Reading and Interpreting Medical Research

EID 424
Bioengineering Applications in Sports Medicine
Prof. Kremenic
Sept 13, 2021

1

Overview

- Classes of studies
- Ethical research
- How to publish research
- Basic stats
- Reading and interpreting papers

2

Types of Papers

- Invited
- Review
- Original research

3

Invited Paper

- Author(s) are experts/at cutting edge in field
- Editor of journal asks them to submit paper on their expertise
- Largely summarizes their past work and looks to the "state of the art" in a particular area
- Beware the author that thinks a little too highly of himself

4

Review Paper

- Authors extensively summarize current and past research on a particular topic
- Can have well over 100 references
- Author(s) often experts in particular area
  - Often invited
- Extremely useful for getting up to speed in a particular field

5

Original Research Paper

- Authors are presenting information on original data they have collected
- Including development of tool/treatment
  - In this type of study, the methodology may be more interesting than the results

6

Types of Original Research Studies

- Retrospective
  - Chart review
- Prospective
- Randomized vs. paired (matched)
- Blinded vs. non-blinded
- The gold standard
  - Prospective, double-blind randomized study

7

Retrospective Studies

- Study of already-existing data
- Relatively easy to perform
  - Data are already there, just have to look it up
  - Chart review
    - Pull out charts on former/current patients, look up required information, test results, etc.

8

Problems with Retrospective Studies

- No control over confounding variables
- Need *lots* of data to show any kind of population effect
  - Nurses study
    - Huge dataset has over 100,000 respondents
    - Lots of good science
    - And…
      - Correlation between french fry consumption as a child and breast cancer?

9

Prospective Studies

- Examine the effect of a treatment vs a group receiving different/no/placebo treatment
  - Treatments decided ahead of time
  - Protocol for all subjects determined
- Can be difficult
  - People must consent to participating
  - All treatments (or lack thereof) must be ethical

10

Advantages of Prospective Studies

- Designed to answer a particular question
- Design should eliminate factors that may confound the issue

11

Randomized Studies

- Subjects are randomly assigned to groups
- Keeps investigators from choosing subjects who are more/less likely to respond to treatment
- Should lead to a relatively equal distribution of subjects between groups
  - Roughly same age range in groups
  - Roughly same body types
  - But, it may not

12

Matching Subjects

- Eliminate chance of demographics of one group being different from demographics of another
- Not easy to match large groups for
  - Gender
  - Height
  - Weight
  - Fitness level
  - Etc.

13

Blinding of Investigators

- Investigators examining subjects should not be aware of their treatment status
  - Could lead to biasing results of their examinations
  - It's hard to be *completely* objective

14

Double-Blind Studies

- Both investigators and subjects are unaware of which subjects belong to which group
  - Control for placebo effects
    - Can be 30%

15

The Research Gold Standard

- Prospective, double-blinded randomized study
  - Best way to show an effect of treatment, etc.
  - Designed to eliminate confounding factors, control covariates
  - Randomization should avoid bias in selection of subjects
  - Blinding of investigators and subjects should ensure integrity of results, account for placebo effects

16

Hydroxychloroquine [Plaquenil] and COVID-19 Controversy

- Many positive reports
- Is it an effective treatment for COVID-19?
- Most studies had no control group
- In most positive studies it was used in conjunction with a kitchen sink of other treatments
  - Azithromycin [antibiotic, "Z pack"]
  - Corticosteroids [powerful anti-inflammatory drugs]
- **FYI, hydroxychloroquine is NOT a useful treatment for COVID-19**

17

Ethical Research

- Research should not compromise treatment
  - Difficult to challenge the "gold standard" for treatment
  - Nor should study participants receive more attention
- Built-in protections for subjects
  - HIPAA
  - Informed consent
  - Institutional review board
  - Often required by journals in order to publish results of research

18

HIPAA

- Health insurance portability and accountability act
  - Passed in 1996
  - Been to the doctor in the past 25 years?
- Mainly deals with how your medical records may be exchanged/discussed among medical professionals, insurance companies, etc.
- Any subject data in a study must be de-identified
  - No way to link study data to a particular person
  - Data must be kept secure

19

Informed Consent

- Research subjects must be aware of and *understand*:
  - All possible consequences of their participation
  - Any benefits they may accrue
  - Should be *written* and *signed*

20

Institutional Review Board (IRB)

- Should review and approve all research at an institution before it is carried out
  - Including the informed consent form
- Made up of a combination of people
  - Scientists familiar with the topic
  - Other scientists
  - Non-scientific members of the institution
  - Members of the community

21

Publishing Your Research

- Carry out a quality study with a significant result
- Find the appropriate journal
  - Quality, well-respected
  - Read by people in the field with which you are dealing
- Write up research in the appropriate format
- Submit to journal editor for review

22

Peer Review

- Editor will send paper to several reviewers who should be experts in the field
  - Should be blind to the source of the paper
    - Not always possible
    - Reviewers often familiar with other researchers in the field
  - Accept as-is
  - Accept pending revisions
  - Revise and re-submit
  - Reject

23

Problems With Peer Review

- Can be hard to publish controversial, but correct findings
  - People may have trouble accepting something that goes against theory that they believe (or espouse)
- Dependent on ethics of reviewers
- Best system available

24

Writing Your Paper

- Know the relevant literature
- Know the relevant literature
- Know the relevant literature
- Know the relevant literature
- Write for the appropriate audience
- Always define acronyms before using them
- Use appropriate grammar
- Know the relevant literature

25

Suggestion for Paper Structure

- Abstract
- Introduction
- Materials and methods
- Results
- Discussion and conclusions

26

Writing Your Paper: The Abstract

- Short summary
  - Usually around 150 words
- Should include:
  - Purpose of study
  - Overview of methodology
  - Overview of results
  - Statement of what study showed
- The sales pitch
  - "What's it going to take to get you to read my paper?"
  - Often the only text publically-available
- Usually, the last thing written

27

Writing Your Paper: Introduction

- Summarize what is currently known about the topic
  - Cite all appropriate literature
    - Make sure you've actually read it as well
- Why is your project needed?
- What question does your project answer?
- State purpose at end
  - Along with any hypotheses

28

Writing Your Paper: Materials and Methods

- Describe what was done
  - The reader should be able to reproduce your study
- Subjects involved
  - Demographics
  - Inclusion/exclusion criteria
  - How were they divided into groups?
- Apparatus used
  - Hardware
  - Software
- Describe how analysis performed
- What statistics were used

29

Writing Your Paper: Results

- Charts and tables often the best
  - Yes, even in scientific literature everybody looks at the pictures first
- Text should elaborate on interesting results
- Do not speculate about the meaning of the results
  - Just state them

30

Writing Your Paper: Discussion and Conclusions

- Summarize results and provide interpretation
  - What's it mean?
- Include how your results relate to the existing literature on the topic
  - Hopefully in agreement with previous work
  - If not in agreement, explain why
  - How do your results extend current knowledge?
- Did you answer your research question?
  - Hypotheses proven/disproven?
- Weaknesses of your study
- Recommendations for further work

31

Introduction to Biostatistics

32

Sensitivity

- Probability of detecting disease/symptom, given that person has disease/symptom
- (number with positive test) / (total number of people with disease)
- *True positive rate*

33

Specificity

- Probability of not detecting disease/symptom given that person does not have disease/symptom
- (number with negative test) / (total number of people without disease)
- *True negative rate*

34

Related Terms

- Positive predictive value
  - Probability that a person has disease/condition given a positive test
- Negative predictive value
  - Probability that a person does not have disease/condition given a negative test
- False positive
  - Person who tests positive but is really negative
- False negative
  - Person who tests negative but is really positive

35

An Example

	Positive Test	Negative Test
Disease Present	150	30
Disease Absent	60	260

- Sensitivity = 150 / 180 = 0.83
- Specificity = 260 / 320 = 0.81
- Positive predictive value = 150 / 210 = 0.71
- Negative predictive value = 260 / 290 = 0.90

36

Real-Life Example

- Mammography
  - 75-90% sensitive
  - 90-95% specific
  - PPV:
    - 20%, women < 50
    - 60-80%, women 50-69

American College of Preventive Medicine Policy Statement

37

COVID-19

- With a completely new, highly infectious contagion, is a *sensitive* or *specific* test more important to prevent spread?
  - What are consequences of a *false negative* vs a *false positive*?
- Very non-theoretical issue: if people can have the virus and be asymptomatic, how does one determine the true positive rate?

38

An Example for a Very Contagious Disease with No Good Treatment

	Positive Test	Negative Test
Disease Present	180	0
Disease Absent	100	220

- Sensitivity = 180 / 180 = 1.0
- Specificity = 220 / 320 = 0.69
- Positive predictive value = 180 / 280 = 0.64
- Negative predictive value = 220 / 220 = 1.0

39

Summary

- Useful to know specificity/sensitivity/etc. in order to understand the purpose of a given test
- Need to establish these values when developing new tests and procedures

40

Controversy: Whole-Body CAT Scan

- Whole-body CAT scan for healthy people to pick up any possible disease
- Problems:
  - CAT scan sensitive, but not specific
  - Findings often equivocal, require lots of follow up testing to confirm
  - Causes lots of unnecessary worry
  - Stress on already-burdened health care system

41

Type I Error

- Detecting a difference between two groups when none really exists
  - False positive
- *P* value of a calculation
  - Expresses chance of a type I error
  - Generally, $P < 0.05$ used to demonstrate statistical significance

42

A Word About P-Values

- P < 0.05 an arbitrary threshold
  - Means 95% certain result is not just a statistical fluke
  - Is this "good enough?"
- This means that 1 of 20 "statistically significant" results is result of random error
  - What happens when dozens of medical journals become hundreds?
- Sterne, Smith. Sifting the evidence—what's wrong with significance tests? *British Medical Journal* 322:226-231, 2001.
  - http://bmj.bmjjournals.com

43

Type II Error

- Not detecting a difference between two groups that does, in fact, exist
  - False negative
- Tends to occur with small sample sizes
- *Statistical power*
  - Power of 80% usually used
  - Requires knowledge of your test population
    - How variable measurements will be
    - Good guessing

44

Sources of Error

- Bias
  - Measurement always erroneously high or low
  - Human tendencies
  - Other systemic effects
    - Calibration of measuring device
- Random error
  - Difficulty in reproducing test conditions
  - Noisy measurement tools
  - Noisy measures
    - Biological systems tend to have large tolerances

45

Comparing Sets of Data

- Independent data
- Dependent (paired) data
- *Normally* distributed!
  - "If there's any justice in the world, everything is Gaussian."
    - -- F. Fontaine, many times
- Comparing means

46

Independent t-test

- Comparing two different things
  - Independent
- Comparing the means of two different populations
- Standard deviation/standard error give an idea of how spread about the mean data are
  - Variance

47

Paired t-test

- Comparing data that are paired
  - e.g., leg strength before and after rehab in same person
  - Data are not statistically independent
- Mean, std dev/err of groups are *not* interesting for this comparison
- Interested in the mean/std dev of the difference between the pairs

48

t-test Examples

- Compare pain scores (scale of 1-100) in subjects
  - Comparison of people with arthritis to people without arthritis
    - Independent t-test
  - Comparison of people with arthritis to *the same people before they had arthritis*
    - Paired t-test

49

Comparing More Than Two Things

- Example: subjects with ACL reconstruction (anterior cruciate ligament, in the knee)
  - Involved (injured) side vs. uninvolved side
  - Test knee flexion strength at different speeds: 30, 60, 90, 120 deg/sec
  - Two sides, four speeds = lots of t-tests (paired or unpaired in this case?)
- ANOVA (repeated measures)

50

ANOVA

- Analysis of variance
- Way of performing many comparisons
- *Main effects*
- *Interactions*

51

ANOVA Example

- Previous example:
  - 2 x 4 ANOVA
    - 2: uninvolved, involved leg
    - 4: different speeds
  - Main effects [possible explanations]:
    - Leg
      - ACL leg weaker (?)
    - Speed
      - Lower strength at higher speed (regardless of which leg) (?)
  - Interaction [possible explanations]:
    - Between leg and speed
      - Involved leg weaker at some speeds but not others (?)

52

Types of ANOVA

- Repeated measures
  - Previous example
  - All data being compared are dependent
- Mixed model
  - Some comparisons independent
  - Example:
    - Test strength prior to and after some intervention
    - Also compare between genders
    - 2 x 2 mixed model ANOVA

53

More ANOVA

- Use *post-hoc* t-tests to illustrate differences
  - Mean and std dev of all tests with involved leg vs uninvolved leg
  - Use P-value for t-tests
  - This is unneeded if there are only two levels to the variable: e.g., pre vs post, since there is only one comparison to make
- Beware multiple comparisons!

54

Multiple Comparisons

- Comparing variables all day increases chances of finding a significant result
- Correct for multiple comparisons in *post-hoc* t-tests
  - Strict: *Bonferroni* correction
  - Multiply P-value by number of comparisons made
    - If three comparisons, need $P < 0.017$ for significance
- ***Some studies do this, some do not!***

55

Another ANOVA Issue

- Sphericity
  - Variance of effect (SD of difference) is different between different levels of a variable
  - There exist corrections for this
    - e.g., Greenhouse-Geisser
  - ***Very few studies do this!***

56

Linear Regression and Correlation

- Relationship between two variables is linear
  - Let's hope it is
  - We are not good at explaining non-linear things
    - Exponential isn't so bad
- Correlation coefficient explains how much variance in one variable is accounted for by another variable

57

Correlation Coefficient

- Pearson correlation
- Pearson *R*
- *R*
- Correlation between two variables:
  - The value R^2 tells how much one variable explains variance in another
    - Correlations may be statistically significant but not particularly useful

58

Multiple Regression

- Model some data as a linear combination of several other variables
  - E.g., body fat from skinfold measurements
- Can be very useful
- Can be horribly misapplied

59

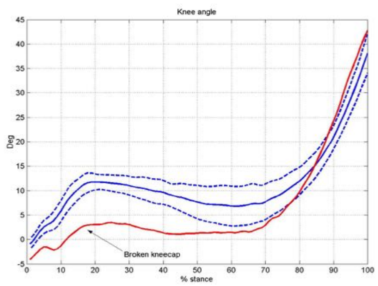
Limits of Agreement

- Where can data be expected to live?
- 68% within 1 st dev of mean
- 95% within 2 st devs of mean

60

Limits of Agreement Example

+ Knee angle during stance phase of gait



61

Difficult Things to Measure

+ Pain
+ Outcome after a treatment/procedure
+ Patient/subject satisfaction
  + Mood ring?

62

Outcome Measurements

+ Questionnaire
  + Score for each question
  + Must be validated
+ "How do you feel?" scale of 1 to 10
  + Not continuous
+ Visual analog scale
  + Mark pain/satisfaction on [10 cm] line
  + Measure distance from start of line to mark

63

My Data Aren't Normally Distributed!!!

- Few samples
  - Difficult to create a bell-shaped curve with five subjects
- Non-continuous data
- Categorical data
  - "yes", "no", "sometimes"
- Use non-parametric statistics

64

Non-Parametric Statistics

- Non-parametric versions of parametric tests
  - Independent t-test: Mann-Whitney test
  - Paired t-test: Wilcoxon test
  - Etc...
- Chi-square test
  - More yes replies in group 1 than group 2

65

A Word About Statistical Significance

- Not necessarily clinically significant!
- Is effect smaller than something that can be measured reliably?
  - e.g., Group 1 has 0.5° more knee flexion than group 2

66

Important Medical Research Trivia

- Do not compare right vs left
  - Why?
- What part of the body is referred to as the "leg"?

67

Reading and Interpreting Papers

- Read with a critical eye!
  - Appropriate prior research cited in introduction?
  - Used correct tools?
    - Measuring devices
    - Stats
  - Used tools properly?
    - Electromagnetic tracker in room with metal studs in walls
  - Correct interpretation of statistics

68

Reading and Interpreting Papers

- Beware the abstract!
  - Sales pitch for the paper
  - Often good to read *last* rather than first
    - Did authors, in fact, demonstrate what they claim in abstract?

69

Sample Study

- *Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial*
  - Leonard Leibovici
  - British Medical Journal 2001; 323:1450-1451
- http://bmj.bmjjournals.com/cgi/content/full/323/7327/1450

70

Sample Study

- Population
  - All adult patients admitted to a university hospital with bloodstream infection from 1990-1996
  - 3393 patients
- Randomized into two groups
  - 1691 in intervention group
  - 1702 in control group

71

Sample Study

- Intervention
  - List of first names of patients in intervention group given to a person
  - Person said short prayer for the well-being and full recovery of this group as a whole
- Variables compared
  - Number of deaths in hospital (mortality)
  - Length of hospital stay
  - Duration of fever

72

Sample Study

- Results
  - Mortality not different between groups
    - 28.1% vs 30.2%, P = 0.4
  - Length of stay in hospital shorter in intervention group
    - P = 0.01
  - Duration of fever shorter in intervention group
    - P = 0.04
- Conclusion
  - "...This intervention is cost effective, probably has no adverse effects, and should be considered for clinical practice..."

73

Sample Study Strengths and Weaknesses

- Prayer said long after illness
  - Is this useful?
  - Does it affect the way patient will be treated in hospital if it is known that s/he will be prayed for afterwards?
    - Perhaps it doesn't matter as long as the prayer is said?
- Why not pray when they come into hospital?
  - Prospective study
  - But, presumably, a supreme being is not limited by causality
- What flavor of prayer?
  - Performed in Israel
  - Christians/Muslims/Buddhists/Druids/Pastafarians et al. are wrong?

74

Sample Study Conclusion

- Proof of existence of divine?
- Proof of existence of Type I error?
- Note: published in December issue
  - BMJ often publishes *interesting/ironic* papers in the "Christmas" or April issues

75

Finding Relevant Literature

- Many databases available
  - Some pay
  - Some free
- PubMed:
  - https://www.ncbi.nlm.nih.gov/PubMed/
  - Free access to MEDLINE and several other databases
- Google Scholar
  - https://scholar.google.com
  - Often leads to PDFs the author may have uploaded

76