

# **Residency Research Manual**



**KU School of Medicine**

**Department of Family Medicine**

## **KU School of Medicine Department of Family Medicine**

### **Greetings!**

Welcome to the KU Family Medicine Residency Research Manual. **The purpose of this document is to provide an abbreviated resource to our residents that will help them achieve success** in meeting their residency research requirement. More than that, we hope that this remains a “living document”, with updates and additions that are included based on input from residents as they complete their program at KU.

The manual includes the bare essentials, so we recommend that residents use this manual as a starting point and reference resource as they complete their research projects. Suggestions for improvement, corrections and any comments are welcome! Please direct them to the Research Director, Dr. Allen Greiner.

All new residents will be provided with a binder that contains a printed copy of the Manual and also includes information about KUMC’s Human Subjects Committee requirements, necessary forms, a section on how to construct effective surveys (including how to access the research division’s SurveyMonkey account), and other handy tip sheets. The Basic Manual also will be available on line and can be found on the KUMC Family Medicine Residency web page, the KUMC Research Division web page and on the KUMC Family Medicine K drive secure server as a read-only document.

We hope you find the manual useful, that you make suggestions of how it can be improved to meet your needs, and that you come back to it from time to time when you are grappling with an issue and want to conduct effective research to answer your clinical or community health research question.

### **Acknowledgements**

Providing an abbreviated, practical, “one stop” reference document to assist residents in conducting research was the vision of **Linda Frazier, MD, MPH**, Department of Obstetrics and Gynecology at KUSM-Wichita. She wrote and organized a research manual for the OBGYN residency program in Wichita and graciously permitted Family Medicine to borrow liberally to produce this version. It is with deep appreciation that we acknowledge her assistance.

## TABLE OF CONTENTS

TOPIC	PAGE
Greeting and acknowledgements	1
KUMC Family Medicine Research Mission Statement	4
Why should you care about research?	4
Resident requirements	4
Residency training competencies	5
Learning objectives	5
Timeline	6
Steps in research project development, implementation and completion	8
Research record retention policy	9
Research support	9
Choosing a topic	10
Study design	10
Study feasibility	11
Refining the research question	11
Developing the research plan	12
Approval of research project	13
Planning for data collection	13
Pre-test and revising the research plan	13
Analyzing the findings	14
Drawing conclusions	16
Disseminating results	16
EndNote®	16
PowerPoint®	17
Faculty and staff resources	18
Recommended readings	19

## APPENDICES

Appendix	PAGE
I. How to calculate 95% CI on a proportion	20
II. How to read an SPSS printout	21
III. Null hypothesis	26

## KUMC Family Medicine Research Mission Statement

To highlight, engage in and promote the value of high-quality primary care research within our academic health center, within our communities, across Kansas and across the nation.

### Why Should You Care About Research?

Research is essential in *improving* the delivery of primary care to patients. Only with research can we identify best practices, assess our effectiveness in improving health, and better understand patient needs. Research also provides practitioners with tools to critique and question published findings and reports that can impact practice. Honing research skills is an essential element to improving the practice of family medicine globally and personally.

In addition to making us better physicians, the residency program in the department of family medicine requires a hypothesis-based research project for program graduation. In the past, these projects often were initiated by residents with advice from assigned faculty mentors, but pulling together the kinds of support needed to prepare for and conduct research was challenging. We launched a new program in 2008 to bridge the Family Medicine Research Division staff and expertise to assist residents and meet the needs of the residency program.



**re·search:** NOUN: 1. a detailed study of a subject, especially in order to discover (new) information or reach a (new) understanding.

*Cambridge Dictionaries Online,*  
© Cambridge University Press 2003.

### Resident Requirements

The following lists the requirements for conducting research during the KU family medicine residency.

1. **Each resident chooses one of the research division faculty to act as a research mentor** and assist with hypothesis generation, literature searching, project design, human subjects approval, and project

- implementation. The research mentor works collaboratively with the resident's clinical mentor to help the resident complete a research project.
2. **All residents must attend a two hour quarterly "Core" session.** The purpose of these sessions is to receive training about various research topics, meet with research faculty, and to co-learn about primary care research topics.
  3. **All residents must complete an institutional online human subjects tutorial** which certifies them for work on research projects.
  4. **All residents, either individually or in paired research teams, must complete an original research project** at least one semester prior to their planned graduation date.
  5. **All residents, either individually or in paired research teams, must make a presentation** of their research findings to the faculty and other residents at a scheduled forum.

### **Residency Training Competencies**

Clinical care excellence requires the physician to weigh medical evidence to determine best management for a given health problem. Research plays a crucial role in this process. Participating in research and scholarly projects assists the resident to gain skills needed to meet several training competencies, especially *Competency 5: Practice-based Learning and Improvement*. Research also contributes to other competencies including *Competency 1: Knowledge* as well as *Competency 2: Patient Care*.

Typically, information about our clinic's patient population and the care provided is gathered and interpreted routinely. Residents help identify specific clinical questions and help specify a study design, assist in data collection and data analysis. Findings are compared with literature retrieved using information technology such as PubMed. Published studies are critically appraised, and the strengths and limitations of the information gathered for the resident's own study are evaluated. Recommendations pertinent to patient care practice are made in the scientific papers written or posters presented.

### **Learning Objectives**

After completing the research project, the resident will be able to:

1. Appraise the strengths and weaknesses of scientific evidence from studies relevant to patient care.
2. Describe the study design features needed to generate valid data in an epidemiologic, case or population health study.

3. Conduct a practice-based or community-based research project and draw appropriate conclusions about potential ways to improve patient outcomes or population health.

These objectives are in keeping with the program requirements for residency education in family medicine put forth by the Residency Review Committee. The program requirements include this statement about research:

“The quality of the educational experience within the department of family medicine is enhanced by an active research environment. Programs should involve residents in research projects.”

It is important to remember that, to achieve these objectives, research has to be recognized as a process. It has been compared to an



hourglass, starting with observations and general questions that get refined, addressed through the collection and analysis of data and finally generalized to contribute to knowledge.

*Every resident is capable of completing a research project that contributes to the improvement of patient care, safety, community or public health, discovery of new information, and advancing the science of family medicine.*

## Timeline

**All residents should choose their research topic during PGY-1**, and no later than the beginning of their PGY-2. Early starts are encouraged. Data collection and analysis should be done during the second year, with **most of the work done no later than the end of PGY-2**. Presentation of research findings must be shared using a PowerPoint<sup>®</sup> presentation format at the annual Resident Research Forum held in April (the specific date will be set by the Residency Director and shared through GroupWise<sup>®</sup> calendaring as well as announced during residency meetings.) Depending on each resident’s project, they will present their findings at the April meeting during their PGY-2 or PGY-3 year.

A model timeline to guide your planning is presented below. While the exact nature of the research project and other residency requirements may dictate deviations from this pace, the timeline provides a pattern of accomplishment that should be emulated.

### Year 1: Identifying and Refining the Research Project

- July - Obtain research requirements, guidelines, and list of mentors and possible research topics
- Oct. - Identify research mentor and define the research question
- Nov. - Complete literature search
- Dec. - Complete written project proposal (including background, objective, and method)
- Jan. - Complete funding assessment
- Feb. - Present proposal to clinical and research mentor, faculty/resident committee if assigned
- Apr. - Complete Institutional Review Board (IRB) submission to obtain Human Subject Committee approval unless exempted
- May – Begin project; collect data

### Year 2: Collecting Data

- Aug. - Meet with research mentor to review progress of data collection
- Feb. - Meet with research mentor to review progress of data collection
- Apr. - Complete data collection
- Apr. - Present data and preliminary analysis to clinical and research mentor, faculty/resident committee if assigned OR
- Apr –Complete and submit written manuscript and/or poster and PowerPoint® presentation to research mentor and Residency office
- **Apr. – Presentation at annual Resident Research Forum**

### Year 3: Analyzing Results, Writing the Manuscript and Presenting at Resident Research Forum

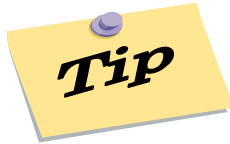
- Sep. - Complete data collection
- Oct. - Present data and preliminary analysis to clinical mentor and research mentor, faculty/resident committee if assigned
- Dec. - Complete data analysis
- Apr. - Complete and submit written manuscript and/or poster and PowerPoint® presentation to research mentor and Residency office
- **Apr. – Presentation at annual Resident Research Forum**

Residents Interested in Fellowships should also consider the following:

- Begin developing research ideas by the end of the first year.
- Complete a research project suitable for publication by the second year.
- Conduct and/or participate in at least one additional scholarly project.
- Submit an abstract to a national meeting.
- Present one or more research projects at the KUMC annual Family Medicine Research Forum.

For more information about possible fellowships, please consult with the Research Division Director as early in your residency as possible.

## Steps in Research Project Development, Implementation and Completion



The Department's research division has examples of previous resident research projects. Take a look at earlier research plan documents, questionnaires, chart audit forms, etc. as you are planning your project. "Learn from those who have gone before..."

As residents begin developing their required research project, it is helpful to think of the project as a series of discrete tasks (Hulley 1988).

1. Formulate a research question or hypothesis.
2. Perform a literature review.
3. Solicit input from at least one clinical mentor as well as the residency director, Dr. Belinda Vail.
4. Review scientific merit and epidemiologic aspects of project with Drs. Allen Greiner, Kim Kimminau, or Tony Wellever.
5. Consider costs and a budget. If the project requires funding for costs such as blood tests, mailing costs, etc., a grant application to an external agency is typically needed. Prepare a budget and discuss with your faculty mentor, research mentor and research division financial staff (Fanta Kuhlman).
6. Write a research plan, review it with mentors and revise as appropriate.
7. Complete all requirements for Human Subjects Committee approval and any other institutional or hospital policies (with assistance of research support staff).
8. Follow all ethics guidelines for research when conducting the project.  
Note: You must complete the online HSC tutorial found on:  
[http://www2.kumc.edu/researchcompliance/human\\_subjects\\_tutorial\\_inst.htm](http://www2.kumc.edu/researchcompliance/human_subjects_tutorial_inst.htm)
9. Enter data into an Excel spreadsheet or other analytic software (like SAS, SPSS, STATA), and analyze it with assistance of research staff.
10. Interpret results of statistical analyses with the aid of the clinical mentor.
11. Summarize the project's background, methods, results and conclusions in a written report, using standardized Departmental format (described below).
12. Present the results of the project at the annual *Resident Research Forum* (held in April). Presentation at additional scientific meetings such as the annual Kansas Public Health Association meeting or the American Academy of Family Medicine is encouraged.
13. Submit study for publication and presentation at a national meeting if warranted.
14. Provide all original consent forms, data collection sheets and electronic files to the residency office staff prior to graduation.

## Research Record Retention Policy

The Department's research unit must retain records from your project for several reasons:

1. To meet Departmental and Human Subjects Committee requirements.
2. Faculty members, who will remain at our institution, are the principal investigators of all research involving residents.
3. The University requires this. The policy is available at: <http://www2.kumc.edu/compliance/policies.htm>. Research records are defined as:

“Information recorded for the purpose of a research study, regardless of form or the media on which it may be recorded. Research Records may include technical data, computer software, laboratory worksheets, memoranda, notes or exact copies thereof that are the result of original observation and activities of a study, and any records that are necessary for the reconstruction and evaluation of reported results of the research and the events and processes leading to those results. Items which constitute research data under this policy include, but are not limited to: laboratory notebooks, samples of chemicals and materials synthesized during research, field specimens, voucher specimens, computer files or other electronic data, video tapes and audio tapes.”

## Research Support

Technical support available from the Research Division faculty (Drs. Greiner, Kimminau and Wellever) and the research support team includes:

- Consultation on study design features (e.g. control of bias and confounding)
- Assessment of statistical power and sample size needed
- Survey design and administration
- Data entry tips and database development
- Advice on data collection and management
- Scientific writing assistance
- Procedures for protecting human subjects (e.g. confidentiality)
- Other assistance with IRB applications
- Data analysis using statistical software
- Assistance making graphs of data findings (e.g. using Excel or PowerPoint)
- Feedback on drafts of the research report, copyediting
- Assistance submitting abstracts to scientific meetings
- Help with journal article submission.

Do not hesitate to ask for help! That's one of the services the Research Division provides in our department.

## Choosing a Topic

To help select a topic, think about:

- Science: What gaps exist in the medical literature?
- Clinical mentor: What topics or projects are being conducted by a Faculty member with whom you would like to work?
- Personal: What interests you? Have you had a patient with a problem for which the literature is not adequate? Are you interested in applying for a fellowship in a specific field?

## Study Design

Here are example study designs from previous resident projects:

1. **Cross-sectional studies**, e.g., a survey or in medical record review.
2. **Retrospective cohort studies** conducted in our longitudinal patient records.
3. Assessment of value of a **new test** compared to a standard diagnostic procedure.
4. Determining the **inter-observer variability for a test** when read by different physicians or technicians.

The study question will suggest the type of study that should be done.

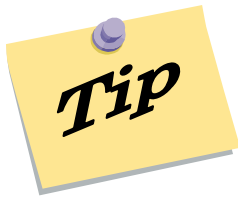
Epidemiologic Concept	Clinical Problem	Study Design Choices
Distribution of illness and population characteristics	<i>Who, What, When, How often?</i>	<i>Assess prevalence or incidence</i>
	Examples: <u>Health problems</u> Is diabetes increasing? <u>Prognostic factors</u> How common is obesity?	Cross-sectional
Determinants of illness or recovery	<i>Why? How can we help patients?</i>	<i>Assess causality or efficacy</i>
	<u>Risk factors</u> Which patients are most likely to have a prolonged hospital stay? <u>Interventions to improve outcomes</u> Prevention Treatment	Cohort or case-control  Controlled trial (definitive) Cohort or case-control (generates hypotheses)

## Study Feasibility

The clinical faculty and research staff will help you consider:

- **Sample size:** How many patients are needed for adequate statistical power?  
How many patients are available in our practice?
- **Time:** How much time will it take to collect the data?
- **Costs:** Are there special costs other than investigator time and routine copying?

The resident should discuss these types of issues with their research and clinical faculty mentors. A focused approach that answers a narrow question is better than a broad investigation of a larger topic with insufficient resources to complete the project. For larger projects, separate the project into Phase I and Phase II, and team up with a resident who will graduate a year or two after you. Collaborating with another institution is acceptable as long as the resident also closely involves a clinical faculty member in our Department, and analyzes the data generated at our institution.



A modest preliminary study can lay the groundwork for a future large study.

For large projects, separate the study into Phase I and Phase II. You do Phase I. Team up with a resident who will graduate a year or two after you for Phase II.

Collaborating with another institution to generate enough sample size is acceptable as long as you closely involve a clinical faculty member in our Department, your data is analyzed here, and all University and Departmental policies are followed.

## Refining the Research Question

After choosing a general topic area, the resident should:

1. Conduct a literature search to evaluate the current research in the area
2. Talk with faculty about their interests and current projects
3. Decide what specific project in which he or she is interested

After the general question has been decided upon, it should be narrowed down into a specific answerable research question or hypothesis. For example:

*General question:* What factors influence whether patients stop smoking?

*More specific research questions:* Is the use of Chantix or Zyban more effective among low-income patients who are attempting to quit smoking? What indicators can be established among patients that smoke that would provide guidance for prescribing them the most effective drug?

## Developing the Research Plan

Resident should prepare a written research proposal in collaboration with their mentors. Research support staff can provide examples of previous proposals to use as models. The proposal should provide a concise review of the relevant scientific literature pertaining to the research topic. The research plan section of the proposal should describe in adequate detail the question to be addressed, the study population(s) involved, the methods to be used and the analytic plan. It should include:

Title

Abstract

Specific Aims (list of main deliverables including number of study subjects)

Background (introduction and literature review)

Methods:

Subject recruitment and selection

Data collection (attach data collection forms)

Statistical methods:

- Data management (spreadsheet software, confidentiality procedures)
- Analysis methods to be used (include control of confounders)
- Statistical tests to be applied
- Sample size and power estimates (for the major study outcomes)

Expected results

Literature cited

Biosketches of investigators (research staff has most faculty biosketches on file).



Use what you write for your research plan's introduction, methods and literature cited in your final paper or poster.

Research staff will show you how to write the statistics section and help with the power estimates.

## Approval of Research Project

After the research protocol has been developed, it should be approved by the resident's research faculty mentor, and it also will be reviewed by the Research Division. Depending on the topic or scope of the project, an additional meeting with other faculty and/or the residency director may be required.

***NOTE: Human Subjects Committee (HSC) approval is needed before the resident initiates the research activity.***

The KU School of Medicine requires justifying the participation of human subjects in research, and promotes the protection of the welfare, rights and privacy of those subjects. The resident (and co-investigators) need to complete the KU on-line human subjects tutorial and other institutional requirements well in advance of the planned research project. In some cases, clearance both by the KUMC HSC and another IRB is needed, which is another reason for residents to start their projects early. The experienced epidemiology research staff in the Department is glad to help residents write and present their proposals in a way that improves the likelihood of clearance by the HSC.



IRB stands for "Institutional Review Board". The Human Subjects Committee (HSC) is designated as the Institutional Review Board (IRB) for the University of Kansas Medical Center. The HSC is responsible for reviewing, approving, modifying, rejecting and monitoring research involving human subjects.

## Planning for Data Collection

Paper forms and surveys are typically used to collect data, and then the resident enters the data into a spreadsheet or other analytic software. If data are entered into Excel or Access, they can easily be exported into an SPSS<sup>®</sup> (Statistical Package for the Social Science, Chicago, IL) if the Excel<sup>®</sup> or Access<sup>®</sup> database was set up in the most efficient way. If not, exporting can be very time-consuming and prone to keying errors. Please consult with the research staff during design and planning to help minimize errors or need for time-consuming recoding or re-entry of data later on. See page 9 for Human Subjects (HSC) and other requirements that apply to data collection and storage.

## Pre-testing and Revising the Research Plan

Once departmental and HSC approval have been obtained, the study may begin. It is useful to test the study plan on a small scale to identify process problems before the larger study is conducted. If a medical record audit is conducted, this

pre-test may suggest additional variables to collect. If a survey is administered, unclear questions may need to be revised. If significant changes are needed to the research protocol, it is necessary to apply for a revision through the HSC.

## Analyzing the Findings

Below is a highly simplified flowchart of some common statistical analysis procedures. An appendix (page 21) shows how to read SPSS output, and provides the formula for calculating 95% confidence intervals (CI) for a proportion.

### 1. Descriptive Statistics (one variable at a time):

#### **Categorical variables**

What are they? Items coded as Yes/No or mild/moderate/severe

What do we calculate? **Frequencies, proportions, rates, 95% CI**

#### **Continuous variables**

What are they? Items like age, weight, systolic blood pressure

What do we calculate? **Means, standard deviations, medians**

### 2. Analytic Statistics for Two Variables

#### **Categorical variables:**

**Degree of control of diabetes (excellent, fair, poor) among those with controlled and uncontrolled hypertension**

Is preeclampsia more likely among teen moms?

If you have a *large sample*, generate a *P* value using the **Chi-Square test**

#### **Square test**

If you have *small numbers*, use **Fisher's Exact test**

**Continuous variables:** Is hypertension more common in smokers?

Is the variable *skewed*?

No (bell-shaped curve): Use **independent samples t-test**

Yes (a few non-smokers, a lot of smokers): Use a nonparametric test, e.g. **Mann-Whitney test**

### 3. Analytic Statistics for More than Two Variables

#### **Categorical variables**

Same as for two variables.

#### **Continuous variables**

Is the variable **skewed**?

No: Use **ANOVA** (analysis of variance)

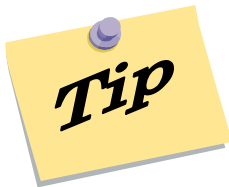
Yes: Use a nonparametric test, e.g. **Kruskal-Wallis test**

#### 4. Controlling for Confounding Variables

Poor control of confounding can ruin a study. An incorrect conclusion is reached because the true cause of the health outcome was something other than what the data showed.

Suppose the data shows that patients who had a prior attempt to quit smoking did better with Chantix than patients who had never tried to quit (i.e., more successful quits occurred with the prior attempt to quit group,  $P < .001$ ). But the patients who had never tried to quit were more likely to be depressed.

The patient's mental health is a confounder, and could be the true reason why the patients with prior attempts did better – it may have had little to do with Chantix.



Be sure you collect enough information about your study population. Then you will have the data to describe key demographic and medical characteristics of the group you investigated. You will also be able to control for key confounders. The audience expects this information during your presentation and in your paper.

There are three main ways to control for confounding.

##### **Subject Eligibility Criteria:**

Only let mentally healthy/non-depressed patients into the study.

*Problem:* You won't find out about depressed patients' needs or their response to Chantix or if depression is prevalent in your patient population, you may not have enough non-depressed patients to do the study.

##### **Stratified Analysis:**

Analyze results for depressed patients separately from non-depressed patients.

*Problem:* There is always more than one confounder (e.g., obesity). So you'd have to subdivide first by mental health/depression status, then by weight. Now you have four groups instead of only two. If small sample size, there will not be enough power to do many subgroup analyses.

##### **Regression Analysis:**

Calculate an odds ratio for the adjusted for the multiple risk factors (e.g. depressed and non-depressed, BMI  $\leq$ 30, age, gender, number of quit attempts, prior smoking cessation medications).

*Problem:* Need statistical software. Although a regression controls for several risk factors at a time, the number of confounders that can be put in a model is limited if sample size (number of patients in the study) is fairly small.

Drs. Greiner, Kimminau, or Wellever will help you write this section of your research plan, and will help you with the statistical analysis after you collect your data.

### **Drawing Conclusions**

The conclusions of a research project should include:

1. What did the results show?
2. Were the results statistically significant?
3. What were the study's limitations?
4. Were the results clinically important?
5. What future research is needed?

### **Disseminating Results**

Residents must prepare a written report summarizing their findings. This written report should be reviewed and approved by the faculty mentor and coauthors before it is submitted to the residency program director.

Resident papers are typically 7-10 single-spaced pages for the introduction, methods, results and conclusions. See the research support staff for examples of previous resident papers and for help with the Departmental style (e.g. subheadings, format for tables and figures).

Place the tables and figures at the **end of the manuscript** (after the references) instead of inserting them in the body of the results section. If your tables, figures or PowerPoint are giving you a hard time, call the research team.

### **EndNote®**

The research team highly recommends that you take a short EndNote® class offered routinely through KUMC Dykes Library to make your writing and reference citation as easy as possible. EndNote® allows you to build a library of literature relevant to your research and it will automatically format those references in a particular journals' reference style that you select. This saves an enormous amount of time and lets you use your literature search information to its full advantage. If you need help or want to access already constructed Endnote libraries on particular topics, let your research mentor know.

## PowerPoint®

The slide presentation for the *Family Medicine Research Forum* follows a standard, 10-15 minute format. Residents should review the content of their presentation with their mentors and co-authors, and practice delivery in advance. The research team will be happy to schedule some “test runs” and give you feedback on your slides and information you want to present. Don't forget to thank your fellow residents if they helped you recruit patients, and others who were of assistance as well as your funder, if you received resources to conduct your project. A standard KUMC PowerPoint template can be downloaded at: <http://www2.kumc.edu/ir/ppt/>. If you need help finding photographs, artwork or specific element that would assist you in making an interesting and compelling presentation, just ask – the research team will be happy to assist.

Residents are also encouraged to discuss with their mentors submission of an abstract to a professional conference. This can be a lot of fun. Building a portfolio of scholarly accomplishments can help when applying for fellowships. Funding for travel to a national meeting may be available if the presentation occurs during residency (but usually not if the presentation occurs after graduation).

Certain projects will merit submission for publication in a professional journal. Your clinical mentor and the research staff will help you with this.

## Faculty and Staff Resources

Our faculty members' research ideas are very good so please brainstorm with them at your earliest opportunity. Studies designed and executed by our faculty members and residents have passed the test of peer review, resulting in numerous recent journal articles and abstract presentations at local, regional and national meetings. Faculty members to consult include but are not limited to:

### RESEARCH DIVISION STAFF:

Allen Greiner, MD, MPH (Director)  
Kim Kimminau, PhD  
Tony Wellever, MA  
Joshua Freeman, MD  
Heidi Chumley, MD  
Jon Delzell, MD

### AFFILIATED RESEARCH FACULTY:

Mugur Geana, MD, PhD (KU School of Journalism)  
Christine Daley, PhD ( Dept. Preventive Medicine and Public Health)  
Ed Ellerbeck, MD, MPH( Dept. Preventive Medicine and Public Health)  
Lauren Aaronson, PhD, RN (Deputy Director, Heartland Institute for  
Clinical and Translational Research)

The contact information for the Department of Family Medicine Research Division is located at 1009 Wescoe Hall, KUMC, 3901 Rainbow Blvd, Kansas Ctiy, Kansas 66160.

Allen Greiner, MD, MPH  
Director, Family Medicine  
Research Division  
Wescoe Hall 1009

Photo here

588-1956 ( Research office  
number)

Department of Family Medicine  
Research Division  
2009

## Recommended Readings

- Gordis L. Epidemiology, 3<sup>rd</sup> edition. Philadelphia, Saunders, 2004.  
*Popular textbook with information on study design, bias, confounding, sample size estimation, kappa statistics, interesting stories from the history of epidemiology and more.*
- Grimes DA, Schulz KF. Epidemiology Series. Lancet 2002;359: No. 9300-9310.  
*A series of articles printed in Lancet, giving a general overview or primer of research methods. It is designed for the clinician.*
- Hulley, Stephen B. and Cummings, Steven R. Designing Clinical Research. Baltimore, Williams & Wilkins, 1988 or 2<sup>nd</sup> edition, 2001.  
*A text on planning and implementing clinical research, for beginning investigators. Has useful tables for estimating the sample size needed to achieve statistical significance when the difference between two rates is an important outcome of a study.*
- Lwanga SK, Lemeshow S. Sample size determination in health studies: A practical manual. Geneva, World Health Organization, 1991:2-5.  
*Has useful tables for estimating the sample size needed to achieve a 95% CI of a certain width when a rate is a key outcome of a study.*
- Neale AV, West P, French L. Surviving Your Resident Research Requirement. JAMA. 1998;280:1802.  
One page overview of advice for the resident.
- Norman GR, Streiner DL. Biostatistics: The Bare Essentials. St. Louis, Mosby, 1994.  
*Explains statistical concepts in plain English with a minimum of equations, and using humorous examples.*
- Rosner B. Fundamentals of Biostatistics, Fourth Edition. Boston, Duxbury Press, 1995.  
*Standard textbook with a detailed explanation of basic biostatistical concepts.*

# Appendix I

## How to calculate 95% CI on a proportion

**Formula:**  $p \pm Z_{1-\alpha/2} \sqrt{p(1-p)/n}$

Proportion = p. For 37.8%, p = .378

Z has to do with the area under the bell shaped curve.

$Z_{1-\alpha/2} = 1.96$  when you want the typical 95% CI.

**Example:** In 37.8% of 185 cases, antibiotic guidelines were followed completely  
“\*” signifies multiplication in example below.

**Lower 95% CI:      Subtract the quantity from p**

$$p - 1.96 * \text{square root of } [p * (1-p)/n]$$

$$0.378 - 1.96 * \text{square root of } [0.378 * 0.622/185]$$

$$0.378 - 1.96 * \text{square root of } 0.00127$$

$$0.378 - 1.96 * 0.0356$$

$$0.378 - 0.070$$

$$0.308 \text{ or } 30.8\%$$

**Upper 95% CI:      Add the quantity to p**

$$P + 1.96 * \text{square root of } [p * (1-p)/n]$$

$$0.378 + 0.070$$

$$0.448 \text{ or } 44.8\%$$

## Appendix II

### How to Read an SPSS Printout

#### A. Frequencies: Percent and Valid Percent

**Example I:** Twin Deliveries at KU Hospital from 1998-2004. What was the race/ethnicity of the mothers of twins? SPSS gives you information to express the results in two ways, either of which is correct.

**Output 1. Using the “Percent” column:** You would say “the race/ethnicity of the moms was: White, 78.2%; Hispanic, 8.3%; Black, 7.3%; Other, 0.9%; **Unknown, 5.3%.**” These add up to 100%.

**Output 1. Mother's race**

		Frequency	<b>Percent</b>	Valid Percent	Cumulative Percent
Valid	White	506	<b>78.2</b>	82.5	82.5
	Hispanic	54	<b>8.3</b>	8.8	91.4
	Other	6	<b>.9</b>	1.0	92.3
	Black	47	<b>7.3</b>	7.7	100.0
	Total	613	<b>94.7</b>	100.0	
Missing	System	34	<b>5.3</b>		
Total		647	<b>100.0</b>		

**Output 2. Using the “Valid Percent” column:** You could also say “Among women whose race/ethnicity was reported, the distribution was: White, 82.5%; Hispanic, 8.8%; Black, 7.7%; Other, 1.0%.” These add up to 100%. You ignore the “Percent” column.

**Output 2. Mother's race**

		Frequency	Percent	<b>Valid Percent</b>	Cumulative Percent
Valid	White	506	78.2	<b>82.5</b>	82.5
	Hispanic	54	8.3	<b>8.8</b>	91.4
	Other	6	.9	<b>1.0</b>	92.3
	Black	47	7.3	<b>7.7</b>	100.0
	Total	613	94.7	<b>100.0</b>	
<del>Missing</del>	<del>System</del>	<del>34</del>	<del>5.3</del>		
Total		647	100.0		

**Example II:** Twin Deliveries at KU Hospital during 1998-2004. What instrument was used in these operative vaginal deliveries?

**Output 3.** This time, it is not correct to use the “Percent” column. That’s because there are missing values for moms who did not have an operative vaginal delivery because, for example, they had a cesarean. So use the “Valid Percent” column.

**Output 3. Instrument used in AVD**

		Frequency	<del>Percent</del>	Valid Percent	Cumulative Percent
Valid	Vacuum	18	<del>2.8</del>	58.1	58.1
	Forceps	12	<del>1.9</del>	38.7	96.8
	Both	1	<del>.2</del>	3.2	100.0
	Total	31	<del>4.8</del>	<b>100.0</b>	
<del>Missing</del>	<del>System</del>	<del>616</del>	<del>95.2</del>		
Total		647	<del>100.0</del>		

**B. Continuous Variables: Checking for Skewness.**

A skewness test is an option that can be chosen when calculating means. If the skewness test says the sample is skewed, a non-parametric test should be used to calculate a P value, instead of the standard tests. The sample is skewed if “Skewness” is > 1.0 or is < negative 1.0.

**Output 4.** The BMI distribution is skewed at both KU Family Medicine clinic and JayDoc as shown below.

**Output 4. BMI Pre-pregnancy**

	Median	Mean	SD	Skewness	N
JayDoc	24.66	26.24	6.74	1.271	20039
Clinic	23.06	24.17	4.99	1.213	13061
Total	23.95	25.42	6.19	1.369	33100

**C. Comparing Proportions: The Chi-Square or Fisher’s Exact Test.**

**Output 5.** The table on the next page shows how many cases, the row percents and the column percents for a group of women with preeclampsia who had Cesareans.

For MgSO<sub>4</sub>, **0 means no and 1 means yes**. 72.1% of women with primary C/S received MgSO<sub>4</sub>, compared to 64.0% of women with repeat C/S.

**Output 5. Mode of delivery if C/S \* MGSO4\_GIVEN Crosstabulation**

			MGSO4_GIVEN		Total
			0	1	
Mode of delivery if C/S	Primary C/S	Count	159	410	569
		% within Mode of delivery if C/S	27.9%	<b>72.1%</b>	100.0%
		% within MGSO4_GIVEN	66.0%	73.7%	71.4%
	Repeat C/S	Count	82	146	228
		% within Mode of delivery if C/S	36.0%	<b>64.0%</b>	100.0%
		% within MGSO4_GIVEN	34.0%	26.3%	28.6%
Total	Count	241	556	797	
	% within Mode of delivery if C/S	30.2%	69.8%	100.0%	
	% within MGSO4_GIVEN	100.0%	100.0%	100.0%	

**Output 6.** The table below shows the P values for comparing the proportions, 72.1% v s 64.0%. In SPSS-speak, “**Asymp Sig.**” means **P value**. The table gives us a lot of statistics. We want to use the **Pearson Chi-Square** unless any cells had expected counts less than 5. If so, we would use the Fisher’s Exact test. For the table below, we would use the P value of **.026**.

**Output 6. Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
<b>Pearson Chi-Square</b>	4.965(b)	1	<b>.026</b>		<b>.027</b>
Continuity Correction(a)	4.592	1	.032		
Likelihood Ratio	4.874	1	.027		
<b>Fisher's Exact Test</b>					
Linear-by-Linear Association	4.958	1	.026		.017
N of Valid Cases	797				

a Computed only for a 2x2 table

b **0 cells (.0%) have expected count less than 5**. The minimum expected count is 68.94.

#### D. Logistic Regression.

**Output 7.** This is a logistic regression to see how much of the risk of **shoulder dystocia** may be due to **diabetes** in our population. We are using the regression to control for other confounding influences, such as the mother's BMI.

The part of the logistic regression print out in the table below shows how many cases were included in the regression. If a case has missing data for any of the variables in the regression, it was excluded from the analysis. This regression only had 2.9% cases with missing data.

#### **Output 7. Case Processing Summary**

Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	<b>4694</b>	<b>97.1</b>
	Missing Cases	138	2.9
	Total	4832	100.0
Unselected Cases		0	.0
Total		4832	100.0

a If weight is in effect, see classification table for the total number of cases.

**Output 8.** The following table tells us how well our variables help to predict whether the patient had shoulder dystocia. The Cox & Snell R square value tells us that the regression model explained 7% of the variance in the data. Since that's not very much, it means we don't have information on some of the other important risk factors. Or maybe many of the true risk factors have not been discovered.

#### **Output 8. Model Summary**

Step	-2 Log likelihood	<b>Cox &amp; Snell R Square</b>	Nagelkerke R Square
1	1526.705(a)	.070	.212
2	1526.928(a)	<b>.070</b>	.212
3	1527.150(a)	.070	.212

a Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

**Output 9.** Results are shown on the next page. Unfortunately, they come in SPSS-speak. On the table below, the text originally in the SPSS print out is in the smaller, non-bold font and the translation that has been inserted is in the larger, bold font.

- Sig means P value.
- Exp(B) means adjusted odds ratio
- Upper/Lower are the 95% CI for the adjusted odds ratio.

- The brief variable name of each potential risk factor is listed instead of what it stands for, but there is another page that provides the translation.
- How the variables were coded determines what was the “reference value” to which the other levels of the variable were compared. For example, each level of parity was compared to  $\geq 3$ .

Findings: Diabetes was independently associated with a 22.6% increase in the risk of shoulder dystocia (Adj. OR 1.226), but the association was not statistically significant ( $P=.631$ ).

### Output 9. Variables in the Equation

		B	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
			<b>P Value</b>	<b>Adjusted OR</b>	Lower	Upper
Step 2(a)	<b>labor2 (vs. spontaneous)</b>		.002			
	labor2(1) <b>Augmented</b>	.536	.004	1.709	1.190	2.454
	labor2(2) <b>Induced</b>	.603	.001	1.828	1.289	2.594
	<b>delmode1(1)AVD (vs. CS)</b>	4.259	.000	70.742	26.055	192.069
	<b>paratot4 (compared to <math>\geq 3</math>)</b>		.000			
	paratot4(1) <b>0</b>	-.157	.641	.854	.441	1.655
	paratot4(2) <b>1</b>	.333	.341	1.395	.703	2.768
	paratot4(3) <b>2</b>	.647	.083	1.909	.919	3.967
	<b>bmi_pcat1a (vs. lean BMI &lt; 18.5)</b>		.052			
	bmi_pcat1a(1) <b><math>\geq 30</math></b>	1.033	.014	2.810	1.228	6.428
	bmi_pcat1a(2) <b>25-29.9</b>	.906	.028	2.475	1.102	5.558
	bmi_pcat1a(3) <b>18.5-24.9</b>	.704	.079	2.022	.921	4.437
	fet_intol_any4_var(1)	.638	.000	1.893	1.415	2.531
	<b>Diabetes(1)</b>	.204	<b>.631</b>	<b>1.226</b>	<b>.534</b>	<b>2.814</b>
	Constant	-8.717	.000	.000		

a Variable(s) entered on step 1: year, labor2, delmode1, paratot4, bmi\_pcat1a, fail\_progress\_avd\_cs\_reason, fet\_intol\_any4\_var, Diabetes.

Note: the rows for “year” on the table above were deleted, and several columns were deleted to make the table simpler and smaller.

## Appendix III

### Null Hypothesis

The null hypothesis is a hypothesis which the researcher tries to disprove, reject or nullify. The 'null' often refers to the common view of something, while the alternative hypothesis is what the researcher really think is the cause of a phenomenon.

The null hypothesis,  $H_0$ , is an essential part of any research design, and is always tested, even indirectly. The simplistic definition of the null is as the opposite of the alternative hypothesis,  $H_1$ , although the principle is a little more complex than that.

An experiment conclusion always refers to the null, rejecting or accepting  $H_0$  rather than  $H_1$ . Despite this, many researchers neglect the null hypothesis, which is poor practice and can have adverse effects.

#### DEVELOPMENT OF THE NULL

Up until the 1500's most people thought that the world was flat (At the time: The Null hypothesis). Columbus challenged this idea with the alternative hypothesis: The world is round. Then most people thought that the earth was the center of the universe (Null hypothesis). Copernicus had an alternative hypothesis that the world actually circled around the sun, thus being the center of the universe. Eventually, people got convinced and accepted it as the null.

Later someone proposed an alternative hypothesis that the sun itself also circled around the something within the galaxy. This is how research works - the null hypothesis get's closer to the reality each time, even if it isn't correct, it is better than the last null hypothesis.

#### EXAMPLES OF THE NULL HYPOTHESIS

A researcher may postulate a hypothesis:

H1: Tomato plants exhibit a higher rate of growth when planted in compost rather than in soil.

And a null hypothesis

$H_0$ : Tomato plants do not exhibit a higher rate of growth when planted in compost rather than soil.

It is important to carefully select the wording of the null, and ensure that it is as specific as possible. For example, the researcher might postulate a null hypothesis:

H0: Tomato plants show no difference in growth rates when planted in compost rather than soil.

There is a major flaw with this null hypothesis. If the plants actually grow more slowly in compost than in soil, an impasse is reached. H1 is not supported, but neither is H0, because there is a difference in growth rates.

If the null is rejected, with no alternative, the experiment may be invalid. This is the reason why science uses a battery of deductive and inductive processes to ensure that there are no flaws in the hypotheses. Many scientists neglect the null, assuming that it is merely the opposite of the alternative, but it is good practice to spend a little time creating a sound hypothesis. It is not possible to change any hypothesis retrospectively, including H0.

## SIGNIFICANCE TESTS

If significance tests generate 95% or 99% likelihood that the results do not fit the null hypothesis, then it is rejected, in favor of the alternative. Otherwise, the null is accepted. These are the only correct assumptions, and it is incorrect to reject, or accept, H1.

Accepting the null hypothesis does not mean that it is true. It is still a hypothesis, and must conform to the principle of falsifiability, in the same way that rejecting the null does not prove the alternative.

## PERCEIVED PROBLEMS WITH THE NULL

The major problem with the null hypothesis is that many researchers, and reviewers, see accepting the null as a failure of the experiment. This is very poor science, as accepting or rejecting any hypothesis is a positive result. Even if the null is not refuted, the world of science has learned something new. Strictly speaking, the term 'failure', should only apply to errors in the experimental design, or incorrect initial assumptions.