Jumping into Statistics: Introduction to Study Design and Statistical Analysis for Medical Research Using JMP Pro Statistical Software

FALL 2022/SPRING 2023

DR. TERRIE VASILOPOULOS, DR. CYNDI GARVAN,
& DR. PENNY REYNOLDS

Meet the Instructors

TERRIE VASILOPOULOS, PHD

Research Associate Professor in Anesthesiology and Orthopaedics and Rehabilitation

CYNTHIA GARVAN, PHD

Research Professor in Anesthesiology

PENNY REYNOLDS, PHD

Research Assistant Professor in Anesthesiology and Veterinary Medicine

Course Objectives

- Review fundamentals of study design and research methodology
- Understand how to choose best statistical test for your research question
- Practice basic statistical analysis use JMP Pro Software

Course Topics

- Life Cycle of Research and Asking a Good Research Question
- Choosing the Right Study Design for Your Research
- Clinical Trial Design
- Populations, Samples, and Hypothesis Testing in Medical Research
- Introduction to Data Types
- Best Practices in Data Collection and Database Management: Getting Started with SAS JMP Pro
- Summarizing and Visualizing Data
- Statistical Methods and How to Choose Them
- Risk Assessment Methods
- Introduction to Regression and Correlation
- Time-to-Event (Survival) Analysis
- Methods for Clinical Diagnostic Testing

Life Cycle of Research and Asking a Good Research Question

9/7/2022

Learning Objectives

Participants will be able to:

1) Understand research process and its iterative nature

2) Identify necessary components of a successful research questions

3) Create an actionable research question

Life Cycle of Research and the Scientific Method

Why is this topic important?

- Understand how new knowledge (aka evidence) is created
- Help you to see how the life cycle of research is connected to the scientific method
- Encourage you to develop as a scientist

The scientific method is the foundation for conducting sound research.

“THE ESSENCE of scientific inquiry is to discover truth: Ask questions, seek evidence, develop hypotheses, conduct experiments and validate findings.”

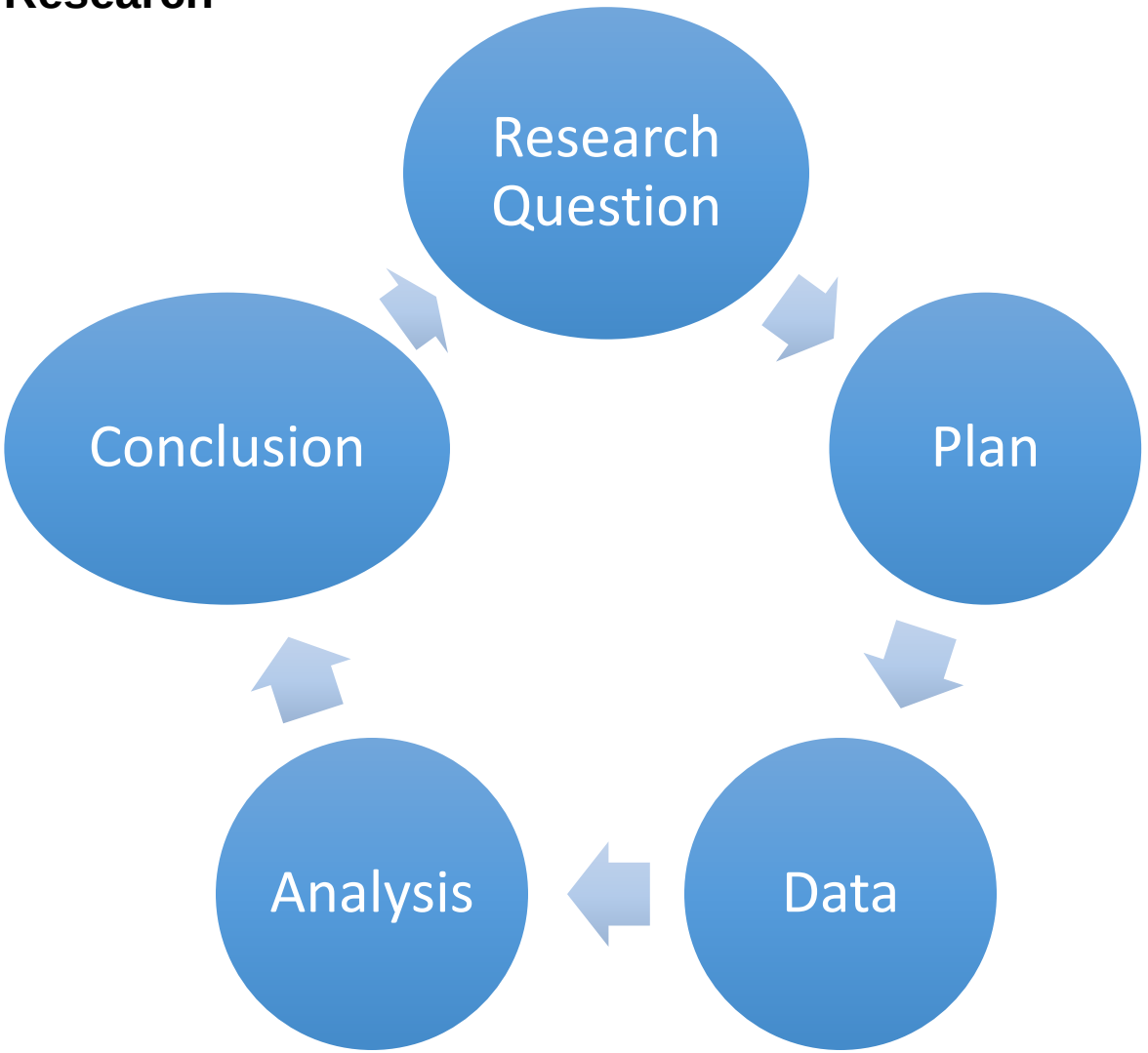
- Washington Post Editorial Board article on discovering origins of COVID-19

Easier said than done.

Simplexity

The Life Cycle of Research

We discover truth with rigorous and reproducible research.



Wild CJ, Pfannkuch M. Statistical thinking in empirical enquiry. International statistical review. 1999 Dec;67(3):223-48.

Research starts with identifying a problem or asking a research question or formulating a hypothesis.

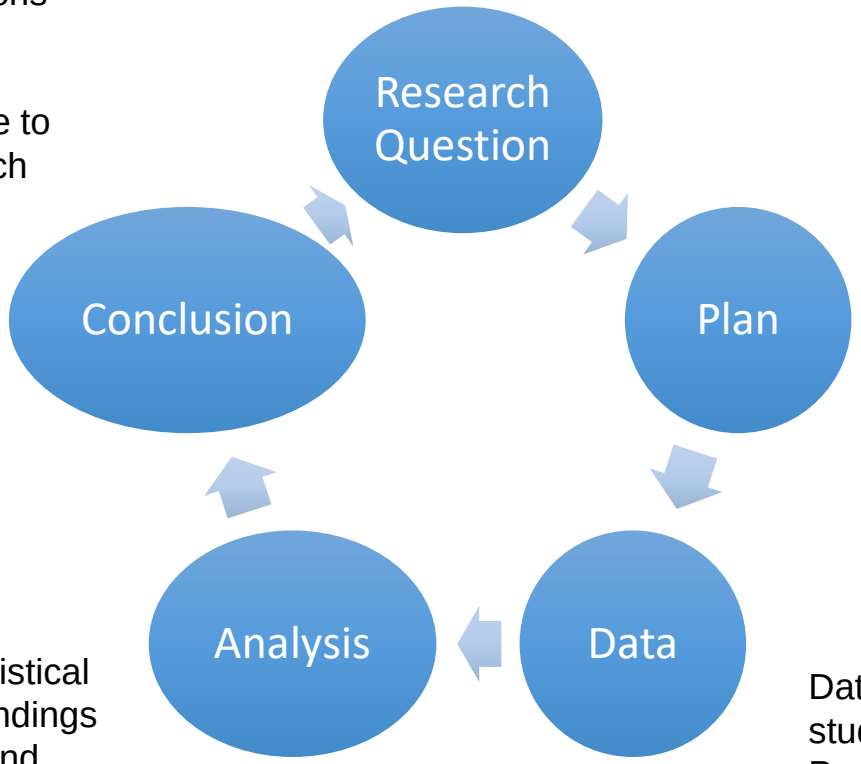
The Planning Phase includes:

- finding out what is already known about the research topic
- formulating a hypothesis
- selecting a study design
- choosing instruments to measure variables of interest
- developing a statistical analysis plan and conducting a power analysis to ensure adequate sample size

Data are collected as the study is implemented. Performing data checks to ensure data quality is an important part of the Data Phase.

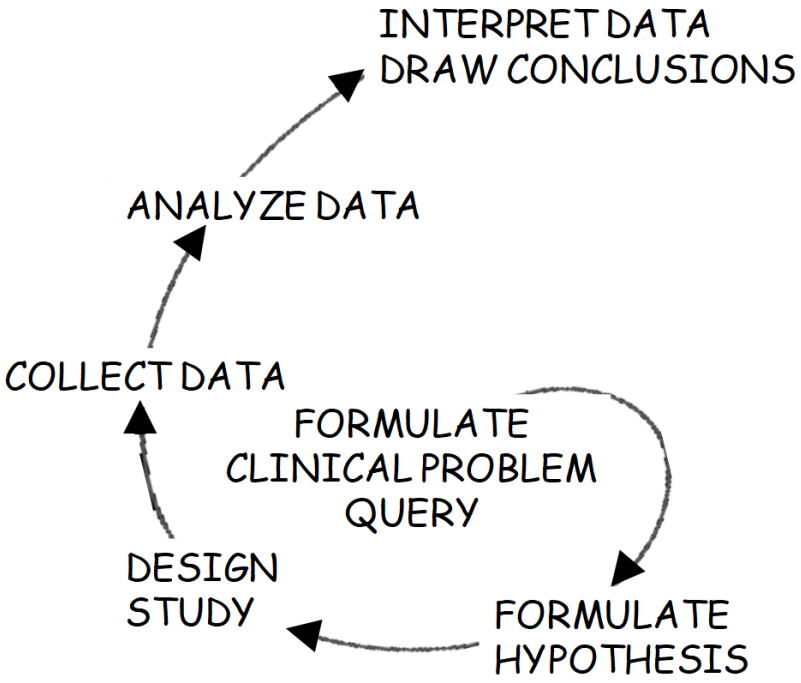
In the Analysis Phase, statistical methods are applied and findings are summarized in tables and graphs.

1. Based on findings from the Analysis Phase, conclusions are drawn.
2. Results of statistical analyses form of evidence to address the study research questions.
3. Study findings are disseminated as manuscripts.
4. Often new questions are generated as a study is completed and the cycle begins anew.



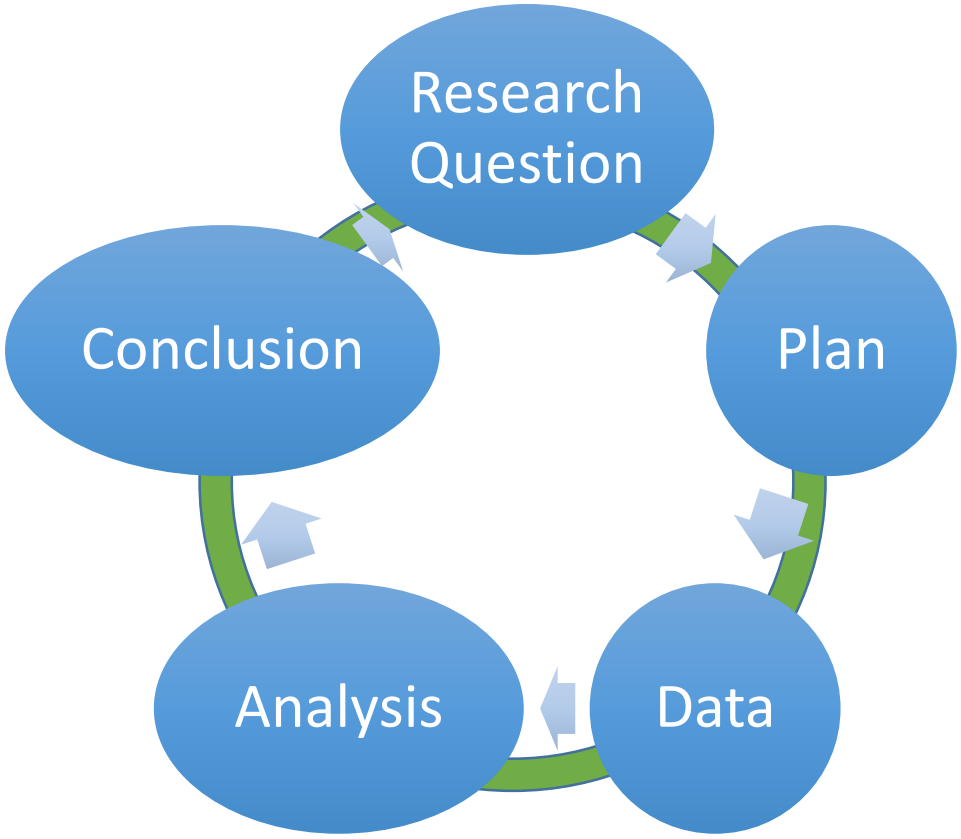
The Life Cycle of Research

The Scientific Method



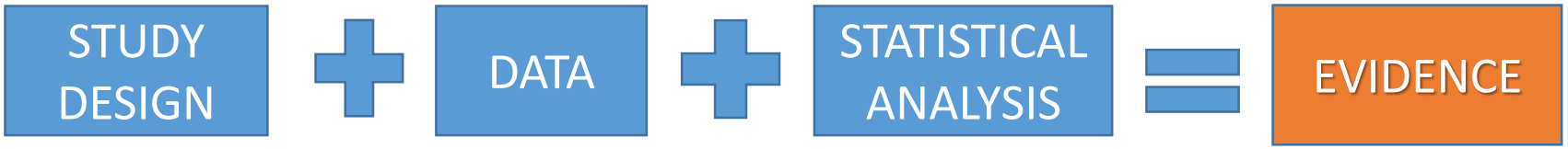
Medical practice is grounded in evidence based medicine (EBM) which is based on evidence generated from the application of the scientific method.

The Scientific Method



The scientific method is the agreed upon framework established by investigators to align the research question, hypothesis, study design, measurement, and data analysis to produce evidence to acquire new knowledge or for correcting and integrating previous knowledge.

The Scientific Method



- Data by itself is information which may not be meaningful.
- Statistics is the logic which underlies the Scientific Method.
- The scientific method combines study design, data, and statistical analysis (logic) to generate knowledge (evidence).

A cautionary tale: Information is not knowledge

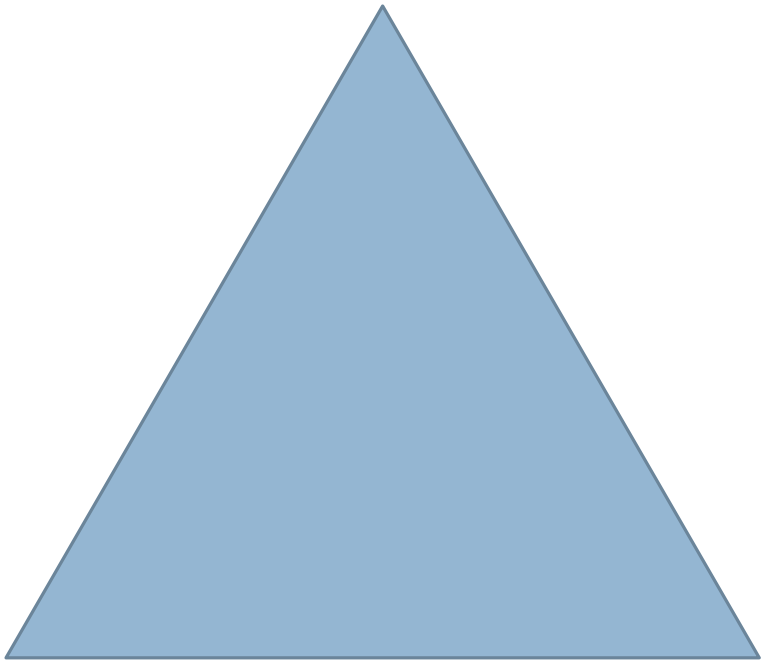
George Gallup (1901 – 1984) is often credited as the developer of public polling. In 1936, his organization, the American Institute of Public Opinion (Gallup Poll), achieved national recognition by correctly predicting, from the replies of only 50,000 respondents, that Franklin Roosevelt would defeat Alf Landon in the U.S. Presidential election. This was in direct contradiction to the widely-respected *Literary Digest* magazine whose poll based on over two million returned questionnaires predicted that Landon would be the winner. Not only did Gallup get the election right, he correctly predicted the results of the *Literary Digest* poll.

Although the sample size of the *Literary Digest* poll was over 2 million, the scientific method was not applied to predict the election. Those who participated in the *Literary Digest* poll were wealthier than the typical American voter. Big data is not enough to produce quality evidence - a truth known since 1936.

What does logic have to do with it?

Deductive Logic

Theory

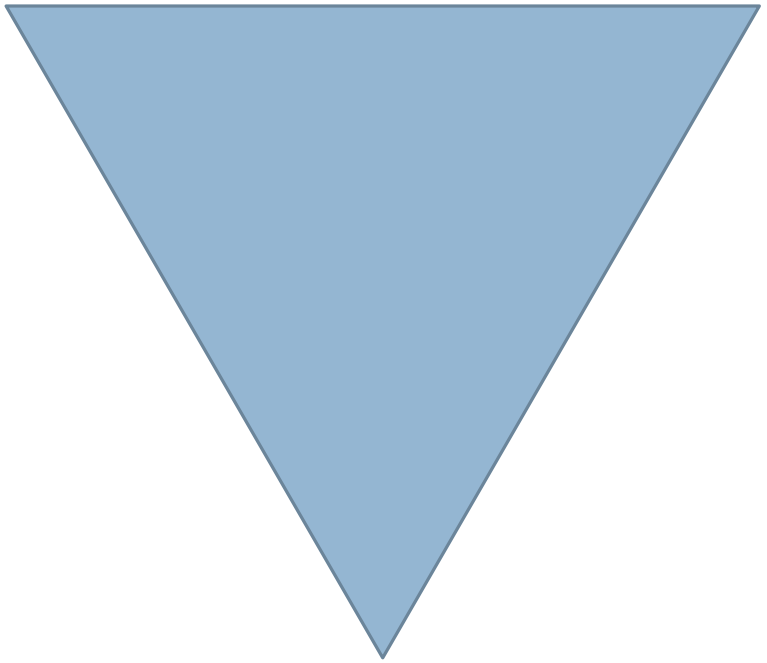


Facts

Top down reasoning

Inductive Logic

Facts (Observations)

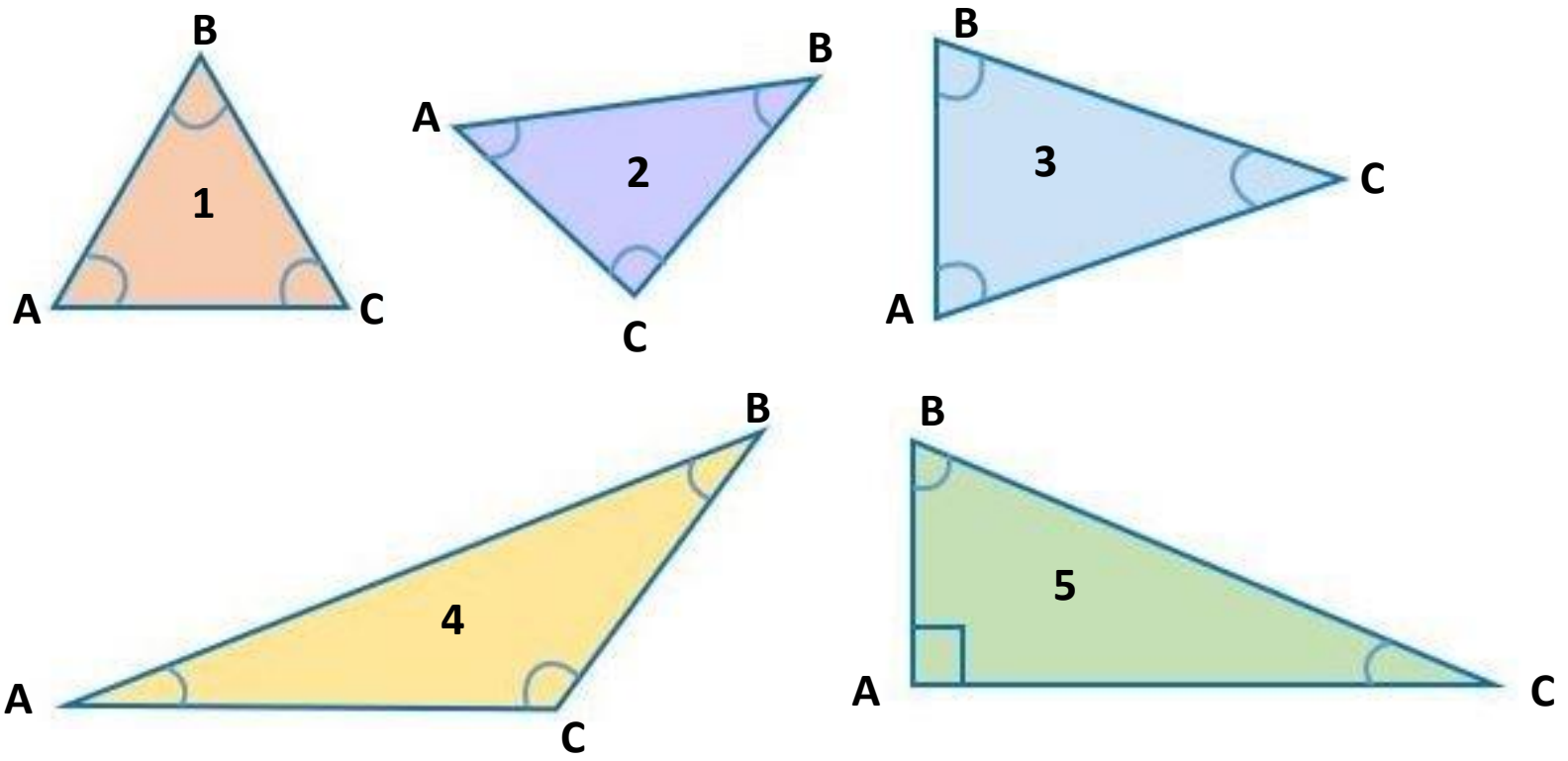


Theory

Bottom up reasoning

Example of Inductive Reasoning

Suppose that each student in an elementary geometry class is given a set of 5 plastic triangles and a protractor. The instructions direct the students to find and record the measures of the angles and then to find the sum of the measures for each triangle.



Triangle	Measure A	Measure B	Measure C	Sum
1	60	59	60	**179**
2	56	58	67	**181**
3	80	50	50	**180**
4	36	40	106	**182**
5	45	45	90	**180**

The object of the lesson is to make it plausible that the sum of the measures of the angles of a triangle is 180° - If the students reason that this is a correct conclusion on the basis of the measurements they have made, this is an example of inductive reasoning.

Example of Deductive Reasoning

Preliminary definition: An integer Z is an even number if there is an integer W such that Z = 2W.

Hypothesis: X and Y are even numbers
Conclusion: X + Y is an even number

Proof:
Since X is even, there is an integer U such that X=2U
Since Y is even, there is an integer V such that Y=2V

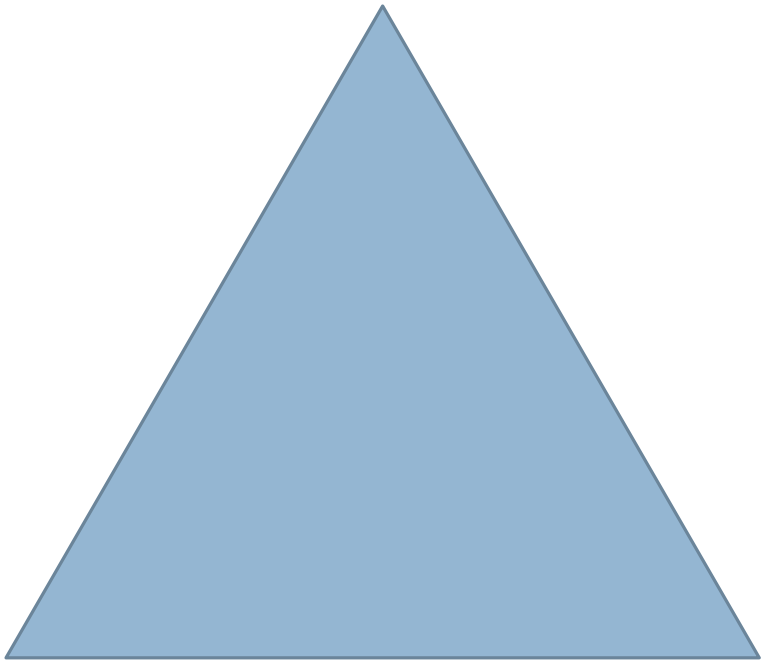
Then X + Y = 2U + 2V
= 2(U + V) = 2(some integer)

Which implies that X + Y is an even integer.

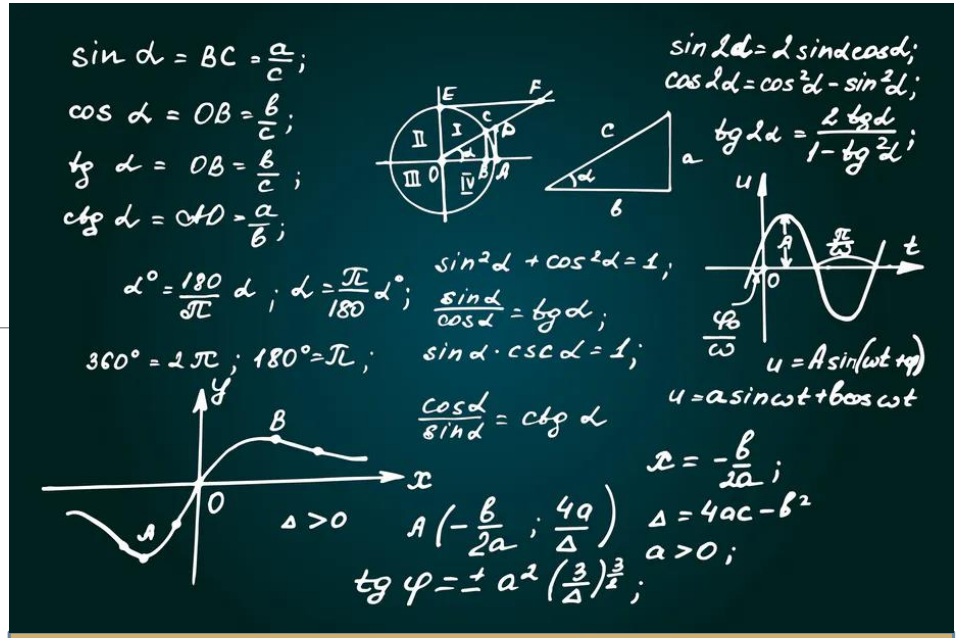
- QED (Quod Erat Demonstrandum, a fancy way to show off you just logically proved something)

Deductive Logic

Theory



Facts

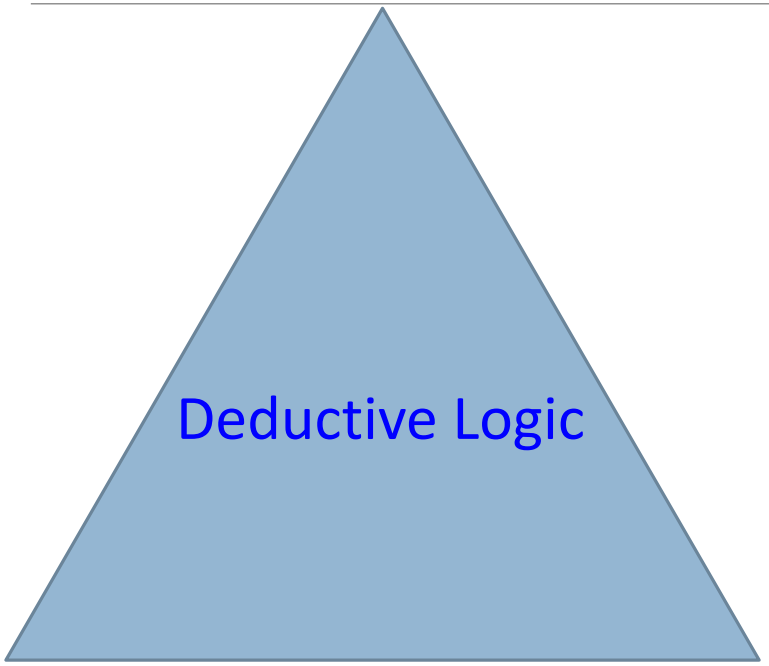


Mathematics and Statistics are sciences where the majority of knowledge is discovered using deductive reasoning.

Medical Knowledge

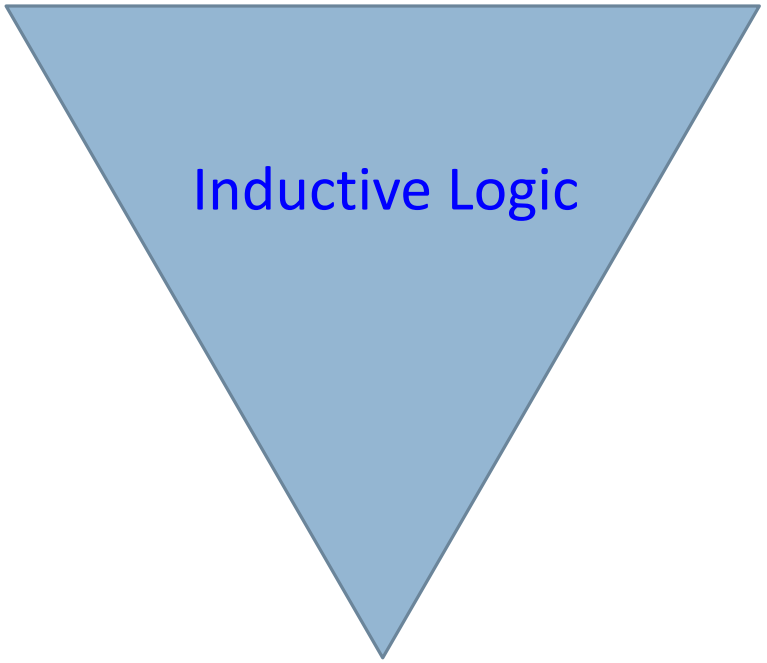
Boyles Law

As volume of a gas increases, its pressure decreases.



As a diver ascends to a depth with less water pressure, nitrogen gas expands. This expanding nitrogen gas can form tiny bubbles in his blood and tissue and cause decompression sickness.

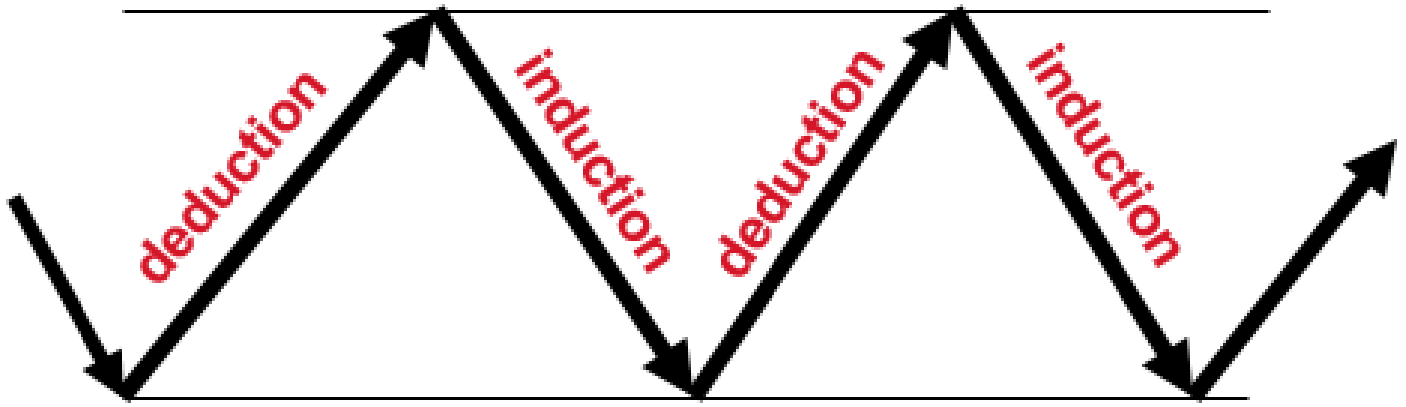
Vaccine RCT



Pfizer & Moderna vaccines are 95% effective

Science's Iterative Learning Process

Data (facts, observations, phenomena)



THEORY

Summary Tip

The first step to becoming a Clinician Researcher is to become a Clinician Scientist.

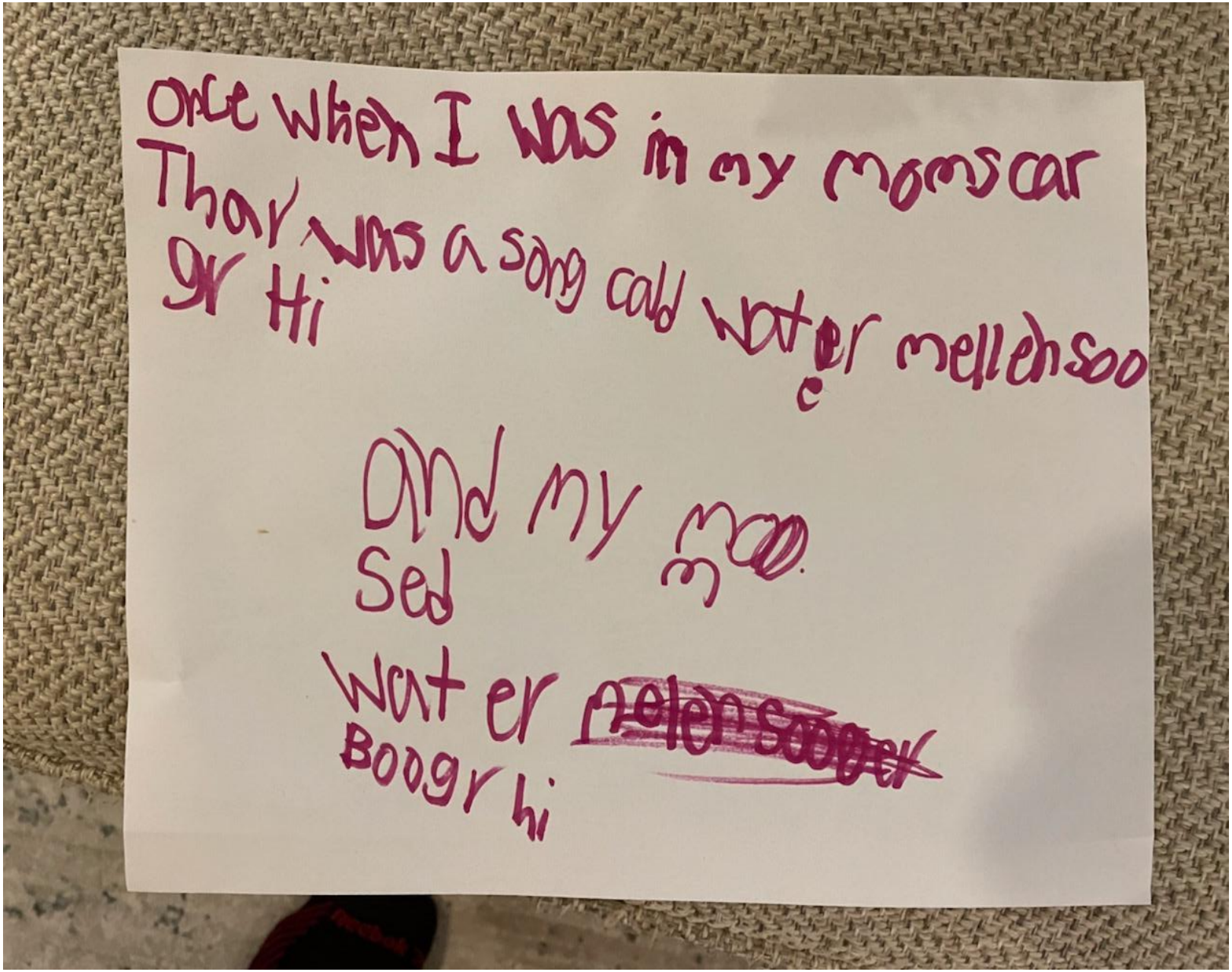
How to Ask a Good Research Question

Why is it important to ask a good research question?

- Identify gaps in knowledge
- Focus your research
- Develop testable objectives and hypotheses

A good research question should essentially be a guide to your project

My granddaughter Claire (7 yo)


Once when I was in my moms car
Thar was a song cald water mellen soo
9r Hi
and my mom
Sed
Wat er
Boogr hi

To ask a good research question – just ask

Don’t worry about the formulation. Focus on what you want to know. Examples:

1) What causes an ulcer?

2) Why do some premature babies become blind?

3) Why is it taking longer for me to do my crossword puzzle 3 months post op?

What causes an ulcer?

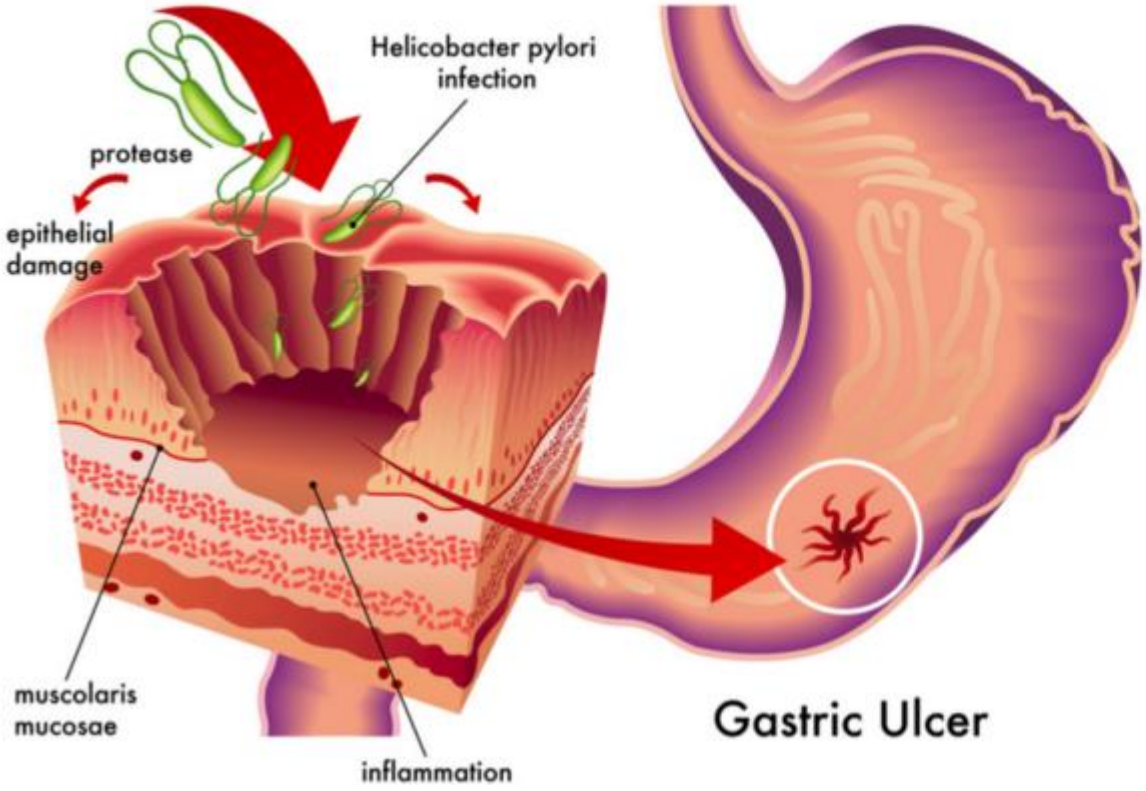
Up until the discovery of *H. pylori* in 1982, clinicians attributed peptic ulcers to lifestyle choices. This could have included consuming a diet rich in spicy foods and an inability to properly manage emotional and personal stress. A paper published in 1967 even reported that peptic ulcers appeared in families with dominant and obsessional mothers.

What causes an ulcer?

Nearly 40 years ago, Australian physicians Barry Marshall and Robin Warren claimed that stomach ulcers were caused by a bacteria called *Helicobacter pylori* and not by excessive acidity in the stomach.

To test his theory, Dr. Marshall ingested the bacteria. He then documented both the formation of his stomach ulcers and their cure following treatment with a combination of antibiotics and stomach-acid-neutralizing medicines.

What causes an ulcer?



Why do some premature babies become blind?

On February 14, 1941, Dr. Stewart Clifford, a pediatrician in Boston, made a house call to check on one of his patients, the three-month-old daughter of a young rabbi. The girl had been born several weeks premature, weighing just four pounds at birth, but had been doing well in the months since her birth. Unfortunately, something now seemed wrong. There was a grayness in the pupils of the girl's eyes—and she appeared to have lost her ability to see.

Just days later, another of Clifford's patients, this one a seven-month-old baby, was discovered to have the same symptoms. That baby had also gone blind. And just as with the prior case, the child had been born prematurely.

By 1942 several other cases of premature babies going blind, with the same symptoms—the appearance of gray masses inside the babies' eyes—were reported in the Boston area.

Clifford contacted Dr. Theodore Terry, professor of opthalmology at Harvard Medical School, and asked him to look into the mysterious cases. Terry studied five of the cases, and wrote an article about the condition in American Journal of Ophthalmology. When eye doctors around the country saw the article, similar stories were reported outside the Boston area.

Most ophthalmologists studying retrolental fibrofasia (RLF) were convinced that the condition was related to the fact that all its victims had been born prematurely. That was understandable.

But Terry wasn't convinced: if this was just another premature birth complication, why hadn't it been seen before?

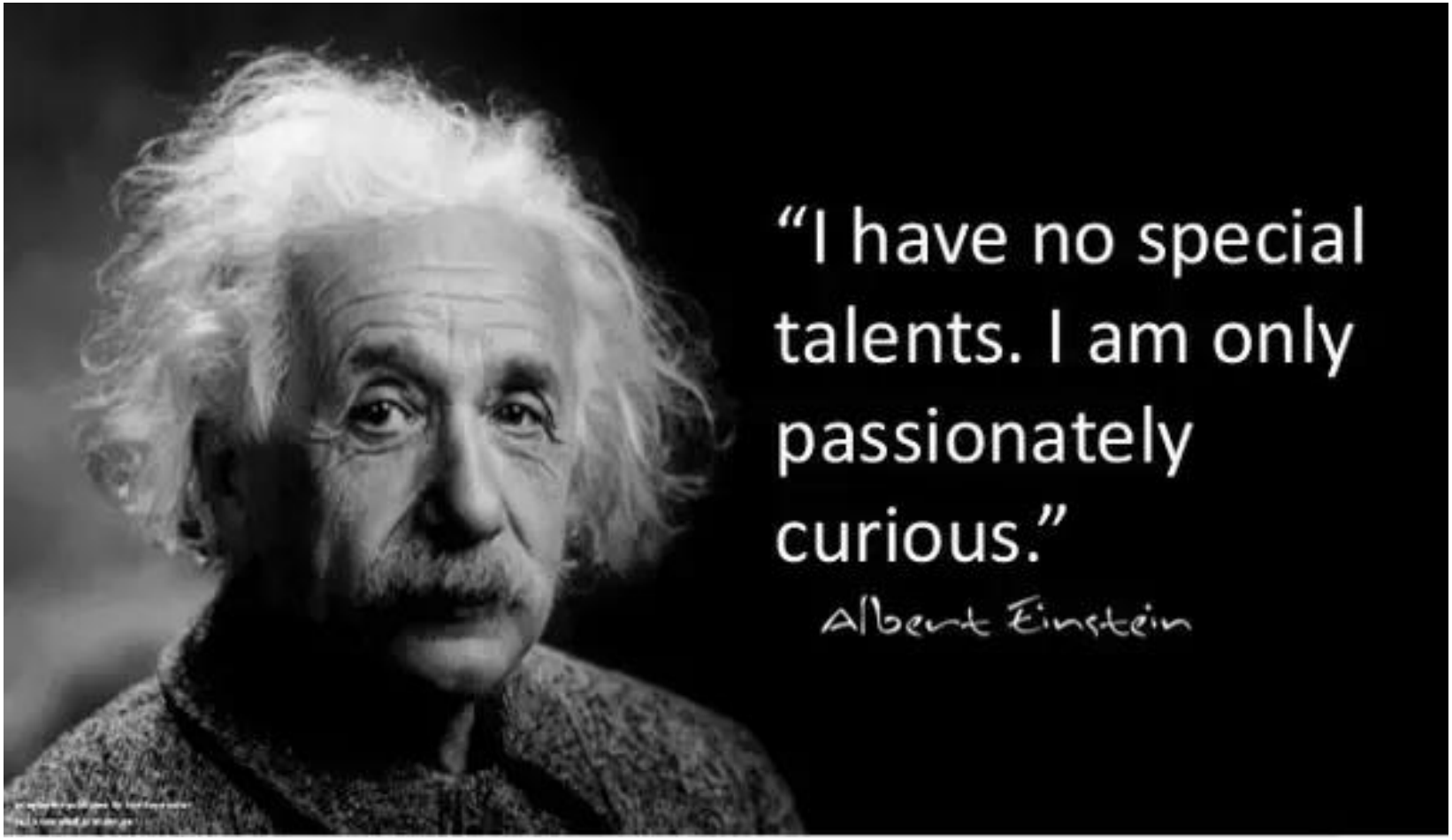
And one of the most puzzling aspects about RLF: it was occurring almost exclusively in modern, developed nations where the care of premature babies had improved and the infant mortality rate for preemies had fallen dramatically over the previous decades.

In 1949 doctors started examining a treatment for premature babies that prior studies had overlooked: incubators. Incubation devices had been used to keep delicate babies warm since the late 1800s, but in the 1930s, a new kind was developed—airtight incubators that could maintain abnormally high air-oxygen levels.

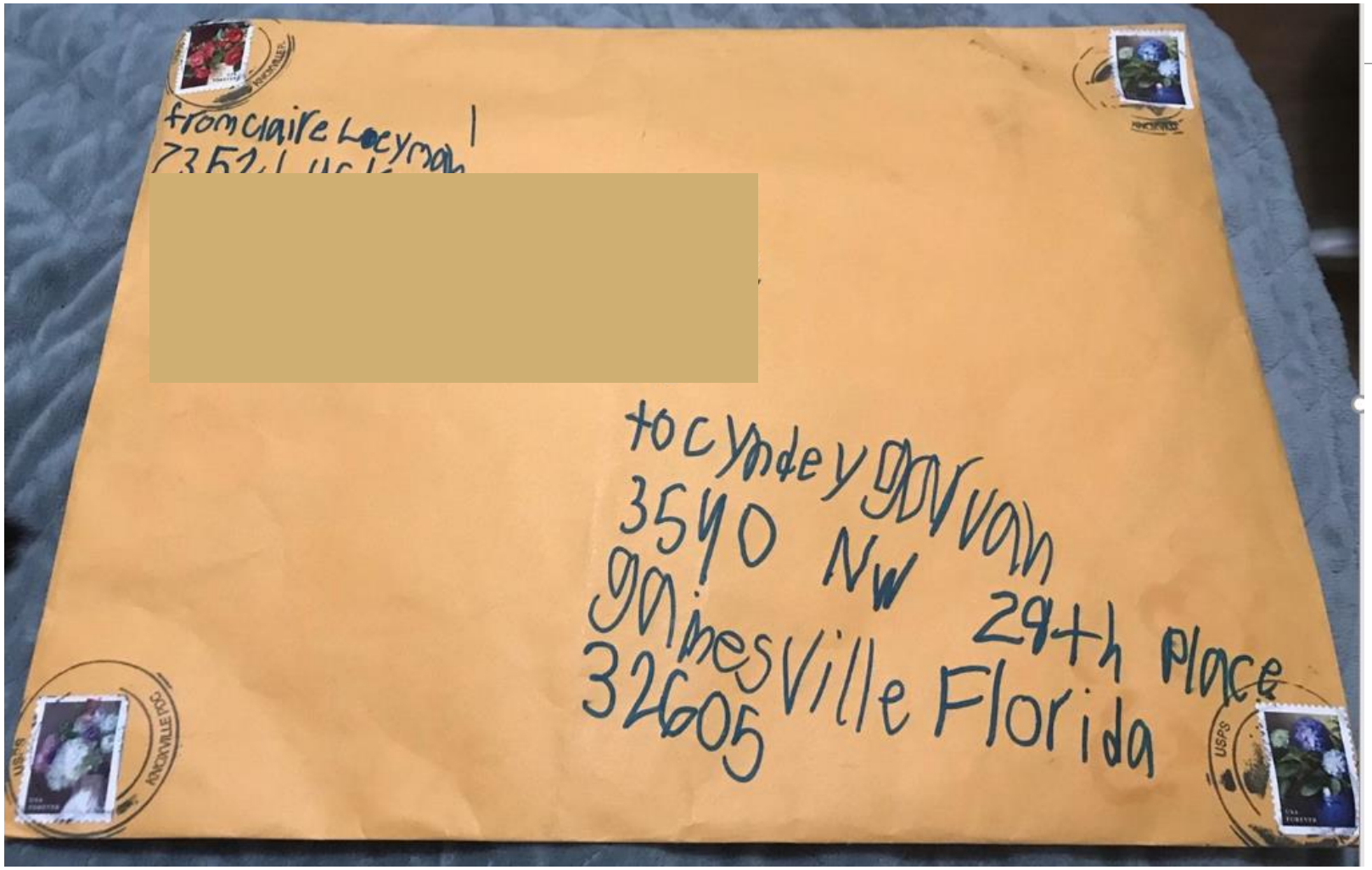
The treatment had been credited with lowering the mortality rate…but had there been an unknown price for that lower rate? By the early 1950s, a lot of doctors thought that might be the case.

Why is it taking longer for me to do my crossword puzzle 3 months post op?

Investigation of post operative cognitive dysfunction began when a physician noticed he could not complete his crossword puzzle as quickly as he could before his surgery.


"I have no special talents. I am only passionately curious."
Albert Einstein

After asking your question you can worry about formatting.



How to formulate a good research question?

- **FINER** → overarching framework
  - Feasible, Interesting, Novel, Ethical, Relevant
- **PICO(T)** → operationalize (testable)
  - Population, Intervention, Comparison, Outcome, (Timing)

Complimentary methods

Feasible	• Sample size, expertise, time and money • Manageable Scope
Interesting	• Intrigues investigators, peers, community
Novel	• Confirms, refutes, or extends previous findings
Ethical	• Likely approved by IRB/IACUC
Relevant	• To scientific knowledge, clinical/health policy, future research

Population	• Who are your patients? • Gender, age, ethnicity, medical history
Intervention	• Studying effect of what? • Drug, procedure, behavior, condition, risk factor exposure
Comparison	• What is comparison/control group? • Placebo, no treatment, standard of care, no exposure
Outcome	• Measure that may be affected by intervention? • Disease occurrence, progression, improvement, mortality, efficacy
Timing	• Over what time frame will study take place?

NIH Specific Aims

The Aims

In this section, you will describe briefly each of the aims you will use to test your hypothesis. Ideally, the aims should be related, but not dependent, upon each other. If you do this, the failure of one aim (or an unexpected result from one aim) does not negatively influence any other aim or prevent the completion of the other aims.

Within 2-4 sentences each, you should describe the experimental approach and how each aim will help answer your larger hypothesis. A typical NIH R01 grant will have between 2 and 4 Aims. Plan to describe each aim in a separate paragraph. Additionally, these tips may help you to formulate your aims sections:

- Give your aim an active title that clearly states the objective in relationship to the hypothesis.
- Include a brief summary of the experimental approach and anticipated outcomes for each aim.
- If you have room, you may wish to include a sub-hypothesis (the small portion of the overall hypothesis) and a small description of the pay-off of each aim. Including these is helpful to creating the impression that each aim is valuable, testable, and independent of the others.
- To make it easier for the reviewers to clearly read and understand each aim, it is often helpful to use headings and/or bullets to delineate each specific aim.

Example Specific Aims

Aim 1 will establish an innovative mouse model for HTLV-1 Tax tumorigenesis. Targeting vectors containing silenced wild-type or mutant Tax genes will be knocked in to the Rosa26 locus of C57BL/6 mice. These mice will then be crossed with homozygous Lck-CRE mice, thereby excising the stop cassette and generating mice that express wild-type or mutant Tax proteins specifically in T cells.

Aim 2 will examine the effect of mutations that disable specific biological functions of Tax on Tax-mediated tumorigenesis. Tax can bind to and regulate the activity of members of the SRF, CREB, NF-kB and PBM protein families, each of which has been implicated in oncogenesis. Mice established in Aim 1 will allow us to compare for the first time the tumorigenic potential of wild-type and mutant Tax proteins in an effort to identify pathways that are required for Tax tumorigenesis.

Color Key: Aim Title | Experimental Strategy | Outcome or Impact

Figure 3. The Aims Section. Sections of the paragraph have been color coded to highlight each critical component. Note the active voice in the titles of each aim and the use of boldface text to highlight the titles.

Formulation important for:

1. Designing the study
2. Translating research question into testable hypothesis

Ideally…

Summary Tips

- Perform thorough literature review
- Examine current trends and technological advances in topic
- Seek input from experts, colleagues, mentors, and collaborators to refine ideas
- Use FINER to develop research question
- Focus question with PICOT to develop testable question
- Develop research hypothesis (later lecture)
- Develop clear, well-defined primary and secondary (if needed) aims
- Before beginning study, re-review to ensure question and objectives are answerable, feasible, and relevant

QUESTIONS?

References

Gauch Jr HG. Scientific method in practice. Cambridge University Press; 2003.

Motulsky H. Intuitive biostatistics: a nonmathematical guide to statistical thinking. Oxford University Press, USA; 2014.

O'Neil C. Weapons of math destruction: How big data increases inequality and threatens democracy. Crown; 2016 Sep 6.

Ringenberg LA, Presser RS. Geometry. Benziger-Wiley Mathematics Program series. Beverly Hills, Calif.: Benziger. 1971.

Wachter RM. The digital doctor: hope, hype, and harm at the dawn of medicine's computer age. New York: McGraw-Hill Education; 2015 Apr 1.

Wild CJ, Pfannkuch M. Statistical thinking in empirical enquiry. International statistical review. 1999 Dec;67(3):223-48.

https://www.todayifoundout.com/index.php/2018/05/curious-case-blind-baby-epidemic/

https://www.biosciencewriters.com/NIH-Grant-Applications-The-Anatomy-of-a-Specific-Aims-Page.aspx