Research Methods 101

Paul A. Thompson, Ph.D. [1]

[1]Director, Methodology and Data Analysis Center
Center for Health Outcomes and Prevention Research Sanford Research/USD
Professor, Department of Pediatrics, Department of Internal Medicine
Sanford School of Medicine of the University of South Dakota

June 3, 2013

Scientific Research

What do we mean by "scientific research"?

With Scientific research, we . . .

- determine facts
  - A fact is a larger truth
  - It is something true about a large group
- assess relationships between the facts
  - Are two facts related?
  - What kind of a relationship might exist?
- make causal statements about how the facts work together
  - Does Fact A $\rightarrow$ Fact B
  - Does Fact A $\rightarrow$ Fact C $\rightarrow$ Fact B

Example: The weight of a child

The weight of a child concerns parents and physicians

What about the weight of a child?

- Fact: How much does a 10 YO child weigh?
  - How can we determine this?
  - Should other things be considered?
- Relationship: Does weight affect activity?
  - Do light children play more?
  - Are heavier children less involved in activity?
- Causal: Children who play outside weigh less
  - Activity $\rightarrow$ Weight
  - Weight $\rightarrow$ Activity

Example: Sleep patterns

What does sleep do?

Sleep at night

- Fact: How often does a person wake up?
  - What is normal sleep?
  - Are there different patterns in situations?
- Relationship: Does sleep affect weight?
  - Heavy persons sleep differently?
  - Sleep patterns affected?
- Causal: Good sleep → weight
  - Children who play outside weigh less
  - Improving sleep improves weight control

Example: Alcohol use

Alcohol is a commonly used drug

Alcohol study

- Fact: How much do people drink?
  - When they drink, how much do they drink?
  - Do people fall into patterns of drinking?
- Relationship: Does alcohol use affect sleep?
  - Do drinkers weigh more than non-drinkers?
  - Is sleep different after drinking?
- Causal: Reducing alcohol → better sleep
  - Changing alcohol use changes sleep?
  - Non-drinkers sleep differently?

What are my challenges here today?

- DTYGTSE
- Not getting too far ahead of things
- Cultural issues
- Community-based participatory research
- Ensuring that methods are at the level of the user
- Presenting info from 3 grad courses in 1 talk in 1+ hours

Where do we start?

Scientific research starts with an observation or question

Question . . .

Are smart phones smarter than our kids?

- Kids use smartphones today
- What does this do to attention?
- Does this affect schoolwork?
- Technology is overwhelming knowledge

Fact: Cellphone use in children

A fact is a careful observation
Scientific research begins with scientific observations

Observations

- Casual
  - See or hear or smell
  - Many observations better
  - **data** $\neq$ **anecdote** $*$ 2
- Scientific
  - Planned carefully
  - Each step documented
  - Cases chosen carefully

What differs?

- Casual
  - Not systematic
  - Subject to bias
  - Arbitrary
- Scientific
  - Observations planned
  - Circumstances controlled
  - Bias minimized

Relationship: Cellphone use and activity

Once we have the facts down, we can think of relationships

Cellphones

- Children with cellphones ...
  - play more?
  - play less?
  - play differently?
- Groups of children ...
  - Similarities cellphone use?
  - Discuss cellphone use?
  - Play with all phones?
- Parents of children ...
  - Regulate usage?
  - Discuss usage?

Data collection

- Children and cellphones ...
  - Measuring play
  - Time, effort
  - Types of play
- Groups of children ...
  - Assessing similarity
  - Measuring discussion
  - Measuring play
- Parents of children ...
  - Measure
  - Discuss usage?

Causal: Increase Cellphone use $\rightarrow$ lower activity

Changing cellphone use

- Baseline measure
  - Cellphone use
  - Activity of children
  - Different types of children
- Intervention
  - Children in groups
  - Amount of cellphone use
  - Measure activity
- Endpoints
  - Time on phone
  - Compliance
  - Activity amount

Logic of the process

- Measure at baseline
  - How much phone?
  - Activity?
  - Which groups?
- Assign amount to groups
  - Low, med, high
  - Allow use to settle
  - Measures of activity
- Differences at endpoint
  - How much phone?
  - Activity?
  - Which groups?

Steps of scientific method-I

1. Devising an idea ...

1. This is often difficult
2. Knowing what's going on is important
3. The "zeitgeist" in your area is key

Steps of scientific method-II

2. Check other research . . .

1. Has research been done?
2. Perform new experiment
3. Redo old
4. Extend old
5. Correct incorrect stuff

Steps of scientific method-III

3. Hypothesis formulation . . .

1. The "formalized question"
2. What research will test
3. The idea must be shaped
4. A testable question is needed
5. Testability and measurement
6. "Operationalization"
7. "Threats to inference"

Steps of scientific method-IV

4. Run the experiment . . .

1. Assign the cases to treatments
2. Perform the treatment
3. Gather information to test the hypothesis

Steps of scientific method-V

5. Analyze data and draw conclusions . . .

1. Follow statistical analysis plan
2. Descriptive statistics
3. Tests of hypotheses
4. Use tests to evaluate hypothesis
   1. Data SUPPORTS hypothesis
   2. Data FAILS TO SUPPORT hypothesis

Steps of scientific method-VI

6. Write up the results

1. Publications and presentations, etc
2. Key part of process

What is the study?

Possible studies

- Effect of parent knowledge on cellphone use
- Effect of usage restrictions on cellphone use
- Effect of cellphone availability on usage
- Effect of cellphone use on attention span
- Effect of cellphone use on learning
- Effect of cellphone use on other activities

Many possible choices . . .

Basic designs in research

"Design" in research = plans for data collection

Components of design

- Time
  - One time observations
  - Two time observations
  - Multiple observations
- Dependence
  - Cross-sectional
  - Repeated measures
  - Cross-over
- Randomization
  - Used
  - Not used

Experimental designs-I

Quasi-experimental designs are often used

Cohort Study (QED)

- Basic idea
  - Group of subjects without outcome
  - Watch over time
  - See how many develop outcome
  - Retrospective (from records) or prospective (from subjects)
- Positive factors
  - Easy to do
  - Retrospective cohorts are common
- Negative factors
  - Lots of "threats to inference"
  - Not conclusive, merely suggestive

Experimental designs-II

Cohort Study-Example

Framingham Heart Study

- Began in 1948
  - 5200 adult subjects
  - Watch over time
  - See how many develop outcome - heart disease
  - Prospective
  - 1000+ papers from study
- Positive factors
  - Heart disease risk factor
  - Blood pressure is involved in stroke
  - Lipid profile research
- Recent developments
  - Diversity of cohort is larger

Experimental designs-III

Single observation designs also used

Cross-sectional designs

- Basic idea
  - Use group of subjects
  - Observe incidence of variables
  - Examine conditions during interview
- Positive factors
  - Easy to do
  - Hypothesis generating
- Negative factors
  - Lots of "threats to inference"
  - No causal flow can be observed

Experimental designs-IV

Example - cross-sectional design

Heart screening study

- Basic idea
  - People get a heart screening
  - Get family history, personal history
  - Screen for heart disease, other conditions
  - Use medical records to get more info
  - Provide FHS and calcium score for patients
  - Advise very ill persons to seek immediate attention
- Provides a good idea of current status
- Cannot do long-term followup

Experimental designs-V

Interrupted time series

Longitudinal design-one intact group

- Basic idea
  - Take intact group (class, club, workplace)
  - Measure current status
  - Provide an intervention
  - Measure final status
- Positive factors
  - Examine effect of the intervention
  - Change means intervention did something
- Negative factors
  - Hawthorne effect
  - No comparison group

Experimental designs-VI

Example - cross-sectional design

Worksite wellness studies

- Basic idea
  - Enroll a workplace
  - Measure baseline information (BMI, stamina)
  - Give an exercize intervention
  - Measure final information (BMI, stamina)
- This can hint at a causal relationship
- Linking in a second workplace in a cross-over design better

Experimental designs-VII

Randomized clinical trial

Longitudinal design-two randomized groups

- Basic idea
  - Take participants and assign randomly
  - Measure baseline status
  - Provide an intervention to Group A, but not Group B
  - Measure final status
- Positive factors
  - Randomization equates groups
  - Change means intervention did something
- Negative factors
  - Sometimes expensive
  - Control groups are difficult to run

Experimental designs-VIII

Example - randomized clinical trial

REPAIR-T1D study

- 54+ newly diagnosed T1 diabetic patients (many young)
- Measure baseline information (BMI, stamina)
- Provide medication vs placebo
- Measure results at B, 6m, 12m
- 2:1 randomization scheme
- Getting ready to examine data

Experimental design for cellphone use study

Changing cellphone use

- Cohort study
  - Children in groups
  - Amount of cellphone use
  - Measure activity over time
- Interrupted time-series
  - Children in groups
  - Measure activity at start
  - Provide intervention for parent monitoring
  - Measure activity over time and at end
- Randomized clinical trial
  - Work with school
  - Assign children at random to groups
  - Group A: Current treatment
  - Group B: Provide daily feedback to parents
  - Measure behavior at start and after 2 month
- Which is best?

Power analysis

Power analysis is used to plan studies

Thinking about power

- Statistical method to plan study
- Combines info about:
  - Study design
  - Planned statistical test
  - Hypothesized differences or results
  - Specific methods vary widely
- Necessary for studies
- Required by many publications

Basic ideas of statistics

What's the big idea?

title

- We want to understand "truth"
- Truth is found in populations - all the cases
- We can only get a sample
- We must measure the sample
- Using logic and math, we draw some conclusion about population
- That's the ENTIRE THING!

I have left out 2,536 texts and 118,382 journal articles in this summary

Statistical tests of hypotheses

title

- We start with an idea, the hypothesis
- There is another idea, the not-hypothesis or alternative
- Either the hypothesis or the not-hypothesis is true
- We set up a test based on the hypothesis
- The data may be consistent or inconsistent with the null hypothesis
- Based on the data, we may
  - REJECT the null hypothesis: Data disagrees
  - FAIL TO REJECT the null hypothesis: Data agrees
- That's the ENTIRE THING!

I have left out 2,536 texts and 118,382 journal articles in this summary

The bad word in statistics

We never say one dirty nasty word PROOF

title

- Using statistical methods, you can never prove something
- You collect data
- After analyzing it, you can say if
  - the data SUPPORTS or
  - FAILS TO SUPPORT
- Your data can be weird, biased, flawed

"I proved that snargle dingbats can gorbalize the nackitaz."
You may have data supporting gorbalization.
Proof of gorbalization is often difficult or impossible to obtain.

Statistical analysis concepts: Facts

Facts are a huge part of statistics: Descriptive statistics

Descriptive statistics

- Describe data values
  - Middle value: mean, median, proportion
  - Max and min
  - Variability: variance, IQR, standard deviation
  - Percentile values: value at 25%
- Incorporate population estimates
  - Confidence intervals (95% CI)
  - MOE (survey-speak)

Statistical analysis concepts: Relationships

Relationships: heart of most statistical tests

Relationships

- Correlation methods
  - Correlation itself
  - Regression with multiple variables
  - Different types of dependent variables
- Group differences
  - t-test
  - ANOVA
  - Mixed model methods
- Dependent variable must be considered

Statistical analysis concepts: Causal concepts

Causal notions: Design + statistics

Causal analysis

- Scientific ideas: aimed at causal statements
- If conditions are set correctly, causal statements possible
- Depends on several factors:
  - Ability to control independent variables
  - Ability to minimize "threats to inference"
- Causality comes from design
- Statistical analysis determines what can be said

Statistical analysis concepts: Recent developments-I

Recent developments in statistical methods - 1980 on

Recent developments

- Mixed models
  - Analysis of fixed and random factors
  - Fixed: Condition, time
  - Random: Things with a wider range (center, class)
- Multi-level models
  - Analysis of hierarchical data
  - Students in classes in schools
  - Patients seeing doctors in clinics

Statistical analysis concepts: Recent developments-II

Recent developments in statistical methods - 1980 on

Recent developments

- Non-Normal data
  - Normal data easiest
  - Binomial (2 choices) also easy
  - Poisson (count), multinomial (order category)
- Truncated data
  - Survival data
  - Censored data
- Difficult statistics
  - Bootstrap
  - Jack-knife
  - Monte Carlo methods

Statistical analysis concepts: Reproducible Research

Reproducible research methods

- Recent avalanche of medical fraud and incompetence
- Many cases especially in the area of genomics/PM
- Reproducible research:
  - Scripted data analysis
  - Clearly documented analyses
  - Tables and figures with direct analysis rules
- All interactive analysis is very problematic

Conclusions

QUESTIONS?