Conclusion: Studies and Statistics in Society

MATH 348, Vassar College, Spring 2024

Prof. Lee Kennedy-Shaffer

April 22, 2024

Davies, *The Guardian*, 1/19/2017

Outline

Research Study Process

- Refine the research question
  1. Identify question of interest
  2. Review previous literature
  3. Identify possible data sources
- Choose a study design
  4. Determine statistical question
  5. Assess statistical and ethical issues
  6. Determine design and analysis plan
- Conduct the study
  7. Implement study and collect data
  8. Analyze data according to plan
  9. Conduct additional exploratory analyses
- Report out the study
  10. Report results and conclusions
  11. Identify limitations
  12. Suggest future research avenues

Statistical Principles and Design Choices

Design Solutions: Survey Sampling

Questions

- Description of Current World
- Association
- Differences and Disparities

Design Solutions: Survey Sampling

Questions

- Description of Current World
- Association
- Differences and Disparities

Challenges

- Biases: Non-response, Measurement Error, Weighting and Stratifying
- Generalizability: Survey Population
- Power and Precision: Variability, Sample Size

Design Solutions: Randomized Trials

Questions

- Causal Effects and Relationships
- Predictions About Future

Design Solutions: Randomized Trials

Questions

- Causal Effects and Relationships
- Predictions About Future

Challenges

- Ethics and Feasibility
- Biases: Dropout, Measurement Error
- Power and Precision: Variability, Sample Size
- Generalizability: Real-World Effects?

Design Solutions: Observational Studies

Questions

- Associations and Descriptions
- Causal Effects in the Real World

Design Solutions: Observational Studies

Questions

- Associations and Descriptions
- Causal Effects in the Real World

Challenges

- Biases: Confounding, Selection Bias, Recall Bias
- Power and Precision: Variability, Sample Size
- Generalizability: Specifically-Chosen Populations

No Free Lunch

Throughout we have seen tradeoffs:

No Free Lunch

Throughout we have seen tradeoffs:

- Bias vs. Variance
- Ethics/feasibility vs. Bias
- Internal validity vs. External validity/Generalizability
- Cost and Convenience vs. Validity and Precision

Lokesh, *Toward Data Science*, 7/9/2020

No Free Lunch

Throughout we have seen tradeoffs:

- Bias vs. Variance
- Ethics/feasibility vs. Bias
- Internal validity vs. External validity/Generalizability
- Cost and Convenience vs. Validity and Precision

Lokesh, *Toward Data Science*, 7/9/2020

There Is No Perfect Design

So methodologists need to find ways to **minimize tradeoffs**. And those designing studies need to be able to **understand**, **quantify/estimate**, and **select** among the options.

Outline

So Much!

Avenues for study design:

- Designs for new data science analysis methods
- Causal inference methods: types and theories of causality
- Meta-analysis: combining study results
- Communication and interpretation of study results
- Field-specific designs and challenges

Lab Sciences

- Sources of bias and variation
- What can be eliminated by averaging?

Economics and Public Policy

- Quasi-experimental methodology
- Causal inference
- Generalizability of causal estimands
- Inference accounting for uncertainty in causal matching

Business and Management

- Factorial designs
- Understanding interactions
- Designs to make best use of big data

Public Opinion Polling

- Addressing biases
- Relationship between responses and actions
- Combining estimates, forecasting

Psychology and Cognitive Science

- Small-n, high-dimensional data
- Designing for highly correlated data
- Generalizability issues
- Replicability issues

Medicine and Public Health

- Big data and precision medicine
- Confounding, confounding, confounding
- Loss to follow up
- Real-World Evidence

Outline

Historical Perspective

The rise of randomization:

- Ronald A. Fisher, *The Design of Experiments*, 1935
- William G. Cochran and Gertrude Cox, *Experimental Design*, 1950
- David R. Cox, *Planning of Experiments*, 1957
- David R. Cox and Nancy Reid, *The Theory of the Design of Experiments*, 2000

Modern Textbooks

- Max D. Morris, *Design of Experiments: An Introduction Based on Linear Models*, 2017
- George C. Canavos and Ioannis A. Koutrouvelis, *An Introduction to the Design and Analysis of Experiments*, 2009
- George E. P. Box, J. Stuart Hunter, and William G. Hunter, *Statistics for Experimenters: Design, Innovation, and Discovery*, 2005
- John Lawson, *Design and Analysis of Experiments with R*, 2014
- Angela Dean, Daniel Voss, Danel Draguljić, *Design and Analysis of Experiments*, 2017
- Douglas C. Montgomery, *Design and Analysis of Experiments*, 2020

More Information

Theory of Causality and Epidemiologic Designs:

- Pearl, *The Book of Why: The New Science of Cause and Effect*, 2018 (Some Material Available Online)
- Hernán and Robins, *Causal Inference: What If*, 2020 (Free Online)

Econometric/Quasi-Experimental Designs:

- Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*, 1966 (Vassar Library)
- Angrist and Pischke, *Mostly Harmless Econometrics*, 2008 (Some Material Available Online)
- Goldsmith-Pinkham, Yale Applied Methods Course Materials, 2021 (Syllabus, Lecture Videos, R Code, Course Materials Online)

More Information

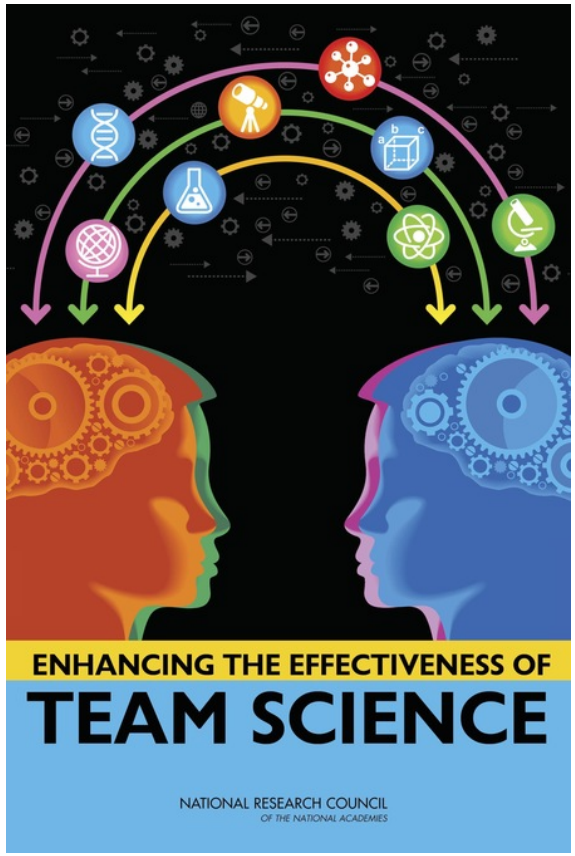
Both:

- Cunningham, *Causal Inference: The Mixtape*, 2021 (Free Online)
- Huntington-Klein, *The Effect: An Introduction to Research Design and Causality, 2021* (Free Online with R, Stata, Python Examples/Code)

Outline

Roles in Research Studies

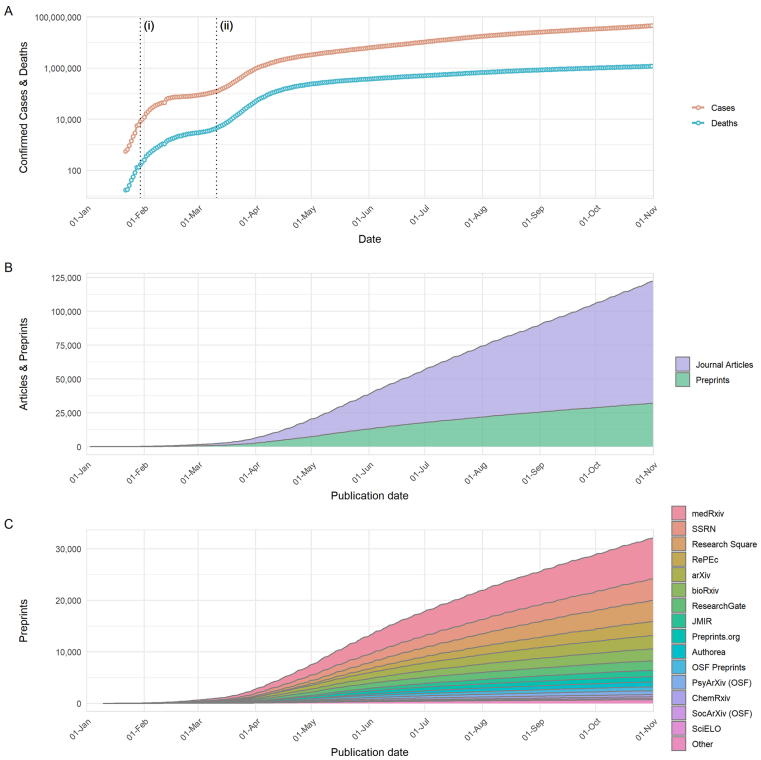
- Investigator
- Analysis Statistician
- Data Manager
- Design Statistician/Methodologist
- Basic Science/Statistics Research
- Scientific Communication



Cooke and Hilton, eds. *Enhancing the Effectiveness of Team Science*, 2015

Outline

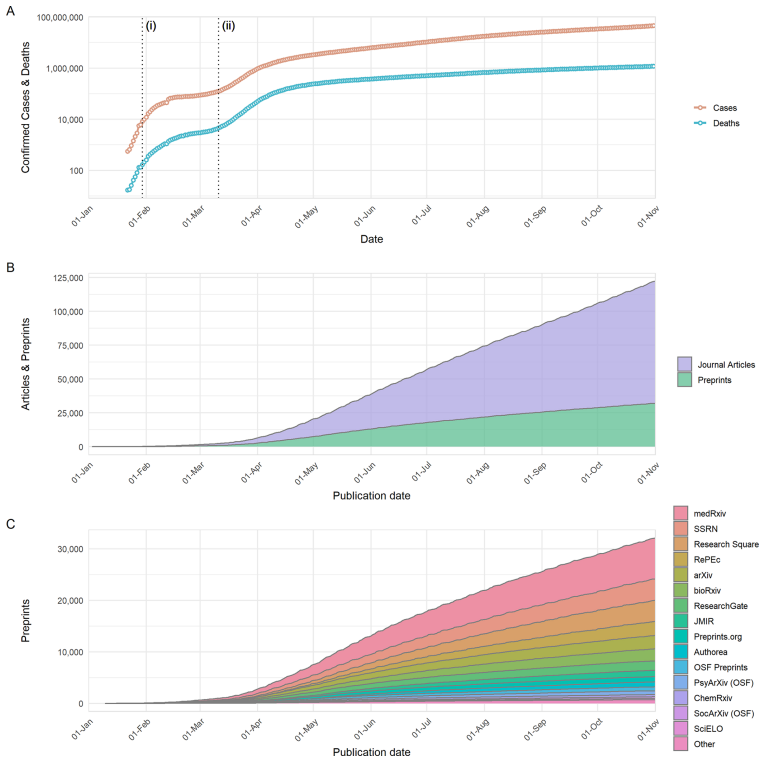
Studies in Today's World



Fraser et al., *PLoS Biology*, 2021

- There's more research than ever before
- There's more importance of and awareness of research than ever before

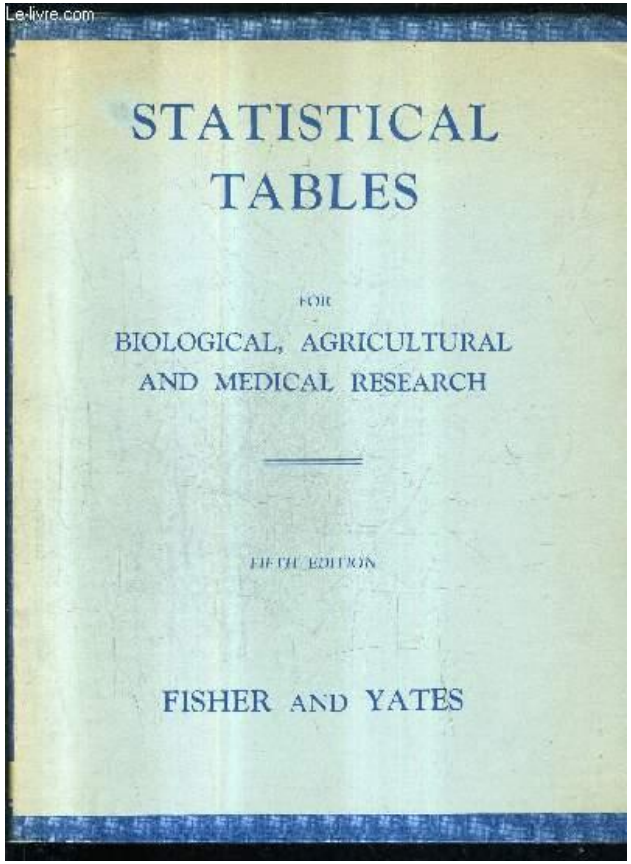
Studies in Today's World



Fraser et al., *PLoS Biology*, 2021

- There's more research than ever before
- There's more importance of and awareness of research than ever before
- Is the research better than ever before?
- Is the research having an appropriate impact?

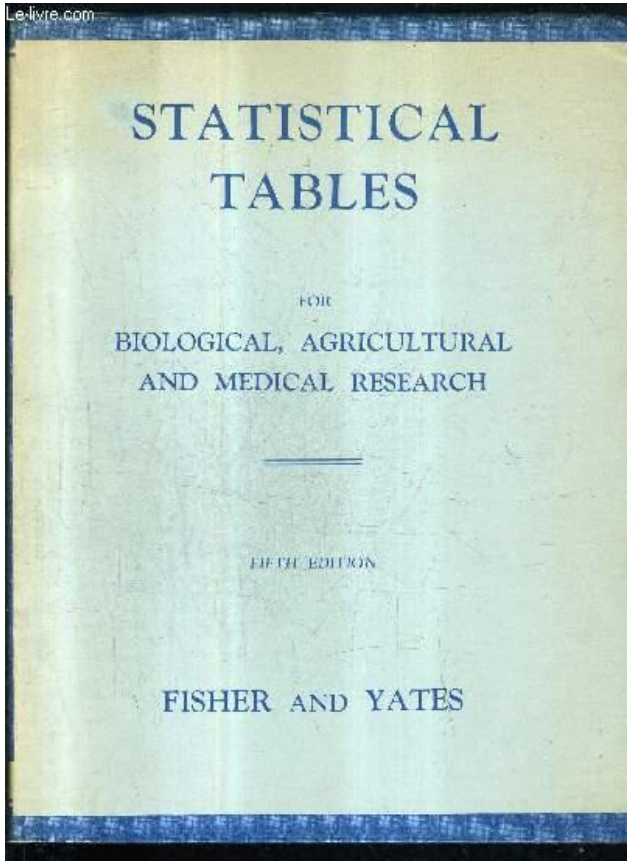
Purposeful Design



- Research used to be DIFFICULT

Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1957

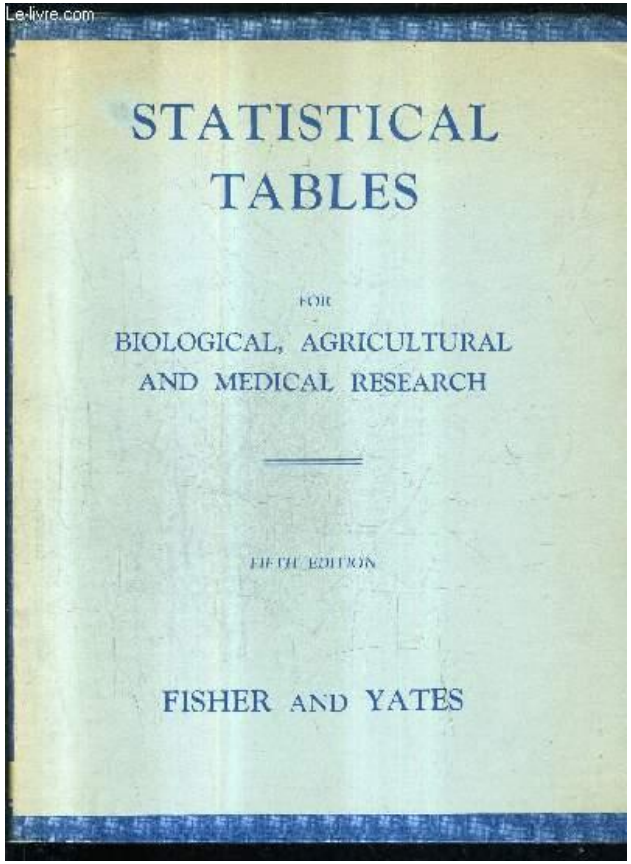
Purposeful Design



- Research used to be DIFFICULT
- Now only good research is difficult

Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1957

Purposeful Design



Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 1957

- Research used to be DIFFICULT
- Now only good research is difficult
- How do we incentivize good, purposefully-designed research studies?

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis
- Noisy information environment

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis
- Noisy information environment
- Low-hanging fruit gone

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis
- Noisy information environment
- Low-hanging fruit gone

Reasons for Optimism

- More opportunities available

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis
- Noisy information environment
- Low-hanging fruit gone

Reasons for Optimism

- More opportunities available
- More informed communication

Optimism or Pessimism?

Reasons for Pessimism

- Lots of bad data analysis
- Noisy information environment
- Low-hanging fruit gone

Reasons for Optimism

- More opportunities available
- More informed communication
- Subtle techniques, bigger data

Measuring What We Want

Do our **exposures/treatments** represent what's possible in the world?

Measuring What We Want

Do our **exposures/treatments** represent what's possible in the world?

Do our **outcomes** measure the big things we really want to learn about?

Measuring What We Want

Do our **exposures/treatments** represent what's possible in the world?

Do our **outcomes** measure the big things we really want to learn about?

How does a desire for **reliable evidence** compete with broader visions of **structural change** and possibility?

Identifying Smaller Effects and Smaller Populations

How do we properly measure **small effects** that are consequential?

Knowing What We Can't Measure

The **limits** of statistics and quantification have never been clearer.

Knowing What We Can't Measure

The **limits** of statistics and quantification have never been clearer.

Understanding study design and its challenges informs the role of **other forms of knowledge and advocacy**.

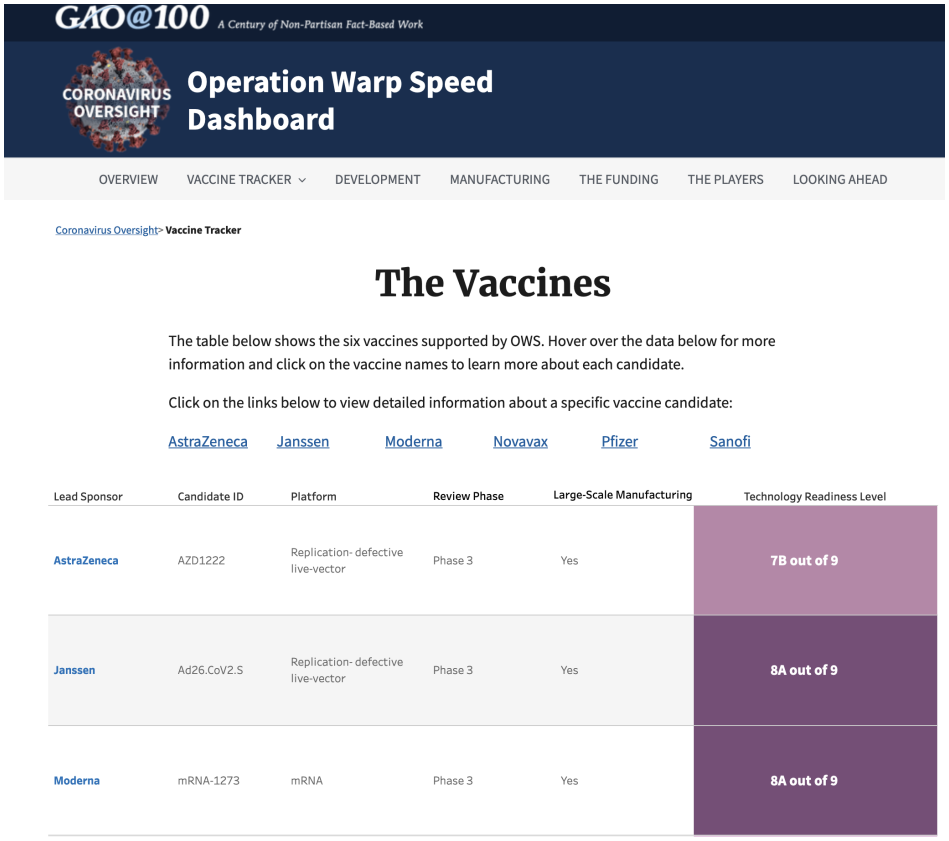
Interpretation and Communication

How do we design, report, and communicate research that can have an **impact on the world**?

Where Do We Go From Here?

Eisen and Tibshirani, *New York Times*, 7/20/2020

Where Do We Go From Here?



GAO@100 A Century of Non-Partisan Fact-Based Work

CORONAVIRUS OVERSIGHT

Operation Warp Speed Dashboard

OVERVIEW | VACCINE TRACKER | DEVELOPMENT | MANUFACTURING | THE FUNDING | THE PLAYERS | LOOKING AHEAD

Coronavirus Oversight > **Vaccine Tracker**

The Vaccines

The table below shows the six vaccines supported by OWS. Hover over the data below for more information and click on the vaccine names to learn more about each candidate.

Click on the links below to view detailed information about a specific vaccine candidate:

AstraZeneca Janssen Moderna Novavax Pfizer Sanofi

Lead Sponsor	Candidate ID	Platform	Review Phase	Large-Scale Manufacturing	Technology Readiness Level
AstraZeneca	AZD1222	Replication- defective live-vector	Phase 3	Yes	7B out of 9
Janssen	Ad26.CoV2.S	Replication- defective live-vector	Phase 3	Yes	8A out of 9
Moderna	mRNA-1273	mRNA	Phase 3	Yes	8A out of 9

Government Accountability Office, *Operation Warp Speed Dashboard*, Dec. 2021

Role of the Statistician

William G. Cochran and Gertrude M. Cox, *Experimental Designs*, 1957, p. 10

"...too little time and effort is put into the planning of experiments. The statistician who expects his contribution to the planning will involve some technical matter in statistical theory finds repeatedly that he makes a much more valuable contribution simply by getting the investigator to explain clearly why he is doing the experiment, to justify the experimental treatments whose effects he proposes to compare, and to defend his claim that the completed experiment will enable its objectives to be realized."