Evidence-Based Decision Making in Healthcare

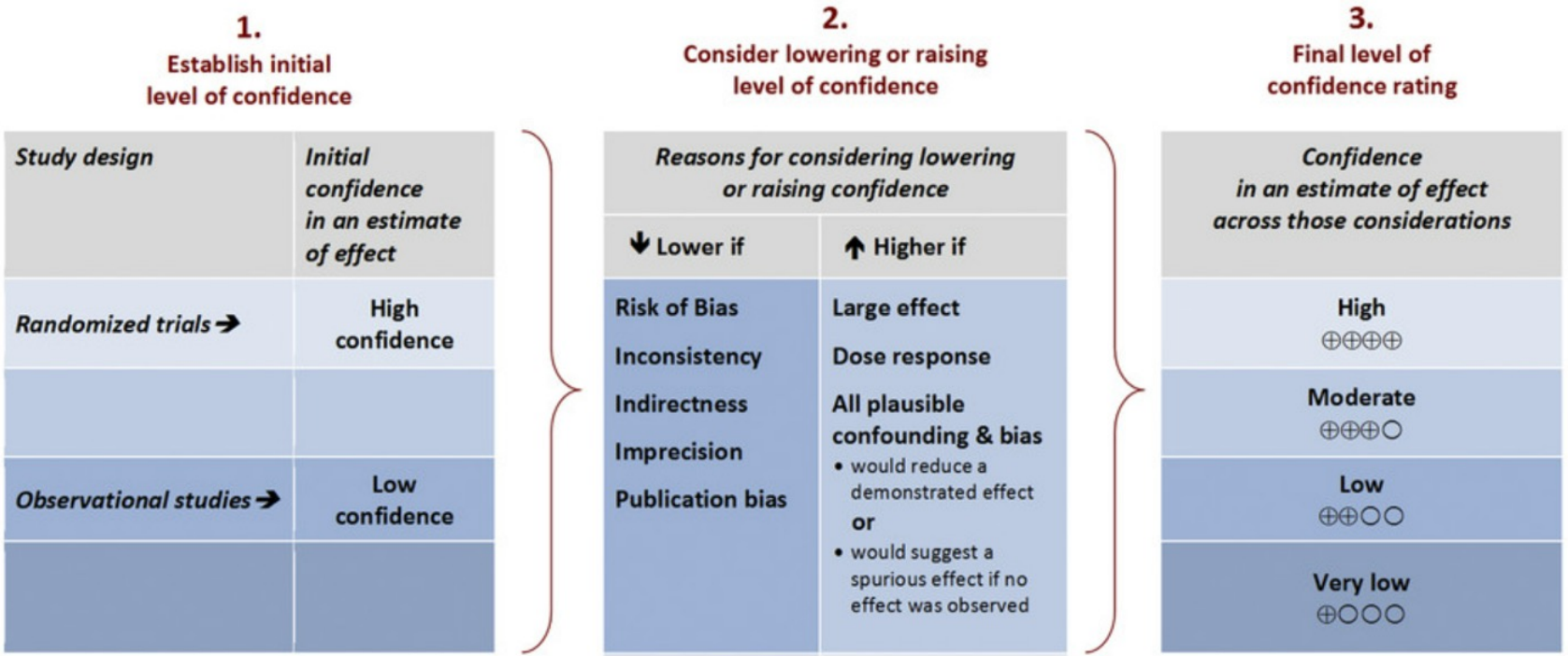
Risk of Bias in Studies and Publication Bias

Dr. Jay K. Varma
https://drjayvarma.com

GRADE

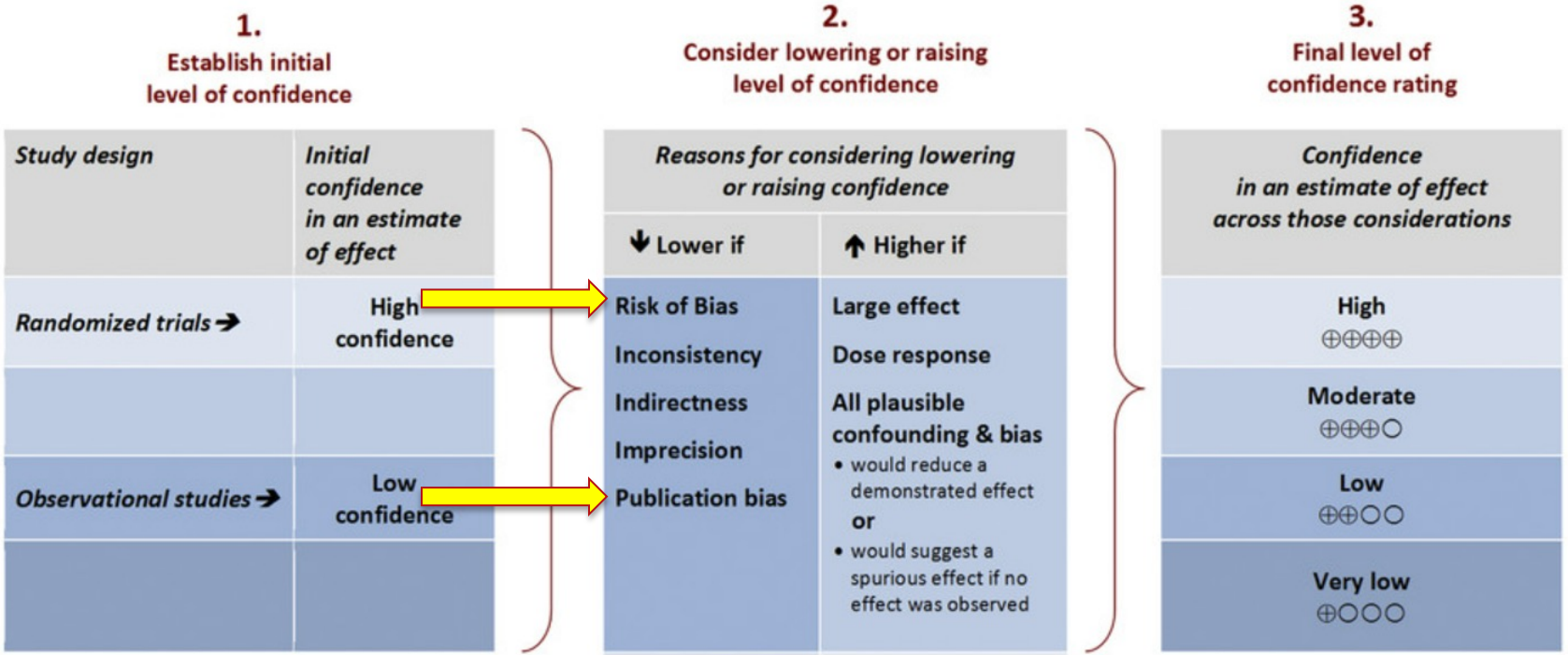
- Summary of evidence and systematic approach to make recommendations
- Reviews quality of evidence with the study design
  - 5 reasons to rate down, 3 reasons to rate up

Rating Quality of Evidence



Source: Environment International 2016

Rating Quality of Evidence



Source: Environment International 2016

Grading Quality of Evidence

Grade	Definition
High 	We are **very confident** that the true effect lies close to that of the estimate of the effect
Moderate 	We are **moderately confident** in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low 	Our **confidence** in the effect estimate **is limited**: the true effect may be substantially different from the estimate of the effect
Very Low 	We have **very little confidence** in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

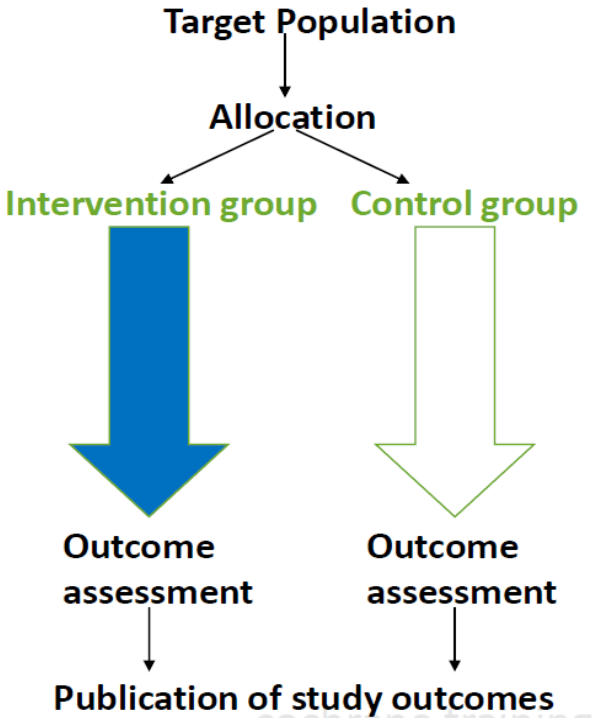
Bias

- A systematic error, or deviation from the truth, in results or inferences
- Risk of bias may be different for each outcome
- Risk of bias should be assessed within each studies and across all studies together
- Risk of bias can downgrade, or reduce, the rating quality by one or two levels

Major Types of Bias in Clinical Trials

- Selection bias, e.g., lack of allocation concealment
- Performance bias, e.g., lack of blinding
- Attrition bias, e.g., incomplete accounting of patients and outcome events
- Selective outcome reporting bias

Trial Flow Chart



Source: Cochrane Training Guide

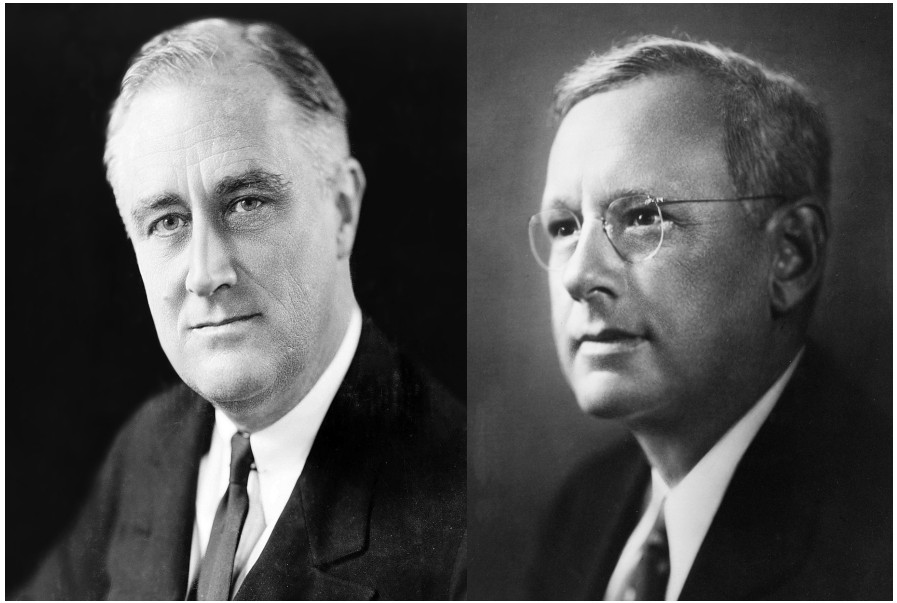
SELECTION BIAS

Selection Bias

- Sampling of a population for a study should, ideally, be randomized, blinded, and representative of the population that will receive the test or treatment
- Limits likelihood that results represent “truth”
- Limits generalizability – can you apply these results to your population in real life and get similar results?

A Famous Example Outside Medicine

- 1936 election
- Democratic President Franklin D. Roosevelt
- Republican Alf Landon
- A referendum on Roosevelt's progress on Great Depression



The Literary Digest

NEW YORK OCTOBER 31, 1936

Topics of the day

LANDON, 1,293,669; ROOSEVELT, 972,897

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballots in the Poll of ten million voters, scattered throughout the forty-eight States of the Union, is now finished, and in the table below we record the figures received up to the hour of going to press.

These figures are exactly as received from more than one in every five voters polled in our country—they are neither weighted, adjusted nor interpreted.

Never before in an experience covering more than a quarter of a century in taking polls have we received so many different varieties of criticism—praise from many; lican National Committee purchased THE LITERARY DIGEST?" And all types and varieties, including: "Have the Jews purchased THE LITERARY DIGEST?" "Is the Pope of Rome a stockholder of THE LITERARY DIGEST?" And so it goes—all equally absurd and amusing. We could add more to this list, and yet all of these questions in recent days are but repetitions of what we have been experiencing all down the years from the very first Poll.

Problem—Now, are the figures in this Poll correct? In answer to this question we will simply refer to a telegram we sent to a

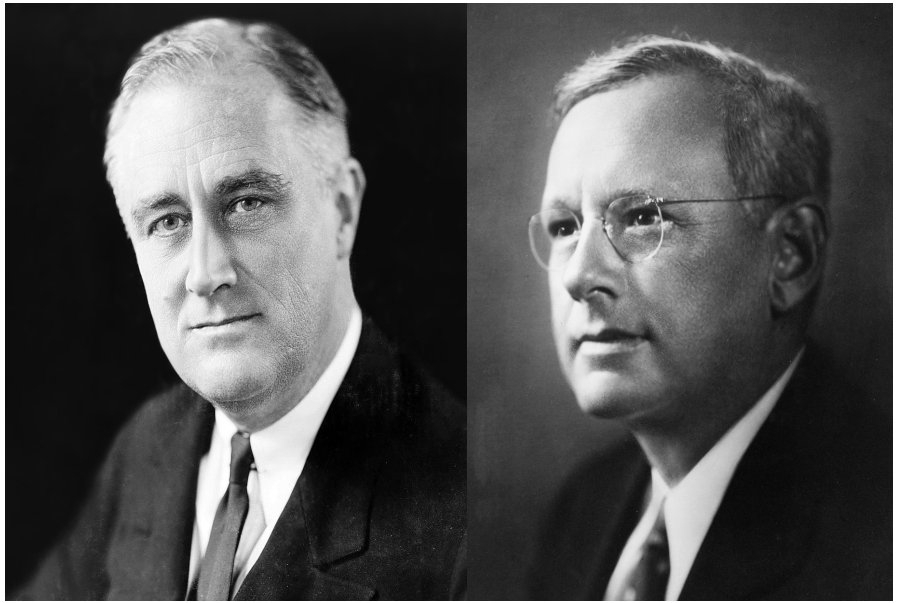
returned and let the people of the Nation draw their conclusions as to our accuracy. So far, we have been right in every Poll. Will we be right in the current Poll? That, as Mrs. Roosevelt said concerning the President's reelection, is in the 'lap of the gods.'

"We never make any claims before election but we respectfully refer you to the opinion of one of the most quoted citizens to-day, the Hon. James A. Farley, Chairman of the Democratic National Committee. This is what Mr. Farley said October 14, 1932:

"'Any sane person can not escape the implication of such a gigantic sampling of popular opinion as is embraced in THE LITERARY DIGEST straw vote. I consider this conclusive evidence as to the desire of the people of this country for a change in the National Government. THE LITERARY

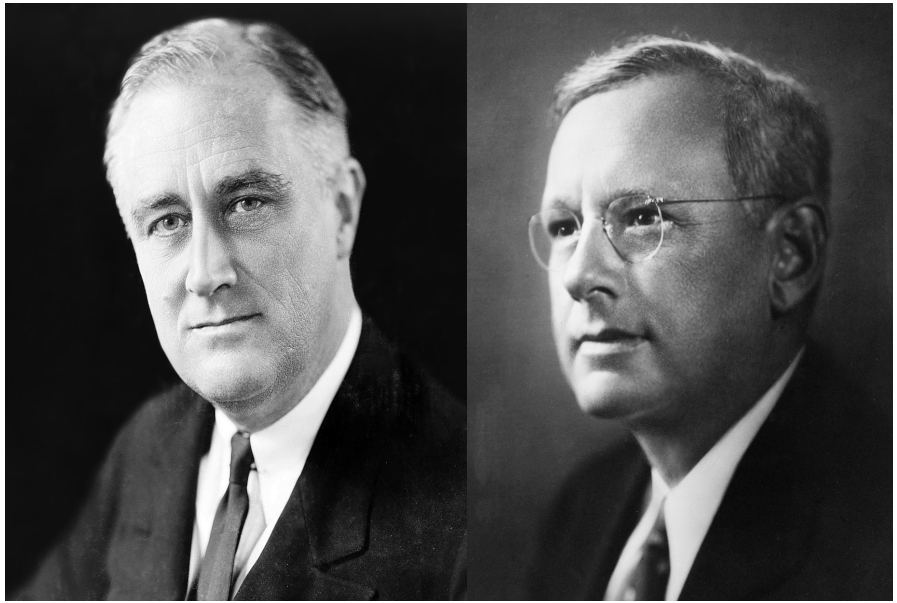
Literary Digest 1936 Poll

- Poll set to ~10M people
- ~2.4M responses

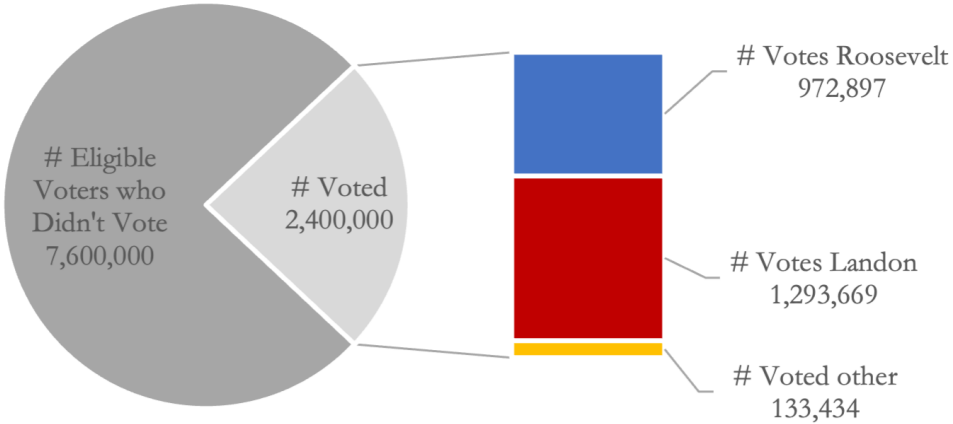


Literary Digest 1936 Poll

- Poll set to ~10M people
- ~2.4M responses
- Landon by a landslide (>60%)

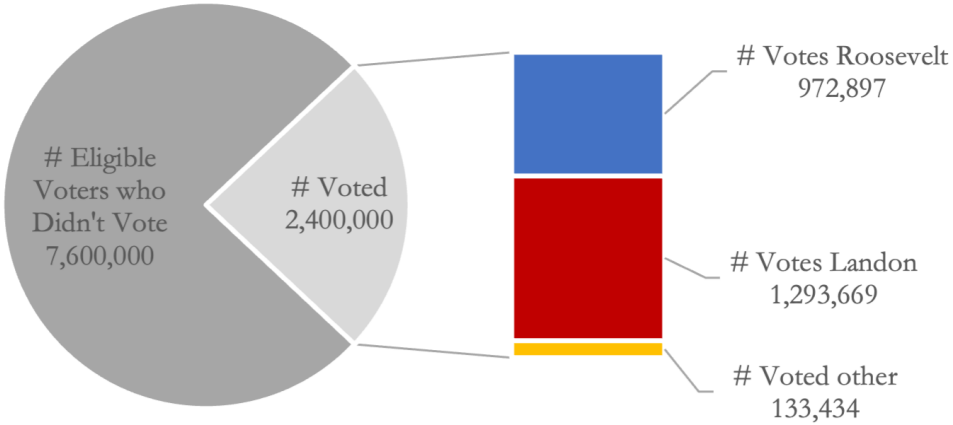


Literary Digest



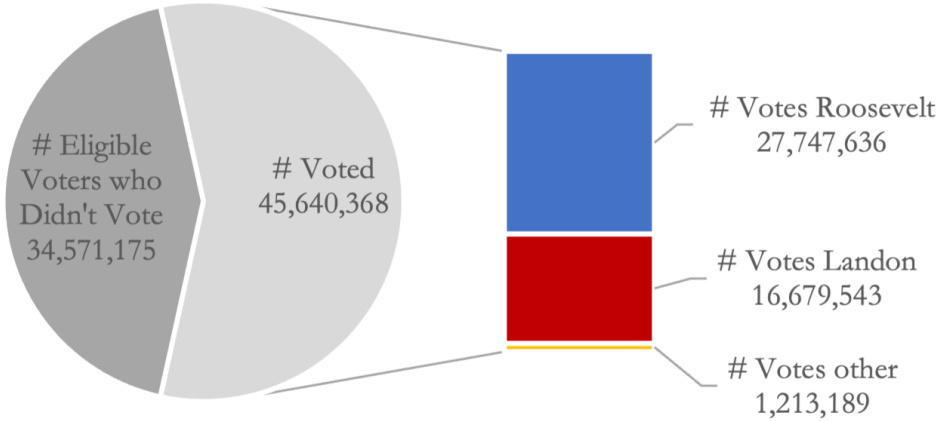
Poll response rate = 24%

Literary Digest



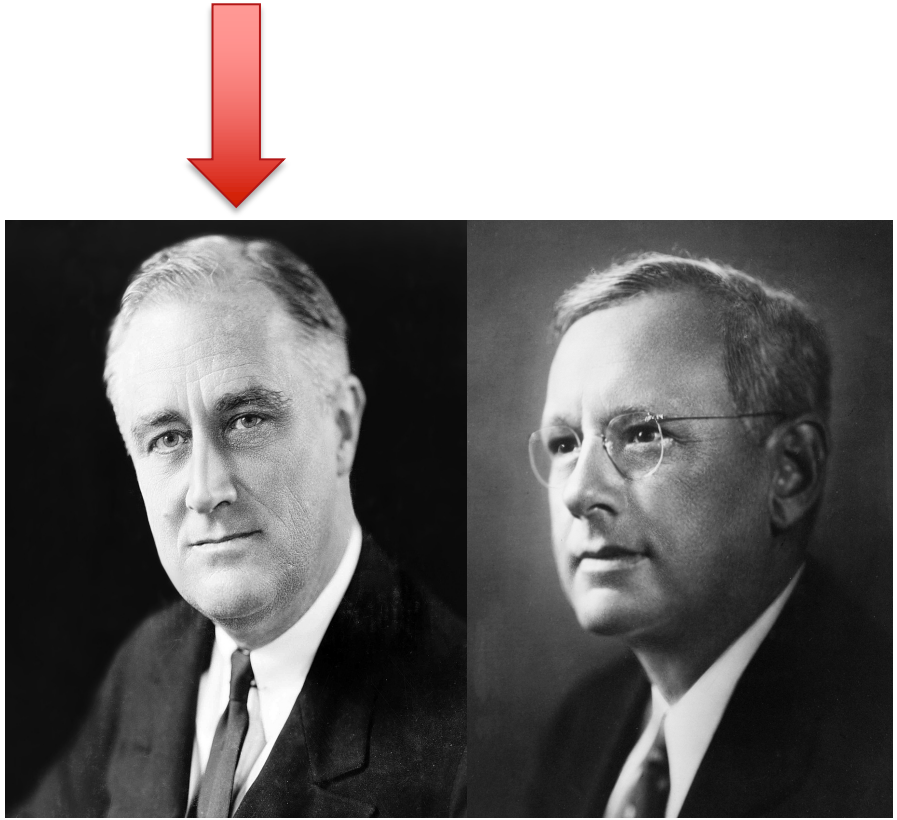
Poll response rate = 24%

Actual Vote



Actual voter turnout = 57%

Actual Result

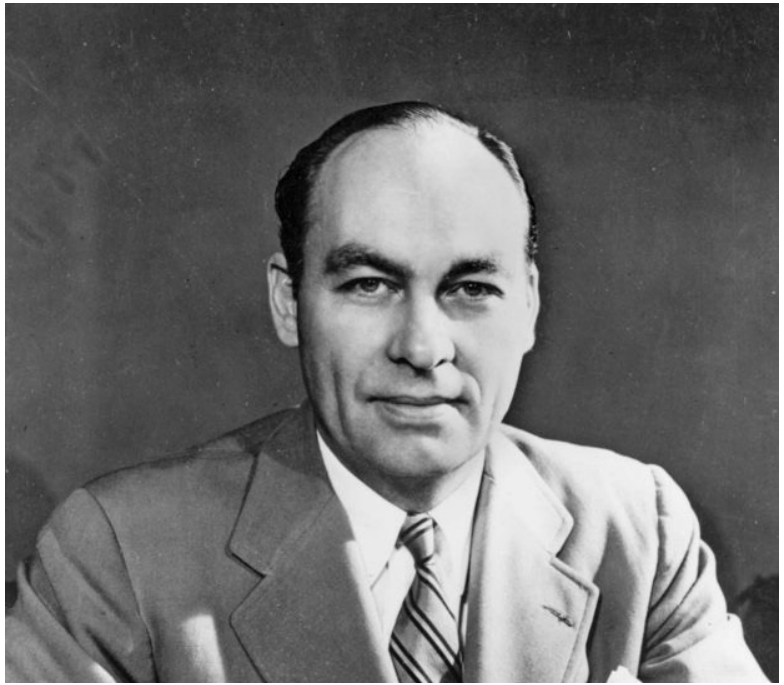


How Did Literary Digest Poll Fail?

- Context: 1936, Great Depression
- Sampled homes using telephone numbers obtained from magazine subscriptions and club memberships
- Who is rich enough to have these memberships?
- Who has a phone?
- Who has a reliable address?
- Who has time and interest to fill out a poll?

Did Gallup Get It Right?

- 1936 Presidential Poll
- Sample size 50,000
- “randomized” sampling
- Predicted Roosevelt Victory 56% (reality would be 61%)
- Limitation: no effort to replicate population proportions



Framingham Heart Study

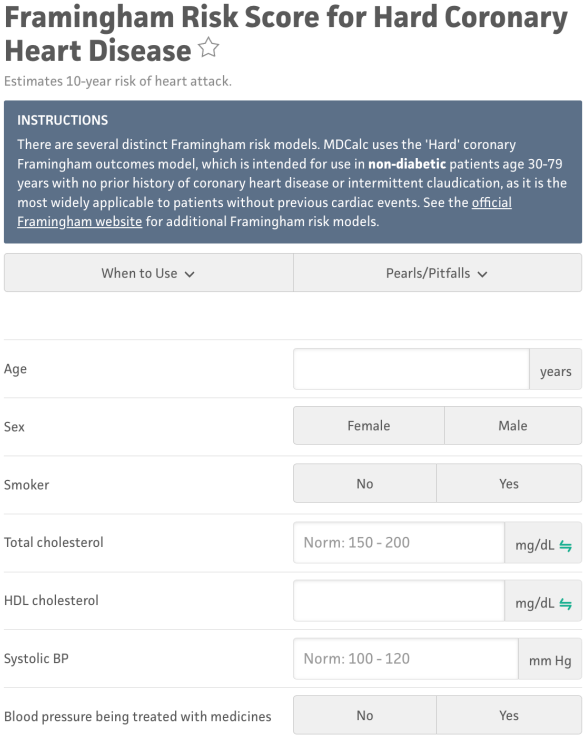
- Intended to identify risk factors for developing cardiovascular disease
- Initiated in 1948
- Initially recruited 5,209 people without cardiovascular disease
- Prospective cohort design

Impact of Framingham Study

- Considered one of the most important prospective observational studies ever conducted
- Source of concept of identifying and quantifying “risk factors” for chronic disease
- Identified smoking, hypertension, hyperlipidemia, and obesity as risk factors
- Identified aspirin as protective factor

Impact of Framingham Study

- “Framingham risk score”
- Over 3000 manuscripts published based on study



Framingham Risk Score for Hard Coronary Heart Disease ☆

Estimates 10-year risk of heart attack.

INSTRUCTIONS

There are several distinct Framingham risk models. MDCalc uses the 'Hard' coronary Framingham outcomes model, which is intended for use in **non-diabetic** patients age 30-79 years with no prior history of coronary heart disease or intermittent claudication, as it is the most widely applicable to patients without previous cardiac events. See the official Framingham website for additional Framingham risk models.

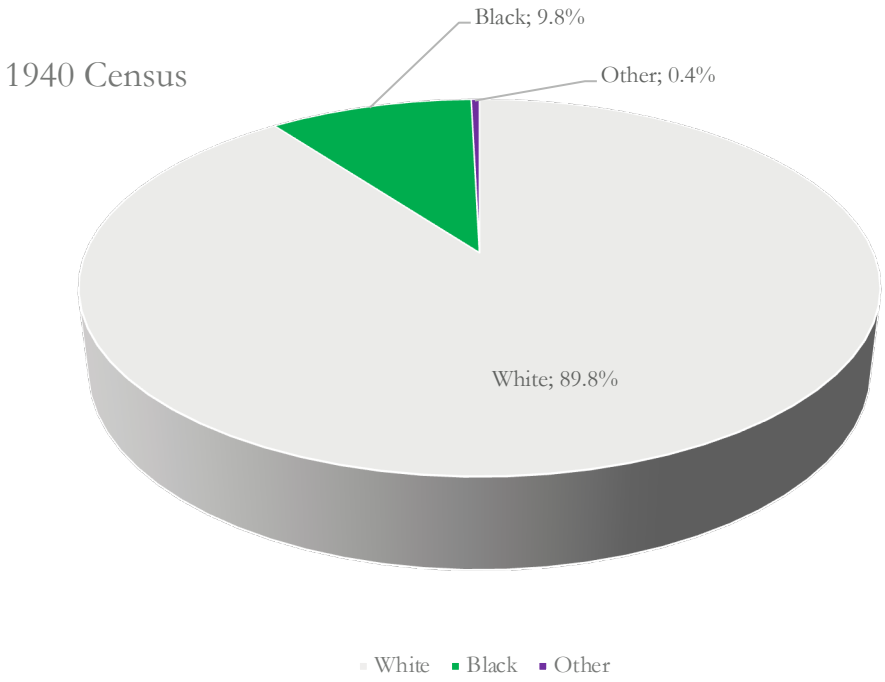
When to Use | Pearls/Pitfalls

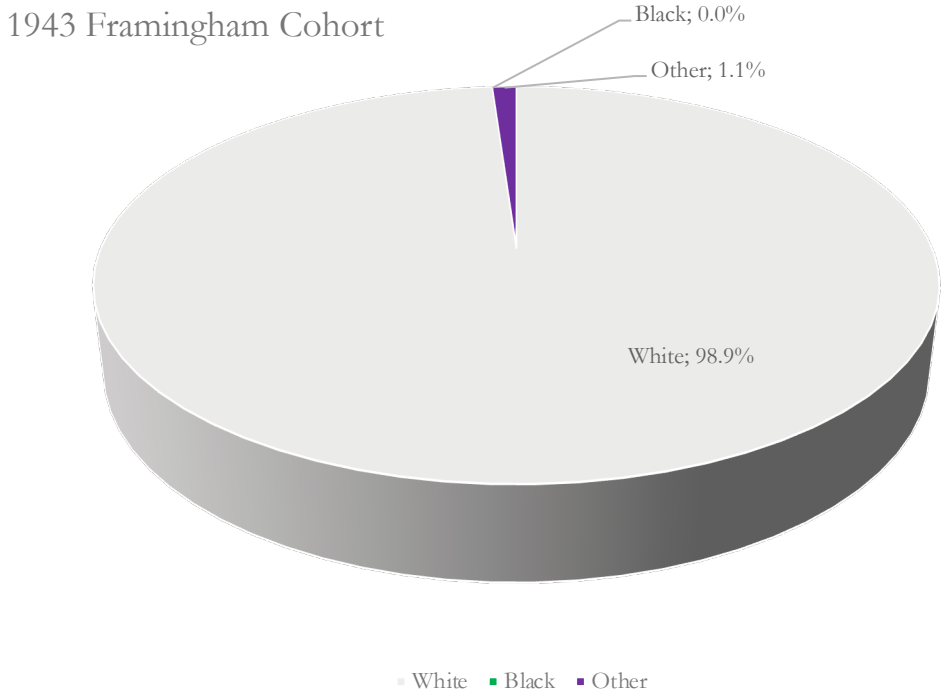
Age		years
Sex	Female	Male
Smoker	No	Yes
Total cholesterol	Norm: 150 - 200	mg/dL ⇋
HDL cholesterol		mg/dL ⇋
Systolic BP	Norm: 100 - 120	mm Hg
Blood pressure being treated with medicines	No	Yes

Limitations of Framingham Study

- Almost entirely white ethnic cohort
- One small geographic area
- Very narrow range of socioeconomic diversity:
  - Initial participants were themselves the very doctors and nurses performing the study
  - Spread to their friends and acquaintances through social networks

Selection Bias in Framingham Study





Generalizability of Framingham Study

- Not a random sample
- Not reflective of national population
- Additional studies funded subsequently, including Jackson Heart Study, to “over-sample” black population

PERFORMANCE BIAS

Performance Bias = Lack of Blinding

- Blinding means people involved in study do not know who is getting intervention and who is getting placebo/control
- Patients
- Researchers
- Recorders and/or adjudicators of outcomes
- Data analysts

Why Is Blinding Important?

- Reduces bias in care, testing, recording outcomes, and interpretation of results
- Such bias is subtle and may not be perceptible
- Ideally study report should specify which groups were blinded, how blinding was maintained, and whether there were violations of blinding

Example from Reading

- 3 clinical trials assessed impact of steroids on all-cause mortality and physical function in patients with spinal injury
- One trial did not blind patients or doctors to steroids vs. placebo
  - Can bias assessment of physical function

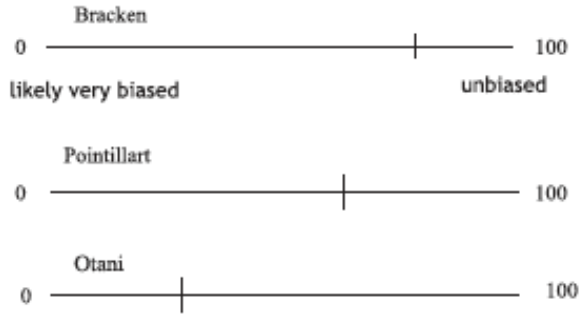


Fig. 1. Validity of three randomized controlled trials addressing the effect of steroids on motor function in acute spinal cord injury.

ATTRITION BIAS

Accounting of Patients & Outcomes

- Loss to follow-up
- Failure to adhere to intention-to-treat or failure to conduct analyses considering only those who adhered to treatment and all patients for whom outcome data are available

Loss to Follow-Up

- Importance of loss to follow-up varies widely
  - Loss to follow-up of 5% in both **intervention and control groups** unlikely to bias study if event rates were 20% and 40%, respectively
- Bias increased if
  - Increased proportion lost to follow-up in relation to intervention and control event rates
  - Differences in loss to follow-up between intervention and control groups

Incomplete Assessment of Outcomes

- Some outcomes should be easier to achieve completeness: all cause mortality (is study patient alive at end of study?)
- Other outcomes require complete assessment, e.g., quality of life

Mitigating Bias During Analysis

- Could loss to follow-up or incomplete outcome assessment change results in an important way?
- Binary outcome: Test a variety of assumptions
  - Code missing as “dead” -> one extremne
  - Code missing as “alive” -> other extreme
  - True effect lies between extremes
- Continuous outcome: Conduct sensitivity analyses

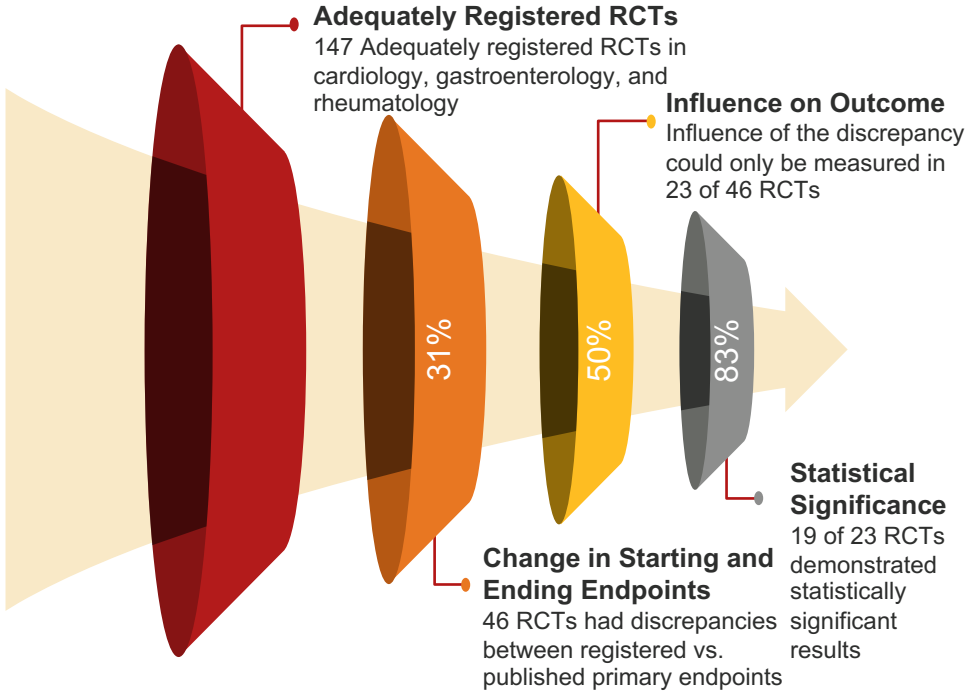
SELECTIVE REPORTING BIAS

Selective Outcome Reporting

- Incomplete reporting of some outcomes and not others – generally to suppressing negative and undesirable findings
- May result in policy or practice using an intervention when it may not be beneficial
- Downgrade quality when using the GRADE framework

Bias in Changing Study Endpoints

- Some investigators change endpoints even after registering a protocol
- Can influence results
- Statistically significant results with a changed primary endpoint can be misleading



Source: Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977e84.

Bias from Unreported Outcomes

- One study found ≥1 efficacy outcome that was not reported in 33% of RCTs in PubMed in 2000
- Fully or partially reported outcomes are 2x as likely to have statistically significant efficacy results vs. unreported outcomes

Source: Chan AW, Altman DG. Identifying outcome reporting bias in randomized trials on PubMed: review of publications and survey of authors. BMJ 2005;330:753.

How to Detect Selective Reporting Bias

- Study protocol from before the study started
- Suspect reporting bias if: 1) the study does not include a key expected outcome or 2) composite outcomes are presented without individual component outcomes
- Consider rating down a body of evidence if suspecting selective reporting bias

An Example of Table Assessing Bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quote: "...participants born on even days were assigned to the experimental group and participants born on odd days were assigned to the control group."
Allocation concealment (selection bias)	High risk	Comment: allocation by date of birth would allow prediction of the allocation sequence.
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "Caffeinated and decaffeinated coffee... was identical in appearance, colour and taste." Comment: it is likely that participants were blinded. Blinding of study personnel was not described.
Blinding of outcome assessment (detection bias) Self-reported outcomes	Unclear risk	Comment: Blinding of outcome assessors was not described.
Blinding of outcome assessment (detection bias) Reaction time	Low risk	Comment: Blinding of outcome assessors was not described, but is unlikely to affect measurement of this outcome.
Incomplete outcome data (attrition bias)	High risk	Comment: outcome data for adverse events were only reported for 53 of 58 participants in the caffeine group. Reasons for loss to follow-up were not described.
Selective reporting (reporting bias)	High risk	Comment: alertness was the primary outcome of the study, but data were not reported. Study protocol was not available to identify any other unreported outcomes. Outcome data were presented for drowsiness although this was not listed as an outcome of interest in the study methods.
Other bias	Low risk	Comment: none were identified.

Source: Cochran Reviews, Chapter 8

PUBLICATION BIAS

Defining Publication Bias

- Selective reporting bias = study published, but negative outcomes not reported
- Publication bias = study not published at all because of negative outcomes
- Systematic reviews when only a few studies are available can overestimate effects, because positive studies more likely to be published initially

Positive Findings

- ~73% published
- ~Published within 4-5 yrs

Negative Findings

- ~41% published
- ~Published within 6-8 yrs

Gray Literature

- Theses, book chapters, compendia of meeting abstract submissions
- Authors may
  - Perceive studies as uninteresting and not see
  - Deterred by rejection from prominent journals
  - Submit studies to local, non-English journals
  - Leads to publishing in obscure journals not indexed at major databases

Mirror Image Phenomenon

A study may be published more than once with slight variations (different authors, presentation) and lead to double counting results in systematic reviews

Randomized Controlled Trials

The risk of publication bias may be higher for reviews that are based on small randomized controlled trials (RCTs), whereas RCTs including large numbers of patients are less likely to remain unpublished or ignored and tend to provide more precise estimates of the treatment effect, whether positive or negative

Discrepancies between results of meta-analyses of small studies and subsequent large trials may occur as often as 20% of the time, and publication bias may be a major contributor to the discrepancies

Although large studies are more likely to be published, sponsors who are displeased with the results may delay or even suppress publication

Observational Studies

- Risk of publication bias is larger for observational studies than for RCTs
- Considered lower standard of evidence, so more likely to be rejected by prominent journals, especially if “negative” findings (eg, no effect shown)

Table 1. Publication bias

Phases of research publication	Actions contributing to or resulting in bias
Preliminary and pilot studies	Small studies more likely to be "negative" (e.g., those with discarded or failed hypotheses) remain unpublished; companies classify some as proprietary information
Report completion	Authors decide that reporting a "negative" study is uninteresting; and do not invest the time and effort required for submission
Journal selection	Authors decide to submit the "negative" report to a nonindexed, non-English, or limited-circulation journal
Editorial consideration	Editor decides that the "negative" study does not warrant peer review and rejects manuscript
Peer review	Peer reviewers conclude that the "negative" study does not contribute to the field and recommend rejecting the manuscript. Author gives up or moves to lower impact journal. Publication delayed
Author revision and resubmission	Author of rejected manuscript decides to forgo the submission of the "negative" study or to submit it again at a later time to another journal (see "journal selection," above).
Report publication	Journal delays the publication of the "negative" study Proprietary interests lead to report getting submitted to, and accepted by, different journals

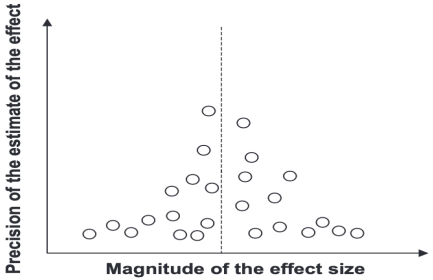
Assessment of Publication Bias

The pattern of study results can indicate bias

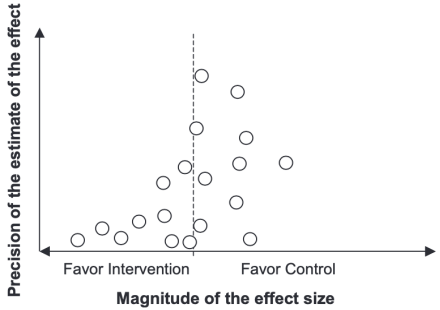
Funnel plots are most popular, but have limitations

Suspicion increases if asymmetrical rather than symmetrical data surrounding the pooled estimate

Funnel Plot



- Larger studies tend to be closer to the pooled estimate
- Effect sizes of smaller studies symmetrically distributed around pooled estimate



- Smaller studies not symmetrically distributed around either pooled estimate or the results of the larger trials
- Trials expected in the bottom right quadrant are missing

Pitfalls of Funnel Plots

- Publication bias not only explanation for asymmetry
- If smaller studies suffer from greater study limitations, they may yield biased overestimates of effects
- A more restricted study population or a more careful administration of the intervention may reveal a larger effect in the small studies

Challenges in Publication Bias

- Difficult to be confident publication bias is absent
- GRADE suggests saying publication bias are “undetected” and “strongly suspected”
- GRADE suggests rating down a maximum of one level (rather than two) for suspicion of publication bias

Example of Publication Bias in Medicine

In 1980, a clinical trial of an anti-arrhythmic drug ‘Lorcainide’ was conducted in 100 patients. Half of the patients were given Lorcainide, half given a placebo

10 patients given the drug died vs. 1 patients given placebo died

Results never published, and development of drug halted. Other companies pursued developed and sold of anti-arrhythmic drugs.

~100,000 patients prescribed these drugs died. Unpublished 1980 results could have prevented some of these deaths.

Example of Publication Bias in Medicine

Seven studies were conducted to compare Reboxetine, an antidepressant, with a placebo. One of these was positive and subsequently published. Other six were negative and left unpublished.

Three trials were published comparing reboxetine with other antidepressants where reboxetine was just as effective; however, three times as many patients' worth of data was collected that showed reboxetine was worse than the other treatments.

Summary

Publication bias can result in substantial overestimates of effect

Suspect when evidence comes from small studies, most of which have been commercially funded

Funnel Plot is most commonly used to assess for publication bias, but has limitations