature. Looking critically at the substantially different procedures of the large number of SI-tests, we have to question whether all these procedures can test the hypomobility of the SI-joint. In reproducibility studies, the observer has to forget the hypothesis of the tests taught by his teachers and has to concentrate on all the different aspects and details of the test procedure as such. For instance, according to the literature, the Patrick test for the SI-joint is supposed to test the mobility of a SI-joint. Looking critically at the test procedure, the observers can decide that the Patrick test, measuring end feeling and motion restriction, only evaluates increased muscle tension of a certain group of muscles related to the hip joint.
The effect of the hypothesis for the reproducibility on SI-tests was illustrated in two studies (Patijn 2000). The first study, which assessed six SI-tests supposed to evaluate SI-mobility, resulted in very low kappa values. In the second study, three tests supposed to test muscle hypertonia and its consequent motion restriction, in different muscle groups around the lumbosacral-hip region, resulted in a kappa value of 0.7.
Whatever tests one selects for a reproducibility study, one has to investigate step by step the whole test procedure and agree about what the test really tests.

Based on this agreement, the observers can define a more plausible hypothesis for the test, which can completely contradict the hypothesis stated in the literature.
Full agreement of the observers about a more plausible hypothesis of a test can lead to better results in reproducibility studies. In reproducibility studies these aspects are essential in the training period of the study format (see figure 10, page 27).

III.6 Blinding Procedures

In every reproducibility study, blinding procedures are essential not only for the patient/observer condition but also for both observers and must be well defined. Be sure that during the study there is no communication between observers, use separate forms for the observers to record their findings. If necessary, be sure there is no communication between observer and the patients.

III.7 Test Procedure and Test Judgement

As already argued under item 5, the observers have to standardise the whole test performance and the way they judge the result of a test. In the protocol format discussed below (see figure 10, page 27), the training period is essential for standardisation in a reproducibility study. The consensus about the definition of the test procedure and its assessment must be discussed in the final publication. To prevent observers' "personal interpretation" during the study, we also advise that the standardised procedures and test assessments are printed on the forms used in the study.

III.8 Selection and Number of Subjects

In reproducibility studies, the primary source population out of which the subjects are selected must be defined and mentioned in the final publication. Selection procedures must be very clear.
In general, for simple reproducibility studies 40 subjects are sufficient. This number of subjects makes this kind of reproducibility study easy and cheap to perform and not restricted to large institutes.

III.9 Statistics in Reproducibility Studies: the Kappa Value

In reproducibility studies with two observers evaluating dichotomous tests (Yes/No), estimation of the kappa values is the method of choice (see below).

III.9.1 Kappa Dependency on Prevalence

In the literature many reproducibility studies judge diagnostic tests with kappa values below 0.6 as clinically irrelevant. However, in the

vast majority of reproducibility studies no information is pre-sented about the corresponding prevalence and overall agreement of the index condition. This is essential, because the kappa value is dependent on the prevalence and the overall agreement.
Published reproducibility studies which present evaluations of tests with low kappa values, as clinically worthless or of minor importance, without mentioning any figures about prevalence and overall agreement, are misleading.

Low kappa values can reflect high as well as low prevalences!!!

Figure 4 shows the dependency of the kappa value on the prevalence.

Note that in case of very low (a) and very high prevalences (b) the kappa value becomes very low.

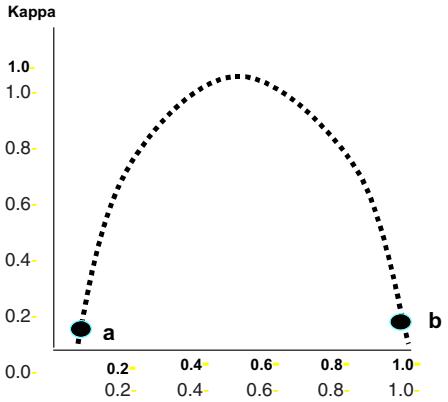


Figure 4. Relation between kappa values and prevalences

III.9.2 Kappa Dependency on Overall Agreement (P_o)

In figure 5 it is illustrated that with a high overall agreement (0.98 in the figure) the maximal kappa value is 1.0 and the minimal kappa value is nearly 0.

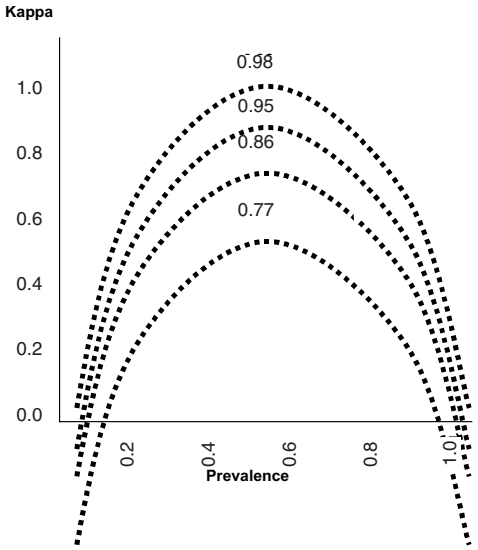


Figure 5. Relation between the different kappa/prevalence curves and the different Overall Agreements ranging from 0.77 to 0.98

The level of kappa values is dependent on the overall agreement P_o of the two observers. The lower the overall agreement in a reproducibility study, the lower the maximal and minimal kappa values become. In figure 5 this relation is shown. Note that in the prevalence/kappa curves with a low overall agreement P_o (0.86 and 0.77), the minimal kappa values become negative.

The dependence of the kappa value both on the prevalence P and on the overall agreement P_o illustrates the fact that a kappa value can only be interpreted in a proper fashion when both prevalence and overall agreements are mentioned in the final publication.

III.9.3 Optimising Procedures for Reproducibility Studies: Influencing the Overall Agreement and Prevalence in Advance to a Level of 50%

When performing a reproducibility study, the end result may be a low kappa value because of two predisposing factors: the overall agreement and the prevalence.
First, an overall agreement of less than 0.80 has the risk of resulting in a low kappa value.
Therefore, in the overall agreement period of the study (see figure 10, page 27), it is essential that observers try to achieve a substantial overall agreement P_o preferably above the level of 0.80. In this way the effect of the P_o on the final kappa value is under control.

Secondly, as shown above, very high and very low prevalences of the index condition result in low kappa values. Therefore we developed a theoretical method to influence the prevalence of the index condition in advance.

In figure 6 the prevalence/kappa curves are presented for the overall agreements P_o ranging from 0.83 till 0.98.

P_O	RANGE PREVALENCE
0.83	0.32 – 0.69
0.86	0.22 – 0.78
0.89	0.17 – 0.84
0.92	0.11 – 0.89
0.95	0.07 – 0.94
0.98	0.02 – 0.96

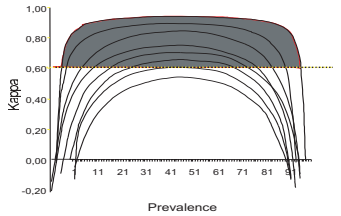


Figure 6. Kappa/prevalence curves of different overall agreements (0.83 – 0.98). The line through a kappa value of 0.60 demarcates the acceptable kappa area above this cut off line (gray area).

Note that the two lowest curves (P_o 0.83 and 0.86) are located beneath the line of the kappa value of 0.6. The curves with a P_o >0.90 have a substantial area (blue) above the 0.6 kappa cut off line.
To prevent unexpected low kappa values, because of unknown and too high or too low prevalences, we prefer to have a prevalence of the index condition near 0.50. The kappa values of prevalence of 0.50 are always located at the top of the curves.
Suppose that in the overall agreement period (see figure 10, page 27) we have achieved an overall agreement P_o of 0.85. We have 40 patients in whom we can study the reproducibility of a test.
Both Observer A as well as Observer B have each selected 20 patients, and each sends his/her 20 patients to the other observer. Each observer sends 10 patients who he judged to have a positive test and 10 subjects which he judged to have a negative test to the other observer. Based on an overall agreement of 0.85, both observers will agree in 85% of the positive and negative judged tests. And disagree in 15 %. In figure 7 the scheme is presented.

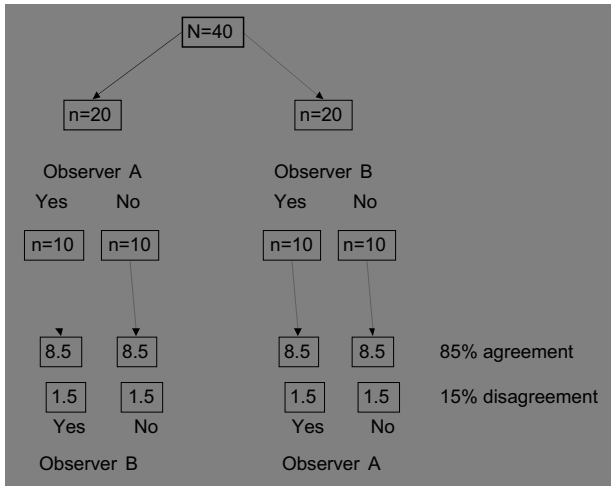


Figure 7. Scheme presenting the number of 40 patients with an overall agreement of 0.85, trying to get a prevalence of the index condition (positive test) of 0.50

Based on the number of patients in which the observers agree and disagree (figure 7), a kappa value can be calculated. In figure 8 a 2 x 2 contingency table shows the results. The prevalence is 0.50 with a overall agreement of 0.85, resulting in a kappa value of 0.70.

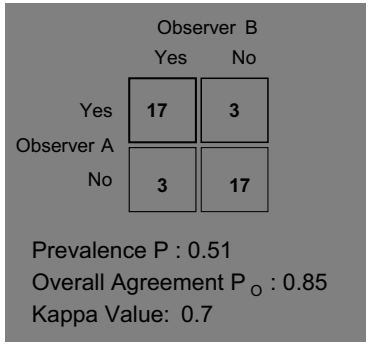


Figure 8. 2 x 2 contingency table based on the results of figure 7

By performing an overall agreement period in a reproducibility study with an overall agreement above the level of 0.80 and subsequently performing a procedure as illustrated in figure 7, one can influence the prevalence in advance resulting in a substantial kappa value of a test procedure. In a recent study, this proposed theoretical format was tested in practice and proved to be right (Patijn 2003, in press).

The easiest way of calculating the kappa value is to use a spreadsheet in which the formulae are integrated. In this way only the basic data has to be filled in and the kappa value is automatically calculated (see appendix 1). On the FIMM website a spreadsheet file can be downloaded.

III.10 Presentation Kappa Studies

In publishing the results of a reproducibility study, all aspects discussed under item 1 to 8 have to be presented. Furthermore, 2 x 2 contingency tables, the overall agreements and the prevalences are essential in a publication. In this way the reader of a paper can easily judge on what data the conclusion is based.
Figure 9 shows an example of a 2 x 2 contingency table. The calculation of the kappa value is also shown.

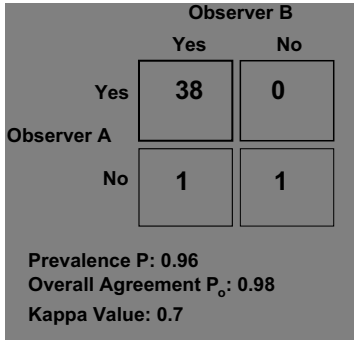


Figure 9. 2 x 2 contingency table of a reproducibility study of 40 subjects

III.11 References Kappa Literature

Barlo W, Lay M I, Azen P, A comparison of methods for calculating a stratified kappa, Statistics in Medicine, 1991; 10(9): 1465–1472

Cohen J, A coefficient of agreement for nominal scales. Educ Psychol Measurement, 1960;20:37–46

Cohen J, Weighted Kappa: Nominal Scale agreement with provision for scaled disagreement or partial credit, Psychol Bulletin, 1968: 70(4): 213–220

Cook R, Kappa, In: Encyclopedia of Biostatistics Eds. Armitage P, Colton T, Publishers Wiley, NY, 1998: 2160–2165

Deursen L L J M, Patijn J, Ockhuysen A L, Vortman B J, The value of different clinical tests of the sacroiliac joint, Ninth International Congress in Manual Medicine, London 1989, Abstr 16

Deursen L L J M, Patijn J, Ockhuysen A L, Vortman B J, The value of different clinical tests of the sacroiliac joint, J Manual Medicine 1990 (5): 96–99

Deursen van L L J M, Patijn J, Ockhuysen A L, Vortman B J, Die Wertigkeit einiger klinischer Funktionstests des Iliosakralgelenks, Manuelle Medizin 1992 vol 30(6): 43–46

Gjorup T, The Kappa Coefficient and the Prevalence of a Diagnosis, Meth Inform Med, 1988;27(4):184–186

Landis R J, Koch G G, The measurement of observer agreement for categorical data, Biometrics, 1977; 33: 159–174

Lantz C, Nebenzahl E, Behaviour and Interpretaion of K Statistics: Resolution of the two paradoxes, J Clin Epidemiology, 1996; 49(4): 431–434

Patijn J, Stevens A, Deursen L L J M, Van Roy J, Neurological Notions on the Sacroiliac Joint, In Progress in Vertebral Column Research, First International Symposium on the Sacroiliac Joint: Its Role in Posture and Locomotion Editors A Vleeming, C J Snijders, R Stoeckart, Maastricht, 1991: 128–138

Patijn J, Brouwer R, Lennep van L, Deursen L, The diagnostic value of sacroiliac test in patients with non-specific low back pain, J Orth Med, 2000; 22(1): 10–15

IV. SEVEN GOLDEN RULES FOR A REPRODUCIBILITY STUDY

In figure 10 a scheme is presented of the different aspects and stages of a reproducibility study on which the Golden rules are based.
Reproducibility studies are easy to perform and not restricted to large institutes like universities. Private practices or other institutes with two or more practitioners in M/M Medicine are very suitable for this kind of study

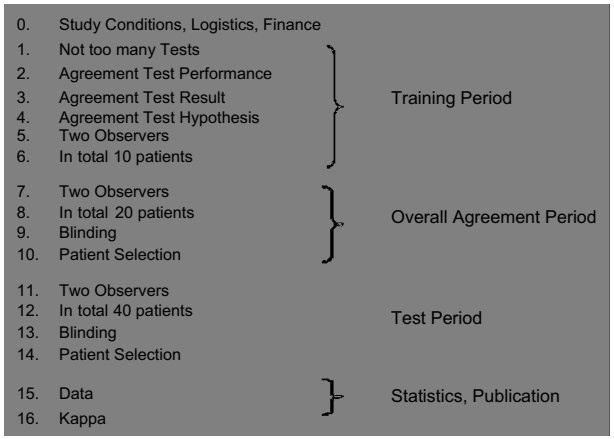


Figure 10. Plan of a reproducibility study

RULE 1 CREATE A CLEAR LOGISTIC AND RESPONSIBILITY STRUCTURE FOR THE REPRODUCIBILITY STUDY.
In a study one single person must be responsible for the entire process of the whole study.
This person is responsible for the logbook of the study. In this logbook all agreements and disagreements are written down and can be used as a reference cadre in group discussions. This person is responsible for the final updated format of the protocol. All participants have to sign this final protocol.

RULE 2 ALWAYS CREATE A TRAINING PERIOD BEFORE PERFORMING A REPRODUCIBILITY STUDY.
In the training period, it is essential for the future observers of a reproducibility study to discuss and define which tests and how many tests they are going to select for the reproducibility study. The decision on how many tests one wants to evaluate is dependent on the aim of the reproducibility study.

In the training period participants have to agree about the detailed performance of the test(s) that they are going to use for the reproducibility study.
20 patients can be used to discuss the precise sequence of procedure of the test(s). Finally, they have to agree about the precise performance of the test and make sure that each observer in a written protocol knows a standardised definition of the test procedure.

It is advisable not to restrict the agreed upon test procedure only to the patients of the study. But, by applying the same agreed upon test procedure to all one's patients visiting a clinic, it enhances the skills of the observers.
Participants have to agree how to define the outcome of the test(s) they are going to use for the reproducibility study. Participants have to perform the test(s) on the same 20 patients and to discuss the precise conclusions of the test(s). Finally, they have to agree about the precise judgement of the test and make sure that each observer in a written protocol knows a standardised definition of the test result. After every new decision, the logbook has to be updated.

Where a combination of tests is being studied, define the minimum number of positive tests for a final positive result of the test procedure.
Participants have to agree about the hypothesis of the test(s) they are going to use for the reproducibility study. Whatever test(s) selected for a reproducibility study, the observers have to investigate step by step the whole test procedure and agree about what the test really tests in their daily practice.

RULE 3 ALWAYS CREATE AN OVERALL AGREEMENT PERIOD BEFORE PERFORMING A REPRODUCIBILITY STUDY.
This period is essential to achieve a substantial overall agreement > 0.80. If the overall agreement is less than 0.80, participants have to discuss their agreements and have to pass the training period again.

RULE 4 ALWAYS USE A BLINDING PROCEDURE IN A REPRODUCIBILITY STUDY.
In the protocol it must be clear how the blinding is achieved not only with respect to the observers but also with respect to the patients. In most protocols, except with items such as pain, blinding is guaranteed when no information is exchanged either between observer and patient or between both observers. Use separate forms for each observer to record their findings.

RULE 5 ALWAYS DEFINE THE POPULATION FROM WHICH THE SUBJECTS ARE SELECTED.
This is essential to show how the selection was made (for example all patients on entrance) and no bias in selection of patients was performed.

RULE 6 ALWAYS MENTION THE DEFINITION OF THE SOURCE POPULATION, THE SELECTION METHOD, THE BLINDING PROCEDURE, THE DEFINITION OF TEST PROCEDURE AND TEST RESULTS IN MATERIALS AND METHODS WHEN PUBLISHING A REPRODUCIBILITY STUDY.

RULE 7 ALWAYS SHOW A 2 x 2 CONTINGENCY TABLE WITH THE PREVALENCE AND OVERALL AGREEMENT FIGURES IN RESULTS WHEN PUBLISHING A REPRODUCIBILITY STUDY.

V. VALIDITY

V.1 Gold or Criterion Standard

After achieving good reproducibility of a test procedure (the extent to which two observers agree about a test in the same population), the validity of a test has to be assessed.

Validity measures the extent to which the test actually does what it supposed to do. More precisely, the validity is determined by measuring how well a test performs against the gold or criterion standard. This is a major problem as well for diagnostics in gen-eral medicine as for diagnostics in M/M Medicine.
In M/M Medicine many characteristic diagnostic procedures, using for instance the end feeling in a passively performed test, are supposed to evaluate the mobility of the anatomical structure being examined. In the vast majority, only a hypothesis is available. For many tests in M/M Medicine, the gold or criterion standard has yet to be developed.

Two kinds of gold standards can be distinguished. First of all there is a gold standard for test procedures. For instances if a test procedure tests the range of motion, or resistance at the end of a passive motion, a gold standard has to be developed that measures in a quantitative way (degrees or N/cm^2) the range of motion or pressure in normal subjects. The evaluation of the quantitative method has also to include a test/retest procedure, to see whether the procedure shows the same data in the same normal subject on two different occasions.
In second instance, both the clinical test procedure and the quantitative method can be compared.

A second kind of gold standard for tests is related to the hypothesis of this test as taught by our teachers (SI-hypomobility) or with a diagnosis. This is the very problem as well for diagnostics in general medicine as for diagnostics in M/M Medicine.
The gold standard for a clinical test can be a radiological, a surgical finding, a post mortem, or a criterion based on data out of a normal population. So far, imaging techniques such as X-ray, CT and MRI are inconclusive in M/M Medicine, because a large number of normal subjects show abnormalities with these techniques.
In special cases, such as the Slump Test, which evaluates dural sac irritation for example from postoperative lumbar adhesions, MRI with gadolinium contrast can act as gold standard.
For some pain-provoking tests in M/M Medicine, the criterion standard is the effect of local anaesthesia in that particular area. The problem with this kind of criterion standard is that one is never sure about the systemic effect of local anaesthetics, and if we are dealing with a referred pain area, if were are sure that the pain is related to the anatomical structure we want to investigate, etc.

In M/M Medicine many tests are used to estimate the mobility of a joint by means of the end feeling. In this case two different policies can be followed. First, one can develop a quantitative method to evaluate the end feeling. In this case the end feeling procedure is validated clinically. Secondly, one can develop a quantitative method to estimate mobility of a joint. In this case, the mobility aspect of a clinical test is evaluated and therefore the real hypothesis of the test.
The list of above-mentioned examples is far from complete, but illustrates the way a gold standard can be developed.
In the absence of a well-defined criterion standard, sometimes a consensus view of experts using some other tests is used as a criterion standard. The problem with the consensus view is that the experts are only agreeing about a test procedure based on hypothesis and the real validity of a test remains uncertain.

In M/M Medicine, before spending much energy to defining gold standards, it is essential that first of all the test procedures are reproducible.

V.2 Sensitivity and Specificity

It has no sense in reproducibility studies to estimate the sensitivity and specificity, when no gold standard is available.

In sensitivity and specificity studies, 100 subjects are sufficient.
The same group of 100 patients is assessed with the test in question and with the gold standard (see 2 x 2 contingency table below). Cases **a** and **d** are correct, cases **c** and **b** are respectively false positive and false negative. A good test has to have few false-positive and false-negative results.
The prevalence of the index condition is illustrated by the formula: (a+c)/n.
It is essential to realise that the prevalence of an index condition can vary in different institutes, countries and from time to time.

The sensitivity of a test is defined as: the proportion of the cases that have the index condition **(a+c)** that the test correctly detects. In formula: **a/(a+c).**

The specificity of a test is defined as: the proportion of the cases that do not have the index condition **(b+d)** that the test correctly detects. In formula: **d/(b+d).**

Both sensitivity and specificity are needed to determine the validity of a test and always have to be presented together in a paper.

	Criterion Standard: positive	Criterion Standard: negative	
Result of Test: positive	**a**	**b**	**a+b**
negative	**c**	**d**	**c+d**
	a+c	**b+d**	**n = a+b+c+d**

V.3 Positive and Negative Predictive Value

To translate the statistics of sensitivity and specificity figures to daily practice, the physician has to know in the individual patient the chances whether a positive test is truly positive as opposed to false-positive. This is expressed in the so-called “positive predictive value of a test”. In the 2 x 2 contingency table above, the formula of positive predictive value of a test is: **a/(a+b)**. One has to realise that the positive predictive value of a test is dependent of the prevalence of the index condition **(a+c)/n.**

Suppose we have 1000 subjects with a sensitivity and specificity of respectively 0.8 and 0.7 and a prevalence of the index condition is 10% (see 2 x 2 contingency table above).

This means that when **n**=1000, then **a+c** = 0.10 x 1000 = 100.

In case of a given sensitivity **(a / (a+c))** of 0.8:

a / (a+c) = 0.8
}→ a/100 = 0.8 → **a** = 80
(a+c) = 100 }→ (80+**c**) = 100 → **c** = 20
(a+c) = 100

If **a+c** = 0.10 x 1000 = 100, n – **a+c** = **b+d** = 1000 – 100 = 900

In case of a given specificity **(d / (b+d))** of 0.7:

d / (b+d) = 0.7
}→ **d**/900 = 0.7 → **d** = 630
(b+d) = 900 }→ (630+**b**) = 900 → **b** = 270
(b+d) = 900

The positive predictive value of a test in this case is
a / (a+b) = 80 / (80 + 270) = 0.22

The negative predictive value of a test is likewise calculated:
c / (c+d) = 270 / (80 + 630) = 0.30

Where there is a larger prevalence of the index condition **(a+c)/n**, the positive predictive value of a test **a/(a+b)** also rises with the same sensitivity and specificity figures. Therefore, the positive predictive value of a test only reflects the prevalence of the index condition and not the property of the test itself.

V.4 Likelihood Ratio

For estimation of the predictive power of a test, independently of the prevalence of the index condition, the likelihood ratio has to be calculated. By definition the likelihood ratio in formula is:

$$\text{Likelihood ratio} = \frac{\text{Sensitivity}}{1 - \text{specificity}}$$

Tests with likelihood ratios close to 1 or <1 are completely useless for daily practice.

First, some remarks about this likelihood ratio and its use in calculating the diagnostic confidence odds.
Normally, we are accustomed to think of percentages like prevalence or true positive figures. The likelihood ratio does not operate on percentages, but on odds based on prevalence and diagnostic certainty.
Odds are the ratio of changes in favour of a condition versus the chances against that condition being present.
For example if a condition has a prevalence of 60%, the prevalence odds of the test being correct is 60 : 40 = 3 : 2. These odds can be changed again into decimal terms. If the prevalence odds are 3 : 2, the chances in favour are 3/(3+2) = 0.6.

By mathematical calculation, the diagnostic confidence odds are calculated by multiplying the likelihood ratio and the prevalence odds.

[Prevalence odds] x [Likelihood ratio] = [Diagnostic confidence odds]

To illustrate the importance of a large likelihood ratio in relation to the prevalence of a condition, an example is shown.

Suppose a condition, has a prevalence of 60% in your practice. Based on reproducibility and validity studies you know that the sensitivity is 0.8 and the specificity is 0.98.

Based on the formula: $\text{Likelihood ratio} = \frac{\text{Sensitivity}}{1 - \text{specificity}}$

the likelihood ratio is 40.
If a patient with a particular condition enters your practice, with a known prevalence figure of 40%, the chance of having this condition is 60%.

The prevalence odds in favour of having the condition are 6 : 4.
The odds for diagnostic confidence is 6/4 x 40 = 60.
Diagnostic confidence odds = 60 : 1.
Diagnostic Confidence is 60/60+1= 0.98 = 98%.

This means that you have improved your confidence from 60% to 98%. This is a good test.

When calculating for the same prevalence of 60%, but with a likelihood ratio of 0.6, the diagnostic confidence will be only 0.47 or 47%. This is less than the chance of 60% of having the condition for a patient when entering your practice. This is a bad test.

Published results of validity studies, trying to advise the daily practitioner which test he has to perform, and only mentioning sensitivity and specificity figures, are worthless. If one knows the prevalence of a certain condition, one can calculate, based on the likelihood figures, the diagnostic confidence.

APPENDIX 1

RELIABILITY of DIAGNOSTICS in M/M MEDICINE

		Observer B Yes	Observer B No	Total
Observer A	Yes	a	b	a + b
	No	c	d	c + d
	Total	a + c	b + d	n

Number of subjects = **n**

Overall agreement $p_o = \frac{a + d}{n}$

Expected chance agreement $p_c = \frac{a + b}{n} \times \frac{a + c}{n} + \frac{c + d}{n} \times \frac{b + d}{n}$

Kappa = $\frac{p_o - p_c}{1 - p_c}$

Prevalence P = (a + [b + c]/ 2) / n

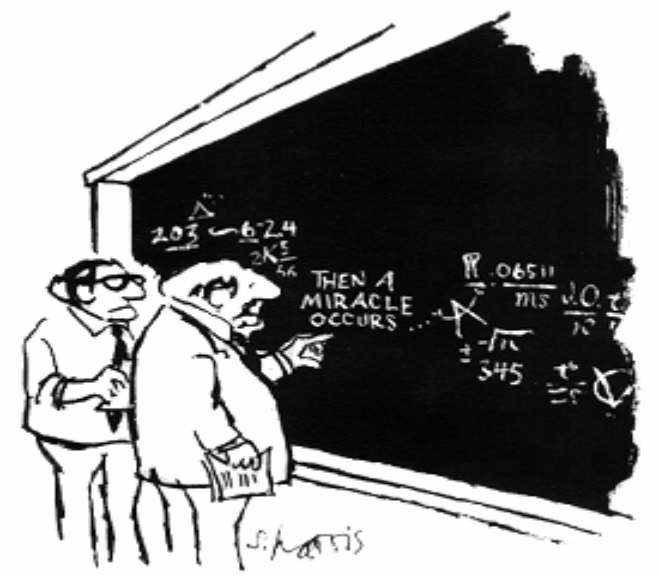
In a spreadsheet the following columns can be defined; see figure above:
Only data **a, b, c, d** has to be filled in:
Column A: data **a** (see 2 x 2 contingency table)
Column B: data **b** (see 2 x 2 contingency table)
Column C: data **c** (see 2 x 2 contingency table)
Column D: data **d** (see 2 x 2 contingency table)
Column E: data **n** Formula =A1+B1+C1+D1
Column F: data **a+b** Formula =A1+B1
Column G: data **a+c** Formula =A1+C1
Column H: data **c+d** Formula =C1+D1
Column I: data **b+d** Formula =B1+D1
Column E: data **a+d** Formula =A1+D1
Column K: Prevalence Formula =A1/E1+B1/2 x E1+C1/2 x E1
Column L: Overall Agreement P_o Formula =J1/E1
Column M: **(a+b)/n** Formula =F1/E1
Column N: **(a+c)/n** Formula =G1/E1
Column O: **(c+d)/n** Formula =H1/E1
Column P: **(b+d)/n** Formula =I1/E1
Column Q: data column M x N Formula =M1 x N1
Column R: data column O x P Formula =O1 x P1
Column S: Expected change agreement P_c Formula =Q1+R1
Column T: $P_o - P_c$ Formula =L1–S1
Column U: $1 - P_c$ Formula =1–S1
Column V: Kappa value Formula =T1/U1

THIS PAGE LEFT BLANK ON PURPOSE

SECTION 1 – Chapter VII
PLANNING TIME, COSTS AND OTHER RESOURCES

Research is complex business. You have to have a high tolerance for ambiguity. You also have to be able to be precise in the midst of uncertainty, because research must be precise. Therefore, you must also have confidence and be clear. Planning is one of the most important aspects of successful research projects. Working with a team is always more effective and has better outcomes than working alone.

One method that works for many people is to go to a room with a white board – or if you like chalk – a black or green board with at least one other person who you consider to be hypercritical. Map out your project on the board. Begin with your question. Don't be afraid. It may be turned inside out and totally different when you finally have the question you will use in your research. On the other hand it may go through a process of dissection and then become itself again – but you will be confident that you have taken all perspectives into consideration.



"I think you should be more explicit here in step two."

Map out your project in a time line. Begin on the far right with the end product date (poster, manuscript, graduation, what ever). Work backwards to now and determine what you will have to accomplish at major time intervals. Talk it out. Work it out. Have a plan.

Write down exactly what you need to do, who you need to do it with, and who needs to do it. Write down when you will meet with key individuals. Write down when you will have materials or information needed. Some research projects have gone unfinished because a student was "waiting for Samuel Becket's Godot" or waiting for that one data file or chart from someone who never received the request. Nothing is completed because nothing can be completed. The despair in the Becket play "Waiting for Godot" is never defined as such, but pervades the interaction between the characters. The two tramps cannot not wait for Godot, because he cannot come.

Perhaps there are no purchases or fees or salaries you will have to pay for in your project. Nonetheless, every research project has a cost. Universities are becoming more cost conscious. Estimate your costs. Include your time/salary, and that of your mentor. Include the cost of someone pulling records or the library making copies of documents. Include the cost of supplies if there are any. Consider the cost of equipment operations or lab supplies. If you are not asking for money all you need is an estimate. It is always a good thing to be able to tell someone that your research project involving abstracting information from existing data or from medical records cost $5,000 for your time and the software use.

Sample budgets are provided at the end of Chapter VII.

SECTION 2 – Chapter VII
Presenting Your Findings

2.A. Posters, Abstracts, and Oral Venues

Most likely you will be asked to develop a poster reporting the results of your study. You might also orally present your poster, or orally present your findings at a workshop or panel or grand rounds presentation.

When you have results it will be helpful for you to practice discussing these at informal settings or with colleagues. This will help you to get acquainted with your own materials and to gather insights, questions, and comments from others as they listen to and consider the work you have achieved.

The major headings a poster are similar to a proposal or manuscript. Normally you begin with an abstract for a scientific poster, or an introduction for an academic or other type of poster, such as a process poster.

The abstract should include why you selected this particular topic in a couple of sentences, then a brief synopsis of your research design, the major findings or results supported by the relevant statistics, a short two or three sentence conclusions/discussion, and any limitations of the study.

Visual aids are an important part of any presentation. Charts and graphs can present a great deal of information quickly and effectively. One or two well-placed figures or tables or images can make your point, whereas too many will clutter your presentation. Each graph should be easily understood. Simplicity highlights key findings. Complicated charts or graphs with many variables will leave your audience wondering about the graph and forgetting your point.

For oral presentations with slides find a good guidebook on power-point slide creation. Keep it simple. Place your notes so you continue to face the audience. Slides should be clean – no paragraphs, only lists and paraphrases.

After presenting your findings orally in several venues, it may be time to present your work for formal publication. Consider several possible journals and acquire the instructions for authors and follow them even if someone advises you differently, follow the rules for the targeted journal and show them these rules. Look at the literature you used to support the research idea and consider those journal venues. These are the journals whose readers would likely be interested in your work.

At the end of Chapter VII we provide several different checklists to help you think about different types of printed and oral reports.

SECTION 3
WRITING A CASE REPORT

What is a Case Report?

A case report describes a clinical situation that the student has identified as unusual or complicated. This may be a clinical situation that has an anomaly, an atypical etiology, or a rare condition. This may be a clinical situation that is frequently found in the literature or in the patient population but that required a novel approach to treatment because of a unique or special requirement of the patient.

A case report is often the first writing a medical professional undertakes. It has a unique function in communications; teaching. Many successful medical writers have accumulated case reports they have written to organize their thinking about their work. Some of these case reports are published later after multiple similar cases are documented, or following a research report that offers insight into the conditions identified earlier. Case reports can contribute to and inspire improved diagnostic procedures, treatment methods or research.

The case report is also attractive because it is about the essence of medicine; the patient. A well crafted case report is uniquely instructive for the physician and for other health care professionals.

Organizing the Case Report

Before you begin writing, do your research on the background. Determine how many previous cases have been reported. Compare your case to published reports. Refer to Chapter II, The Literature Review.

Develop an outline. If you have a target journal, use its instructions for authors.

Possible Outline Structure

- **I.** Introduction
- **II.** Abstract or case summary
- **III.** Report of literature (keeping in mind the available space for your report, summarize, summarize, summarize and then focus on only the MOST salient literature for this report)
- **IV.** Discussion
- **V.** Conclusions
- **VI.** References

Style

Use a narrative style. You are reporting and discussing a medical condition in a human being. Report the findings in a chronological order, in the same way that findings may have unfolded or revealed themselves in the clinical environment.

The title of your report may be catchy, but should above all be scientifically relevant and reflect the true clinical condition about which you are reporting. For example if this case is about a patient with fibromyalgia, the title should include that word, rather than something that may mislead the reader, such as “somatization of chronic depression.” Published papers are indexed by title content and key words.

In the discussion section you may speculate about various findings in the form of “questions for future research” or “because of gaps in the literature it is unclear whether/what, and so forth.”

Details

Facts to Consider

You can use the following list as a checklist for those facts you need to include or consider in your case report.

- ☐ Gender, Age, Race AND Ethnic or culture Conceptions of race, as well as specific racial groupings, vary by culture and over time, and are often controversial, for scientific reasons as well as because of their impact on social identity and identity politics. Most biological and social scientists regard the concept of race primarily as a social construct, while some maintain it has a genetic basis. Check the website of the U.S. Department of Health and Human Services for the latest information.
- ☐ Patient's chief presenting complaint
- ☐ Other health conditions noted
- ☐ History of presenting health condition
- ☐ Date of onset
- ☐ Symptoms manifested – typical and atypical
- ☐ Social history
  - Including but not only occupation, marital status, children, family, living conditions, education
- ☐ Health history (past history is redundant – history IS past)
  - Including but not only surgeries, births, infectious diseases, injuries
- ☐ Genetic history
- ☐ Physical exam findings
- ☐ Medical and Lab tests and findings/values
- ☐ Diagnosis
- ☐ Consultations – reports
- ☐ Treatment prescribed
  - Material and methods, frequency, duration, dose response
- ☐ Final outcome
- ☐ Autopsy report

Medical and Lab Tests

It is not always necessary to report the lab values themselves unless they are of statistical significance or important visually to dramatically portray the findings and illustrate how different or how subtle or obvious the findings were or should have been. Think of this as if you are writing a mystery that unfolds and is being deconstructed. At the end you reveal the culprit or the answer to the mystery. If you use clinical values be sure to compare them to normal or expected values.

It is usually important, but not always, to list the tests performed to develop a differential diagnoses. In other situations you may report that standard tests were performed with negative findings. Always report tests for which findings were abnormal, and their degree of difference from normal.

Social and Health Histories

Use only the family and social history elements that are pertinent to the case and make some statement as to the average or normal status of other clinical history elements. For example, if there is no history of the condition and normally that would be the primary clinical indicator of the condition existing in your patient, this should be stated. If the family history has no relevance to this type of condition you should so state that fact. However, if a condition is only passed maternally, you do not need to describe the father's medical history as it would be irrelevant. Consider what you write to be of value to the extent that it is pertinent to the main purpose of the case report.

If you plan to use illustrations, plan for them early in the process. Always refer to the illustration and discuss it clearly and briefly. Say WHY you included it of all those you might have included. For example if you have used two different methods of measuring the diameter of the carpal tunnel, why would you select only the MRI image and not the ultrasound image? If you have an x-ray that is abnormal, how is it so, and do you need a normal picture for comparison – is the abnormality obvious or subtle?

Tips

Avoid jargon

Always spell out the term and then use a commonly accepted abbreviation. Use a reference such as Dorland's Illustrated Medical Dictionary for universally accepted abbreviations. For example, the expression "no pathology was found," is virtually meaningless according to the dictionary definition of "pathology" from Webster's New International Dictionary (unabridged, second edition).

Use clear language and grammatical structure

Examples of bad grammar

"Babies with a bad history in evidence of rapid progressive hemolytic process were exchanged soon after birth."

"The ankles began to swell toward the end of the day and disappeared by the next morning."

"With adequate relaxation in the patient, on his abdomen with his hips flexed over the edge of the table, we have the deformity facing us,"

Details to include or not to include

Irrelevant detail is exactly that. Detail may be interesting AND relevant, but irrelevant detail is rarely interesting. It is difficult to cut out of your material. Be parsimonious – be ruthless get others to review your material and ask questions or make suggestions on the level of detail you need or can cut.

The italicized material on the following page was taken from instructions to students in a college of osteopathic medicine on how to write a case report.

"A middle-aged housewife strained her back while moving furniture during spring cleaning. She was treated as an outpatient for three weeks, with manipulation and muscle relaxants, and was advised to rest at home between clinic visits. The pain persisted, however, so she was admitted to the hospital on March 5, 1965."

Details of age and race can be given in the section describing the first physical examination, or the case report can simply begin with, "A white housewife, aged 47, strained…" But, notice that the doctor who reads even this short section learns many things that are not actually said: The patient is, probably, an active woman, although not accustomed to constant heavy work. She, likely, is not wealthy, or she would not be doing her own heavy cleaning, but neither is she a pauper, or she would not likely have her own home to clean. She, doubtless, has some pride in the appearance of her home, though the extent cannot be known. The doctor can, probably, guess something about the kind of injury and immediately place the situation beside similar instances he has encountered.

These few words thus recreate for the physician, who reads them, a recognizable picture of a human being, not just another hospital chart of a case of backache. This is always a proper objective.

CAUTION Carefully assess the material. The above section above has several examples of thinking errors. We cannot assume that someone else will infer anything from vague statements. The above material has biased statements, socially fraught with risks in miscommunications. Why?

"A middle-aged" (be cautious in using this terminology as middle-aged to a 28 year old person does not mean the same thing as to a 45 year old person – be specific about chronological age)

"housewife" (again terminology – this is a female – by gender and not a "housewife" by any category of profession or lifestyle)

"strained her back" (you don't know if she strained it or hurt something else – she is complaining of back pain that could be associated with a "strain."

"while moving furniture during spring cleaning" (spring cleaning is an irrelevant detail. Was it WHILE or AFTER or BEFORE moving furniture – what furniture? A lamp? A sofa?)

"She was treated as an outpatient for three weeks" (once a week? Twice a week?)

"with manipulation" (what kind of manipulative medicine? HVLA? Craniosacral?)

"and muscle relaxants" (what kind prescribed?

"Although she was advised to rest at home between clinic visits" (did she??),

"her reported" (not "the") pain persisted.

"Therefore she was admitted to the hospital (better to say inpatient care) on March 5, 1965."

This patient may in fact be a woman who is poor and abused by someone who forces her to move heavy furniture and use archaic methods of cleaning.

OR

She may have fallen on the ice four days earlier and hurt her hip which translated into back pain.

OR

She may have had to care for a grandchild and strained herself picking up the baby from a play area.

What difference does it make if she is wealthy or poor, educated or not, happy or sad. Be VERY careful not to stereotype anyone or make risky assumptions about what the reader will infer. Do not imply anything in a case report. Say what you mean using objective facts not subjective statements – if you want to report observations say they are your observations.

CAUTION Keep the human touch while providing the facts. Compare these two examples below.

"Case 54321, middle-aged housewife was admitted for c.c. of low back pain of 3 weeks' duration"

OR

A white non-Hispanic female, age 42, presented a complaint of low back pain. She reported that it occurred immediately after she had moved heavy furniture alone, in her home during an active session of cleaning house."

Final Checklist

- ☐ Case materials are complete – no pending data
  - No?
    - → Fill in gaps in data.
- ☐ Case materials are accurate – no questions about the accuracy or completness
  - No?
    - → Clarify.
- ☐ Case materials are adequate – sufficient in number and quality
  - No?
    - → Get additional data.
- ☐ My literature search and references have been vetted by an expert in the field
  - No?
    - → Get expert input.
- ☐ My Illustrations are clear, referenced, and explained
  - No?
    - → Make it so.
- ☐ I have all necessary permissions to quote, use instruments, figures, or illustrations
  - No?
    - → Make it so.
- ☐ I have all necessary authorizations and approvals to submit this for publication
  Any student or resident physician reports submitted to the JAOA require permission of the attending physician or the head of the department to be considered for publication.
  - No?
    - → Get them.
- ☐ My reference style is correct for the target journal?
  - No?
    - → Make it so.
- ☐ The paper been reviewed by an expert or reliable and trusted reviewer/editor
  - No?
    - → Seek it and revise as needed. (Ensure it is read by someone familiar with <u>and</u> someone <u>not</u> familiar with the topic/case)

SUMMARY

The biggest criticism of osteopathic manipulative medicine – including neuro-musculoskeletal and physical medicine treatments, is the inadequacy of its evidence-base. Without scientific evidence of how OMM works on the human body, OMM will forever be a cash-only, optional treatment with limited expert practitioners. Other health related professions and technicians are becoming more recognized as offering healing forms of manual therapies. If we want this to continue to be a part of the complete physician's repertoire of available therapies, we must continue to research it rigorously.

Medical students and residents can perform meaningful research that contributes to the scientific knowledge of how OMM works and under what conditions. Guided by the hands of a few dedicated, physician scientists, you can advance the science of osteopathic medicine.

Research develops critical thinking skills that improves clinical decision making. Medicine is a science that is constantly changing with new information on drug therapies and other treatment modalities amassing at a staggering rate. It is essential to effective 21st Century medicine to remain current with research and evaluate it with a critical mind.

The best understanding of research comes from those who have been involved in its synthesis. Thus an essential part of the training process of any physician should be a research project. This manual provides you with tools to organize your ideas and conduct scientifically sound research in osteopathic medicine.

The following pages provide

- Recommended minimum research competencies for medical students and residents
- Examples of survey instruments
- Production checklists for you to use to produce your proposal and reports of your research
- Two sample budgets

Recommended Minimum Research Competencies
in Osteopathic Medical Education

There is a need to set a national standard for research competency in osteopathic medical education. The depth and breadth of competency will increase over the course of medical education. As follows is the list of functional competencies expected of medical students and residents at the designated points in their education.

End of medical school years 1-2:

1) History of osteopathic research
2) Knowledge of research vocabulary
3) Ability to do a literature search
4) Knowledge of basic statistics
5) Understanding of research problems that are uniquely osteopathic (OMT)
6) Awareness of support resources available consistent with level of competency expected

End of medical school years 3-4:

1) Ability to review & summarize journal articles
2) Ability to formulate a research question/hypothesis
3) Awareness of support resources available consistent with level of competency expected

End of post-graduate years 1-3:

1) Understand the process of design and implementation of a research project
2) Ability to critique journal articles
3) Ability to write a manuscript suitable for publication or a grant application
4) Awareness of support resources available consistent with level of competency expected

Endorsed by:

1. AOA Bureau of Research 11-23-02
2. Educational Council on Osteopathic Principles 11-30-01.
3. Osteopathic Collaborative Clinical Trials Initiative Conference III (OCCTIC III) Fall 2001.

Attitudes Toward Osteopathic Principles and Practice Survey – 2001 (ATOPPS)

Please use the scale below to indicate how much you agree with the following statements.

SD Strongly Disagree D Disagree U Undecided A Agree SA Strongly Agree

1.	I believe that the application of osteopathic principles is useful in medical practice	SD	D	U	A	SA
2.	Manipulation and OMT aren't real medicine.	SD	D	U	A	SA
3.	Overall, I think that I would have been happier if I pursued my medical education at a MD school.	SD	D	U	A	SA
4.	I believe that learning osteopathic practices, such as OMT, will give me an advantage over MD's when it comes to patient care.	SD	D	U	A	SA
5.	I intend to keep current about advances in OMT.	SD	D	U	A	SA
6.	If things had gone my way, I would not have chosen to attend an osteopathic medical school.	SD	D	U	A	SA
7.	I anticipate using my manipulative skills.	SD	D	U	A	SA
8.	Attending OMM lectures and labs took away valuable time that could have been better spent studying for other classes.	SD	D	U	A	SA
9.	I have a lot of respect for physicians who use OMT.	SD	D	U	A	SA
10.	I think that manipulative medicine is a useful tool for a primary care physician	SD	D	U	A	SA
11.	Approaching the human body as a unified biological machine is helpful when it comes to understanding patient care issues.	SD	D	U	A	SA
12.	I plan to apply osteopathic principles in my practice.	SD	D	U	A	SA
13.	I'm convinced that exposure to osteopathic principles helps medical students learn better patient care skills.	SD	D	U	A	SA
14.	In all likelihood, after I graduate I won't ever use OMT in my daily practice.	SD	D	U	A	SA
15.	I think that there exists a substantial scientific basis for osteopathic principles.	SD	D	U	A	SA
16.	Looking for neuromuscular symptoms or somatic manifestations of disease can assist in forming a differential diagnosis.	SD	D	U	A	SA
17.	I'm convinced that osteopathic principles distinguish D.O.'s from their M.D. counterparts.	SD	D	U	A	SA
18.	Patients are more than the sum of their physical parts.	SD	D	U	A	SA
19.	I'm convinced that osteopathic practice distinguishes D.O.'s from their M.D. counterparts.	SD	D	U	A	SA
20.	I'm sure I could learn new osteopathic manipulative techniques easily.	SD	D	U	A	SA
21.	I knew that I probably wouldn't like learning OMT even before I started medical school.	SD	D	U	A	SA
22.	It is helpful to have a "big picture" perspective of a patient's history when planning treatment.	SD	D	U	A	SA
23.	In general, men have an advantage over women when it comes to performing OMT.	SD	D	U	A	SA
24.	A certain body type is needed in order to perform OMT competently.	SD	D	U	A	SA
25.	Students who enjoy and excel at OMM are strange.	SD	D	U	A	SA
26.	At first, I didn't think osteopathic principles had much merit, but now I do.	SD	D	U	A	SA
27.	If someone asked me about osteopathic principles, I could clearly explain them.	SD	D	U	A	SA
28.	If OMT works, it's mostly a placebo effect.	SD	D	U	A	SA
29.	Except in simple cases, OMT should be referred to an OMM specialist.	SD	D	U	A	SA
30.	In most clinical cases, I believe I could treat the structural component associated with the patient's illness with OMT.	SD	D	U	A	SA
31.	In general, OMT has a place in the treatment of Cardiovascular diseases (HTN, MI,...).	SD	D	U	A	SA
32.	In general, OMT has a place in the treatment of Respiratory diseases (COPD, Asthma,...).	SD	D	U	A	SA
33.	In general, OMT has a place in the treatment of Gastrointestinal diseases. (Cholecystitis, Pancreatitis,…)	SD	D	U	A	SA

PRODUCTION CHECKLISTS

CHECKLIST FOR QUALITY ISSUES IN THE PROPOSAL

Problem identification

- ☐ Is the research problem/line of enquiry clearly identified?
- ☐ Are the aims and/or objectives of the research clearly specified?

Background

- ☐ Has appropriate literature been examined to contribute to the understanding of the research problem and/or conceptual framework for the study?
- ☐ Have other relevant sources been reviewed?

Approach

- ☐ Are the conceptual framework and theoretical assumptions clearly stated?
- ☐ Are the proposed research methods clearly outlined?
- ☐ Are the project design, methods of data collection and analysis appropriate to the aims of the research?

Feasibility

- ☐ Does the research project appear feasible, given what is know about the proposed research methods, time frame, researcher's track record etc.?
- ☐ Are the intended research outputs appropriate or realistic?

Significance and impact

- ☐ Is the proposed research a new line of enquiry within or across disciplines?
- ☐ Does the research project have a potential social impact, e.g. promote problem solving, policy development or evaluation?

Team projects

☐ If the proposed research is a team project, are the roles of team members clearly spelled out?

Capacity building

- ☐ Does the project contain an element of research capacity building?
- ☐ If the project includes capacity building, is it likely to develop real research-related skills in the students or novice researchers concerned?

Budget

- ☐ Is a detailed budget provided?
- ☐ Is the budget justified in relation to the project's proposed aims and research activities?

Dissemination of research results

- ☐ Is there a plan to disseminate the research findings within the research community?
- ☐ Is there a plan to disseminate the research findings among stakeholders and the wider public?
- ☐ If applicable, is there a plan to report back to the community studied?

This list excerpted from a longer list of criteria used by the National Research Foundation to evaluate proposals.

Scientific Poster Criteria Checklist

General appearance

- ☐ Did I use a template?
- ☐ Does my school have a template I need to follow?
- ☐ Does the conference have a template or other requirements?

Title

- ☐ Is it appropriately scientific?
- ☐ Is it attractive – eye catching?
- ☐ Are all authors listed in proper order with titles and affiliations?
- ☐ Are logos properly used and placed?

Content

Abstract

- ☐ Include reference to IRB or IACUC
- ☐ Include purpose/question
- ☐ Include statement of rationale
- ☐ Include statement of main results or findings
- ☐ Include statement of major findings (primary outcome)

Background and Significance or Introduction

- ☐ Most salient literature
- ☐ Primary value to the science/topic

Aims or Purpose and Questions or Hypotheses

- ☐ Listed, emphasized, stands-out

Research Design and Methods

- ☐ Type of study
- ☐ Primary outcomes
- ☐ Subjects/sources of information
- ☐ Interventions/comparisons
- ☐ Data collection and management

Findings

- ☐ Charts and tables
- ☐ Major findings stand-out – easily read

Conclusions

- ☐ Strengths of study
- ☐ Limitations of study
- ☐ Impact statement
- ☐ Recommendations for future research

References

- ☐ Keep to minimum – more than 10 is too many

Acknowledgements (funding, sponsor, assistance from other people)

How I will dress to present my poster

Will I include materials to give the viewer e.g. a chart, a table, an abstract

Questions to guide your thinking

As you develop a scientific manuscript consider the following material excerpted from guidelines for workshop presentations at the AOA research conference 2003

A Component: Osteopathic clinical description of the condition

1. What are the musculoskeletal or clinical details that characterize this condition?
2. What manipulative medicine treatments/techniques are currently used in treating this condition?
3. What is the rationale for application of OMT in this condition?

B Component: Basic Bio-physiologic foundations for the topic clinical condition

1. How does one describe the scientific basis for the clinical condition?
2. What elements of human physiology are involved in causing or manifesting the condition?
3. What do we know about current traditional treatment of the condition?

C Component: Current knowledge regarding the relationship between the physiological and clinical manifestations of the condition

1. What knowledge is available in published research or other medical and scientific texts regarding the use of OMT for this condition compared to other interventions?
2. What are the outcomes studied and the methods used in available research?

D Component: Future Directions

1. What are the gaps in our knowledge?
2. What questions are raised by the current available scientific and clinical knowledge?
3. What research activities are recommended for the future?

NAME OF PROJECT - YEAR ONE - START DATE - END DATE

PERSONNEL					DOLLAR AMOUNT REQUESTED		
NAME	ROLE ON PROJECT	TYPE APPT. (months)	Total Hours on Project	BASE SALARY	SALARY REQUESTED	FRINGE BENEFITS	TOTAL
NAME	Principal Investigator	12	240	90,000	10,385	2,077	12,462
NAME	Co-Investigator	12	290	58,800	8,156	1,876	10,032
NAME	Biostats Consultant	12	60	124,800	3,900	-	3,900
NAME	Co-Investigator	12	85	186,600	7,626	-	7,626
NAME	Psychiatric Consultant	12	24	186,600	-	-	-
Personnel SubTotal					30,067	3,953	34,020
EQUIPMENT (*Itemize*) Operations							
SUPPLIES (*Itemize by category*)							
OTHER EXPENSES (*Itemize by category*) SUBJECT COMPENSATION $20 each interview-juvenile; $20 each interview-family (n 132)							5,040
SUBTOTAL DIRECT COSTS FOR EACH ANNUAL BUDGET PERIOD(shown here as M&O only)							5,040
FACILITIES AND ADMINISTRATIVE COSTS (INDIRECT COSTS)							
TOTAL DIRECT COSTS FOR EACH ANNUAL BUDGET PERIOD							39,060

Budget Justification

A total of $39,060 is requested for costs associated with this proposed one-year prospective study.

Key Personnel

NAME Associate Professor of Psychiatry at the INSTITUTION, will devote an average of 20 hours per month over the year of the research project in the role of principal investigator. She will provide the overall leadership to the study. She will meet at least bi-weekly with the research team to monitor progress and resolve problems. She will train the interviewers, conduct interviews and collaborate with other members of the research team. NAME will prepare interview packets, collect them, store them, and serve as the liaison with the Institutional Review Boards. She will be supported by the Department of Psychiatry for any additional time needed to ensure the successful completion of the study and dissemination of the results. She will be responsible to ensure data is entered and work with NAME in the cleaning and analysis of data.

NAME, TITLE, as a co-investigator, will dedicate an average of about 5 ½ hours per week to this project. NAME will manage the random selection process, assign each potential subject to a designated interviewer, participate in the selection and training of interviewers, meet with the research team, and collaborate with other key personnel. NAME has participated significantly in the development of this proposal.

NAME TITLE will provide a total of 60 hours of consultation supported by funding from this grant, in the analysis of data and development of the final report. NAME has been involved in the development of the proposal.

NAME, Associate Professor will devote an average of one day per month to this study as a co-investigator for the psychiatric component, including analysis of the MAYSI data, insights into the juvenile social competency measures, and participate in monthly research team meetings.

NAME Assistant Professor and Chair, Department of Psychiatry will contribute his time to this project. He will confer monthly with the research team at the Department of Psychiatry research meeting on its progress. He will participate in the analysis of data and dissemination of findings.

Supplies will be donated by the Department of Psychiatry. This includes the cost of proprietary instruments, paper, envelopes, copying, and other routine research supplies.

Subject compensation of $5,040 for 126 juveniles and their families, is for the time and other costs associated with the families completing the interviews. Target gift cards will be given ($20.00) each per family, and per juvenile.

Note: It is better to use an excel file to calculate budgets. If you have to put your budget in a word document you can copy the excel file – then you know your calculations are correctly verified, thus avoiding costly errors. The ORC has the NIH budget format in excel and makes this available for others to use.

SAMPLE RESEARCH BUDGET

A non-medical research project to be conducted on-campus
100% sponsor agency funding.
Project period 10/01/2003 to 9/30/2006
Budget period 10/01/2003 to 9/30/2004 (Year 1 budget)

	Agency Request
Senior Personnel	
PI Dr. B. Smith @ 2 SM months	14,000
Co-PI Dr. L. Jones @ 1.25 SM months	10,000
Co-PI Dr. C. Thomas @ 2 CY months	12,000
Other Personnel	
Analysis Technician@ 6 CY months	16,000
2 Graduate Students @ 12 CY months	12,000
Total Personnel	64,000
Fringe Benefits (PI, Co-PIs & Technician)	13,000
Maintenance, Operations, Travel	
Equipment	15,000
Travel	4,500
Materials and Supplies	9,500
Consultant	2,500
Subcontract (Azalea University)	5,000
Subcontract (Dogwood College)	35,000
Printing	1,500
Total Direct Costs	150,000
F&A Costs @ 38% MTDC (125,000)	47,500
Total Direct and F&A Costs	197,500

Total Project Cost $ 197,500

SAMPLE BUDGET JUSTIFICATION

Senior Personnel

There are one primary investigator and two co-investigators:
Dr. Barbara Jones, Assistant Professor, will be the principle investigator leading the research. The PI is on a 9-month academic year appointment at $ 63,000 and requests two months of summer salary ($ 14,000).
Dr. Larry Smith, Associate Professor, is a Co-PI and will assist in the project. He is also on a 9-month academic year appointment at $ 72,000 and requests 1.25 months of summer salary ($ 10,000).
Dr. Carol Thomas, Associate Professor, is the second Co-PI. She is on a 12-month calendar year appointment at $ 72,000 and requests 2 months salary ($ 12,000).

Other Personnel

Ms. Julie Chen, a full-time technician currently on staff, will devote 50% of her time to this project ($ 16,000) to prepare samples for analysis and record results.

Support is requested for two graduate students at $ 6,000 each. These students will collect samples under supervision of the PI and Co-PI's.

Fringe Benefits

Fringe benefits are calculated at 25% of salary for the PI, Co-PI's and Technician.
Graduate Students do not receive benefits in accord with standing University policy.

Equipment

Support is requested to purchase an Analysis Machine to be used exclusively on this project. The catalog price quoted by the manufacturer for this item is $ 15,000.

Travel

The PI and Co-PI will travel to two technical conferences at an estimated cost of $ 1,000 per conference. The project will need $ 2,000 for local travel by graduate students and faculty for sample collection (auto mileage and meal allowances). $ 1,500 is budgeted for the PI to travel to an international conference to present research results.

Materials and Supplies

Total request of $ 9,500 is based on estimates of $ 4,500 for consumable lab supplies, $ 3,500 for chemicals for sample analysis and $ 1,500 for specialized sample containers.

Consultant Services

Dr. Howard Phillips, chief scientist for LMC Systems Inc. will act as a consultant in the interpretation of certain sample analysis results. His rate is $ 300 per day for 5 days plus an estimated $ 1,000 in travel costs.

Subcontracts

Dr. Marilyn Johnson of Azalea University will provide specialized statistical analysis work in connection with final compilation of project data ($ 5,000).
Sample collection and analysis for the control group will be performed at Dogwood College under the supervision of Dr. Albert Hunter ($ 35,000).

Printing

Estimated at $ 800 for journal page charges and $ 700 for the production of 30 copies of a final report as required by the sponsor

SOME TEXTS YOU CAN RELY ON

Friedman: Research Tool Kit: Putting It All Together, 1998: Brooks/Cole Publishing

Hulley, Cummings, Browner, Grady, Hearst & Newman: Designing Clinical Research (2nd Edition), 2001: Lippincott, Williams, & Wilkins Publishers (light blue paper back)

Martella, Nelson & Marchand-Martella: Research Methods: Learning to become a critical research consumer, 1999: Allyn & Bacon Publishers

Neutens & Rubinson: Research Techniques for the Health Sciences, 2001: Benjamin Cummings (San Francisco)

Punch: Developing Effective Research Proposals, 2000: Sage Publisher

Statistics for the Utterly Confused by Lloyd Jaisingh, Ph.D. McGraw Hill

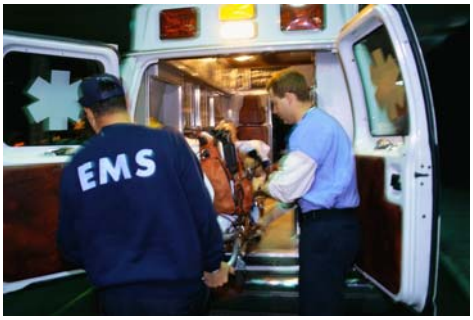
Contact any of the individuals listed in the Forward to this Manual to identify other quality texts that will help you with your research.

This page intentionally left blank

How do you feel?

Successful

or



need resuscitation?

This page intentionally left blank