Critical Appraisal of Medical Literature

Understanding and applying clinical trial results into your practice…

CAPT Ryan Schupbach, PharmD, BCPS, CACP

Vice Chairman, IHS National Pharmacy & Therapeutics Committee

Oklahoma City, OK

Objectives…

- Discuss elements needed in clinical trial dissection and appraisal.

- Interpret important features of sound medical literature.

- Describe basic translational statistics used in formulary management.

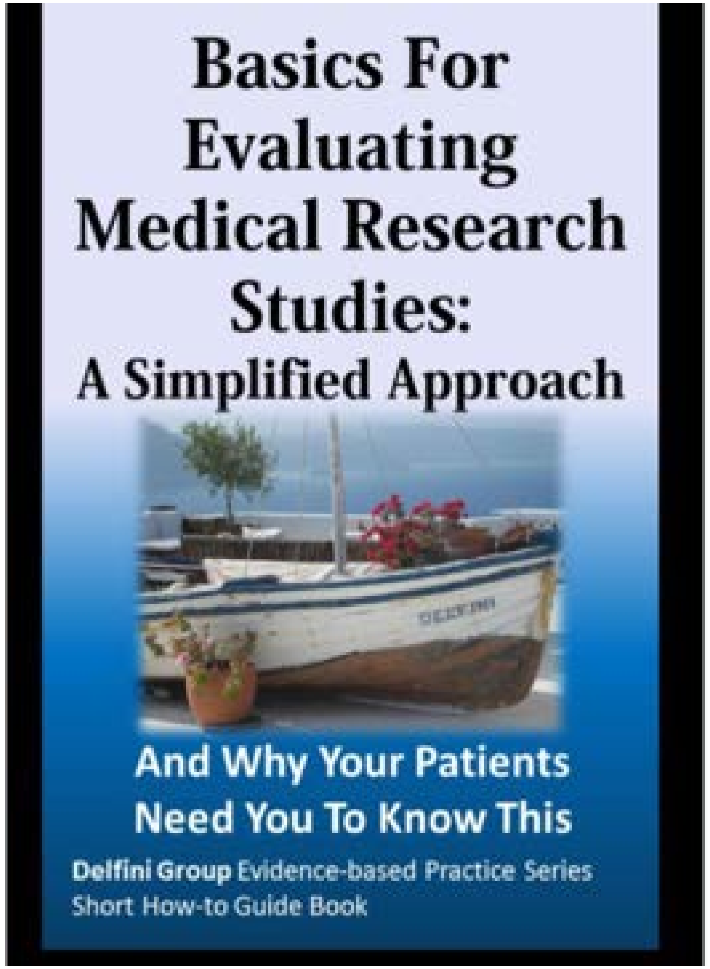
Disclaimer…

- This presentation was developed solely with pharmacotherapeutic applications/interventions in mind.

- May be most applicable when reviewing literature on new or untested medication therapies, where efficacy measures are the primary endpoints and the trial is designed to show superiority of the intervention.

The Delfini Group…

- Used as primary resource throughout the presentation.
- Excellent read and nice guide to evaluating medical literature.
- “Determining if health care evidence is reliable requires critical appraisal for validity and clinical usefulness.”



Basics For Evaluating Medical Research Studies:

A Simplified Approach

And Why Your Patients Need You To Know This

Delfini Group Evidence-based Practice Series

Short How-to Guide Book

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Litany of literature…

- The National Library of Medicine, the world's largest library, publishes approximately 13,000 references each week…
    - In 2010, 1 new medical article was published every 26 seconds.
    - Clinicians need to read ~5,000 articles per day to stay up-to-date.
    - 2017 saw record number of FDA approvals for new drugs.

- How can busy clinicians remain current with this barrage of literature?
    - Many clinicians rely on abstracts which are frequently inaccurate.
    - One study found 18-68% of abstracts in the 6 "top –tier" medical journals contained information not verifiable in the body of the article.

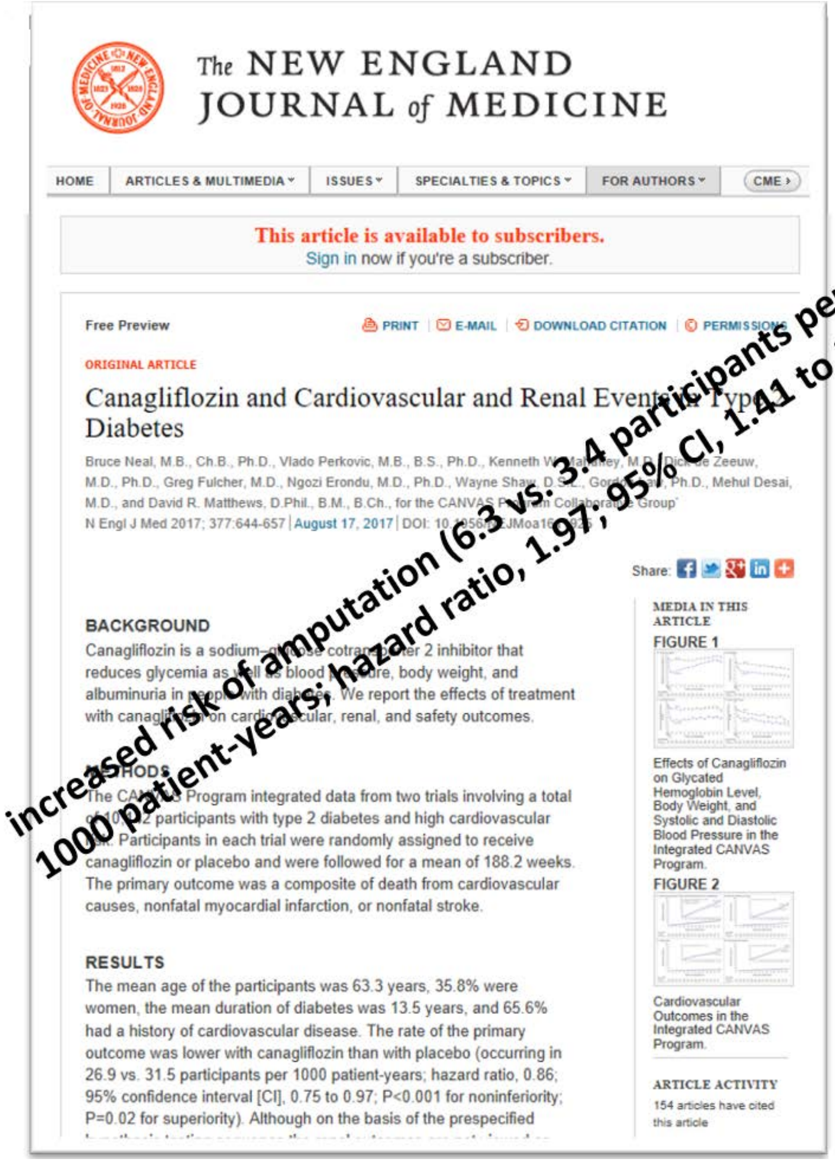
Garba S, Ahmed A, Mai A, et al. Oman Med J 2010; 25(4):311-314.
Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.
Pitkin Rm, Branagan M, Burmeister L. JAMA. 1999;281(12):1110-1111.

The current problem…

- Majority of healthcare decision-makers do not have skillsets to evaluate medical research for validity (closeness to truth) and usefulness.

- Patients deserve to know the <u>benefits and risks</u> of interventions and likelihood of experiencing various outcomes.

- Patient <u>preferences</u> are likely to differ if patients are provided with information on the quality of evidence and amount of risk/benefit.



The NEW ENGLAND JOURNAL of MEDICINE

This article is available to subscribers.
Sign in now if you're a subscriber.

Free Preview

ORIGINAL ARTICLE

Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes

Bruce Neal, M.B., Ch.B., Ph.D., Vlado Perkovic, M.B., B.S., Ph.D., Kenneth W. Mahaffey, M.D., Dick de Zeeuw, M.D., Ph.D., Greg Fulcher, M.D., Ngozi Erondu, M.D., Ph.D., Wayne Shaw, D.Sc., Gordon Law, Ph.D., Mehul Desai, M.D., and David R. Matthews, D.Phil., B.M., B.Ch., for the CANVAS Program Collaborative Group

N Engl J Med 2017; 377:644-657 | August 17, 2017 | DOI: 10.1056/NEJMoa1611925

BACKGROUND
Canagliflozin is a sodium–glucose cotransporter 2 inhibitor that reduces glycemia as well as blood pressure, body weight, and albuminuria in people with diabetes. We report the effects of treatment with canagliflozin on cardiovascular, renal, and safety outcomes.

METHODS
The CANVAS Program integrated data from two trials involving a total of 10,142 participants with type 2 diabetes and high cardiovascular risk. Participants in each trial were randomly assigned to receive canagliflozin or placebo and were followed for a mean of 188.2 weeks. The primary outcome was a composite of death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke.

RESULTS
The mean age of the participants was 63.3 years, 35.8% were women, the mean duration of diabetes was 13.5 years, and 65.6% had a history of cardiovascular disease. The rate of the primary outcome was lower with canagliflozin than with placebo (occurring in 26.9 vs. 31.5 participants per 1000 patient-years; hazard ratio, 0.86; 95% confidence interval [CI], 0.75 to 0.97; P<0.001 for noninferiority; P=0.02 for superiority). Although on the basis of the prespecified

Effects of Canagliflozin on Glycated Hemoglobin Level, Body Weight, and Systolic and Diastolic Blood Pressure in the Integrated CANVAS Program.

Cardiovascular Outcomes in the Integrated CANVAS Program.

154 articles have cited this article

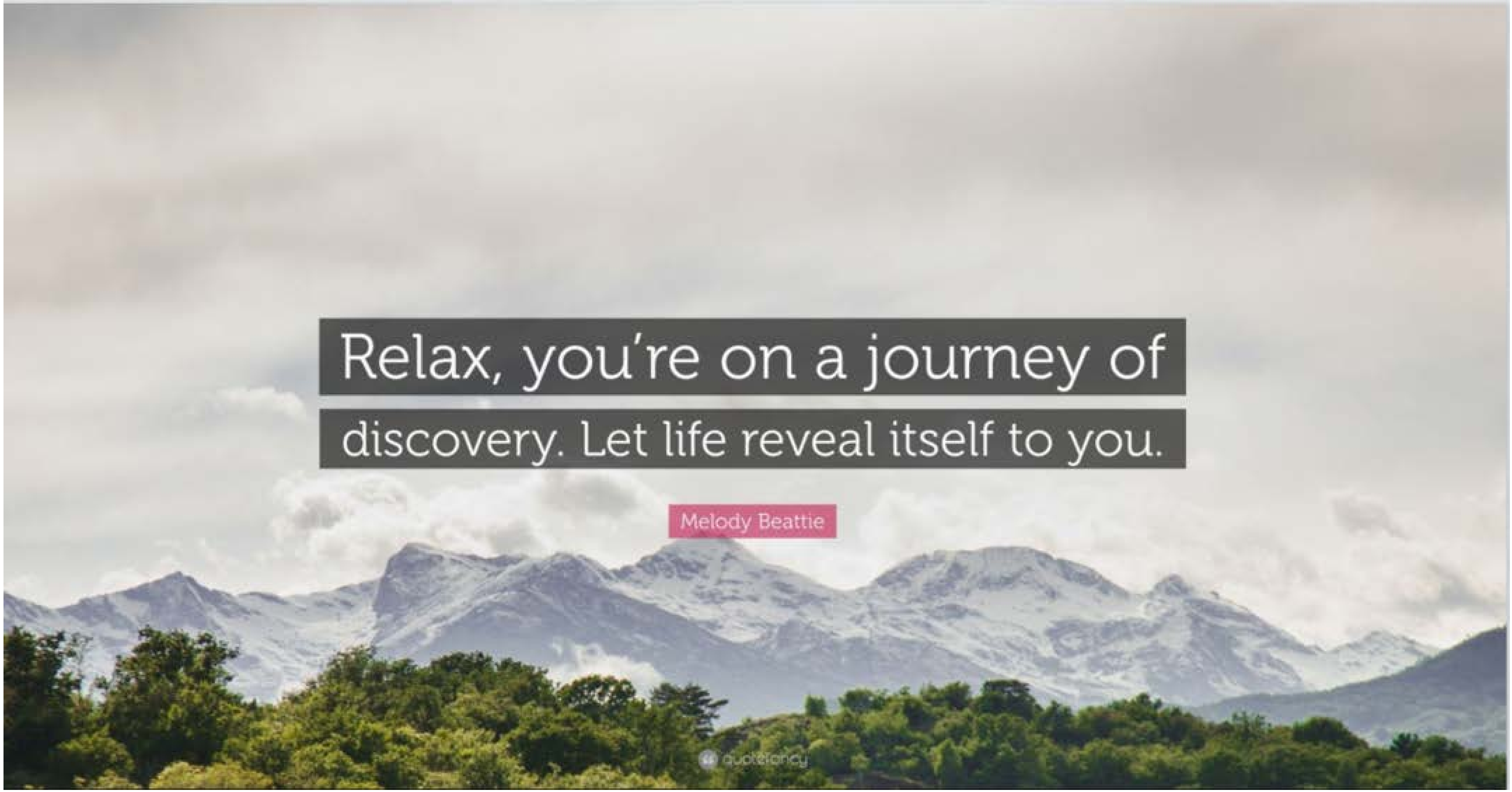
increased risk of amputation (6.3 vs. 3.4 participants per 1000 patient-years; hazard ratio, 1.97; 95% CI, 1.41 to 2.75)

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Critical Appraisal of Published Literature

- Acquiring basic critical appraisal skills is easy; doesn’t involve “heavy lifting” over statistics.
- There is no best way to critically appraisal a trial.
- Critical appraisal helps clinicians conclude beneficial outcomes reported in trials were not caused or distorted by bias or chance.
- Critical appraisal focus:
  - <u>Finding the problems</u> in the study, not the positives of the study.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

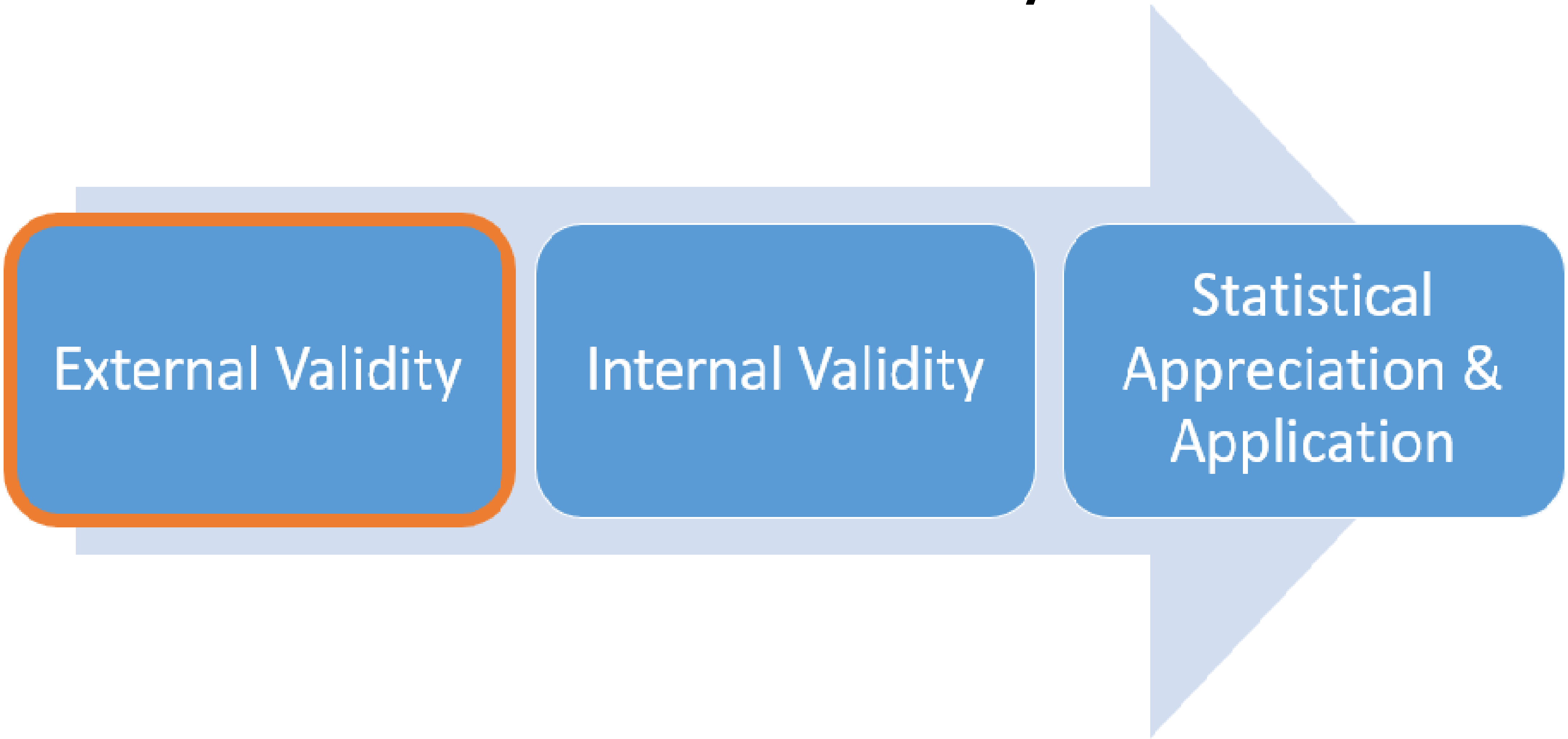


Relax, you're on a journey of discovery. Let life reveal itself to you.
Melody Beattie
quotefancy

“Critical appraisal is inexact and a process of discovery.” – The Delfini Group

https://quotefancy.com/quote/866919/Melody-Beattie-Relax-you-re-on-a-journey-of-discovery-Let-life-reveal-itself-to-you

Medical Literature Appraisal – the Process
External Validity



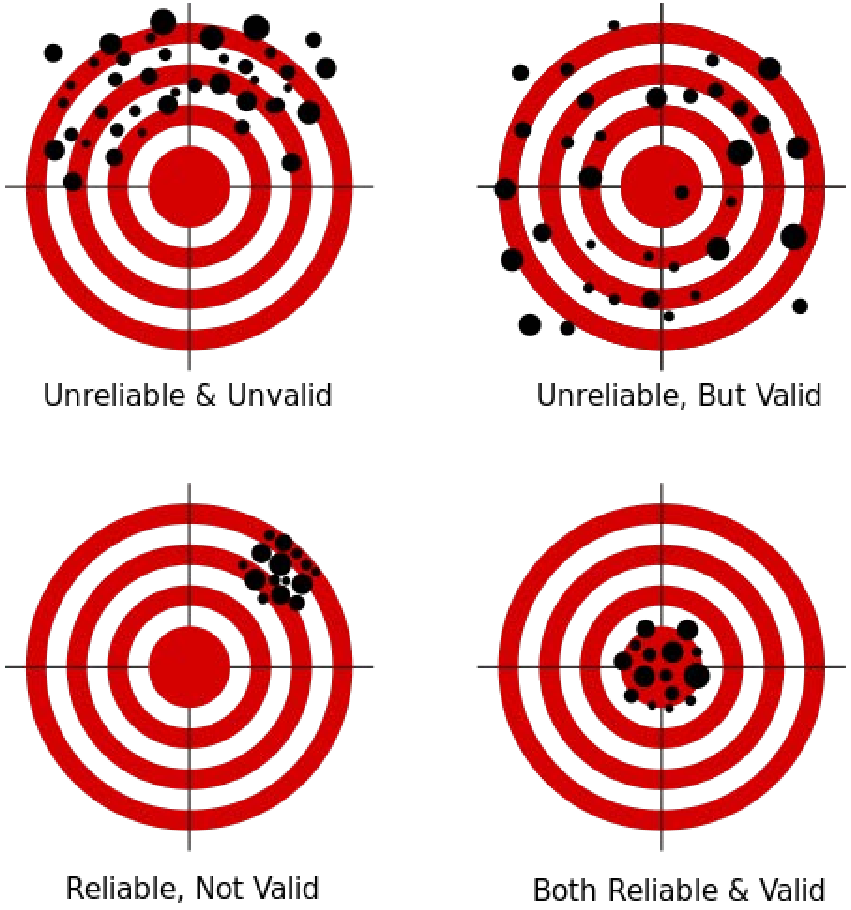
External Validity

Internal Validity

Statistical Appreciation & Application

Definitions: Medical Literature Appraisal (1 of 2)

- **Validity:**
  - The degree to which a study achieves the aim for which it was designed.
  - Does it represent the truth?
- **Reliability:**
  - The degree of consistency between repeated measures of the same thing.
  - If the study was repeated, would the same data be obtained?



Unreliable & Unvalid

Unreliable, But Valid

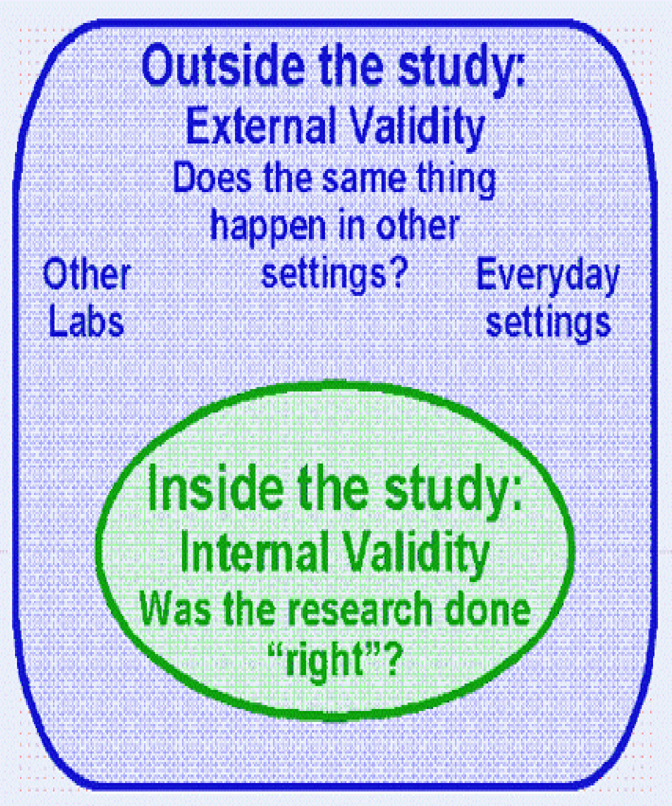
Reliable, Not Valid

Both Reliable & Valid

Garba S, Ahmed A, Mai A, et al. Oman Med J 2010; 25(4):311-314.
https://rampages.us/bplache660blog/wp-content/uploads/sites/7210/2015/07/validity.png

Definitions: Medical Literature Appraisal (2 of 2)

- **External** validity
  - Can the conclusions be applied in settings different to that used in the study?
  - *"Can I apply these conclusions to my patients?"*
- **Internal** validity
  - The ability of the study design to measure what it was intended to measure.
  - *"Can I rely on the conclusions of this study?"*



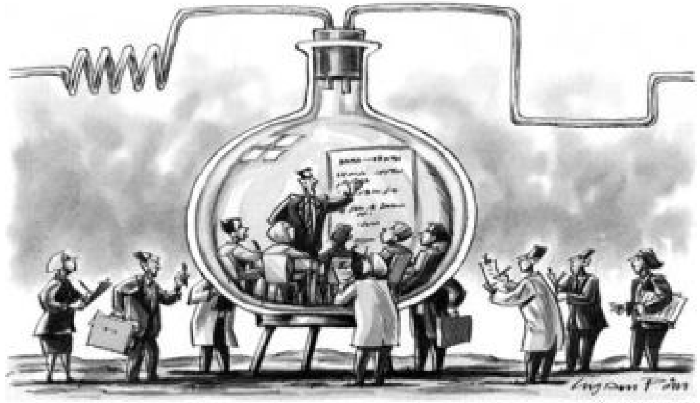
Garba S, Ahmed A, Mai A, et al. Oman Med J 2010; 25(4):311-314

External Validity & Usefulness of Studies (1 of 2)

- **Are the populations studied similar to my patients?**
  - Small percentage of AI/AN patients at study inclusion, grossly underrepresented.
  - Need to recognize genetic variation and drug metabolism.
    - Ethnic polymorphisms can affect pharmacokinetic/pharmacodynamic properties of drugs.
  - International or regionally-specific? Multi-centered or single site study?

  Inspect the inclusion and exclusion criteria.

- **If so, will the results be applicable to my patients?**
  - 394 Taiwanese diabetics in an inpatient hospital in Taipei?
  - Any study from the Department of Veterans Affairs (VA).



Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

https://larspsyll.files.wordpress.com/2013/03/test-tube.jpg?w=360&h=198

External Validity & **Usefulness of Studies** (2 of 2)

- **Are study outcomes meaningful to my patients?**
  - Five "primary" outcomes include:

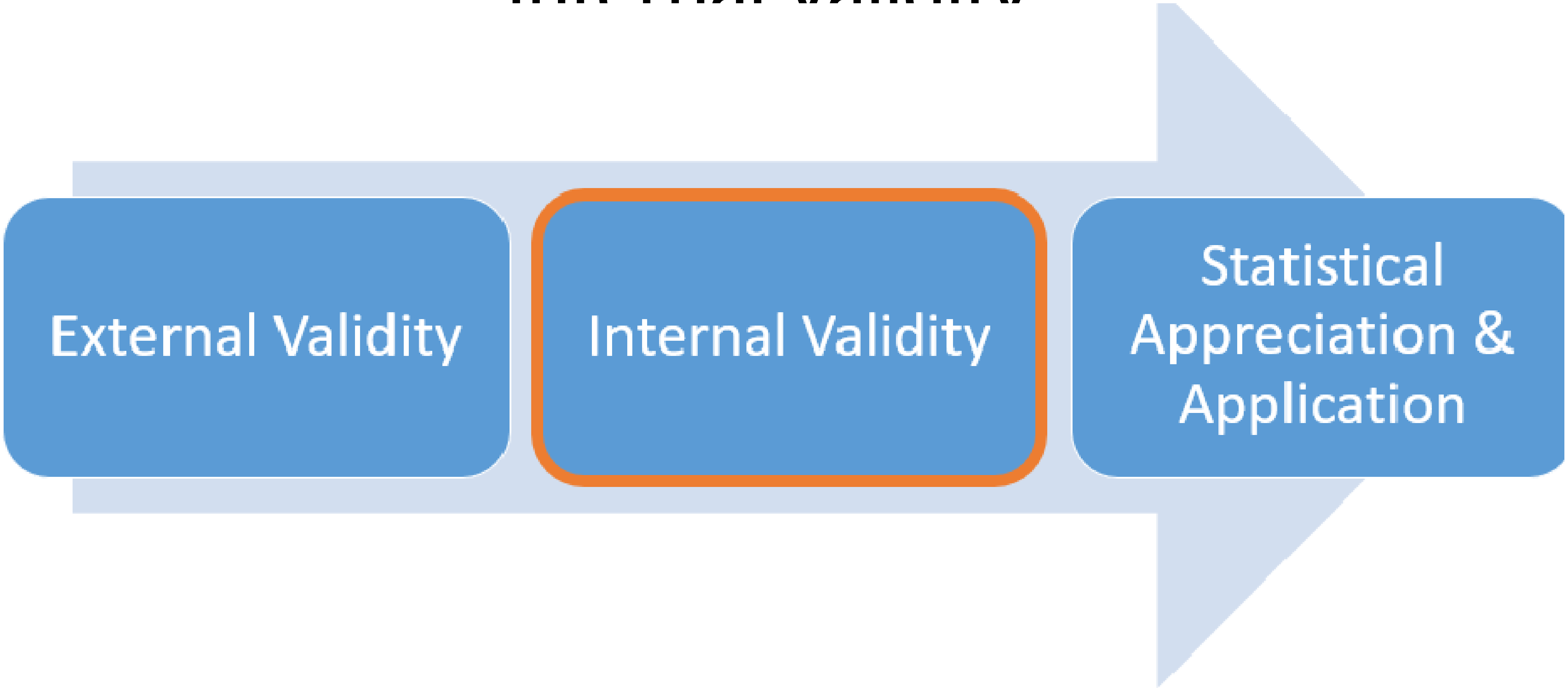
    (1) Morbidity, (2) Mortality, (3) Symptom relief, (4) Quality of Life, (5) Functioning (mental/physical/emotional).

  - Surrogate outcomes: aka "intermediate outcomes"
    - LDL, BP, A1c, imaging results.
  - Composite Outcomes
    - Consist of two or more component outcomes (e.g., death or chest pain)
    - Rise in statistical efficiency (due to rise in event rates) which reduces sample size requirement, cost and time.
    - Patient experiencing any <u>one</u> of the events are considered to have experienced the composite outcome.
    - CAUTION! - Evaluate components collectively <u>AND individually</u>; composite results can be misleading.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.
Cordoba G, Schwartz L, Woloshin S, et al. BMJ 2010;341:c3920.

Medical Literature Appraisal – the Process

Internal Validity



“There are only a handful of ways to do a study properly, but one thousand ways to do it wrong.” - McMaster University

Determining Interval Validity… **Study Design** (1 of 2)

"At the most basic level, study designs are either experimental or observational."

Observational studies:

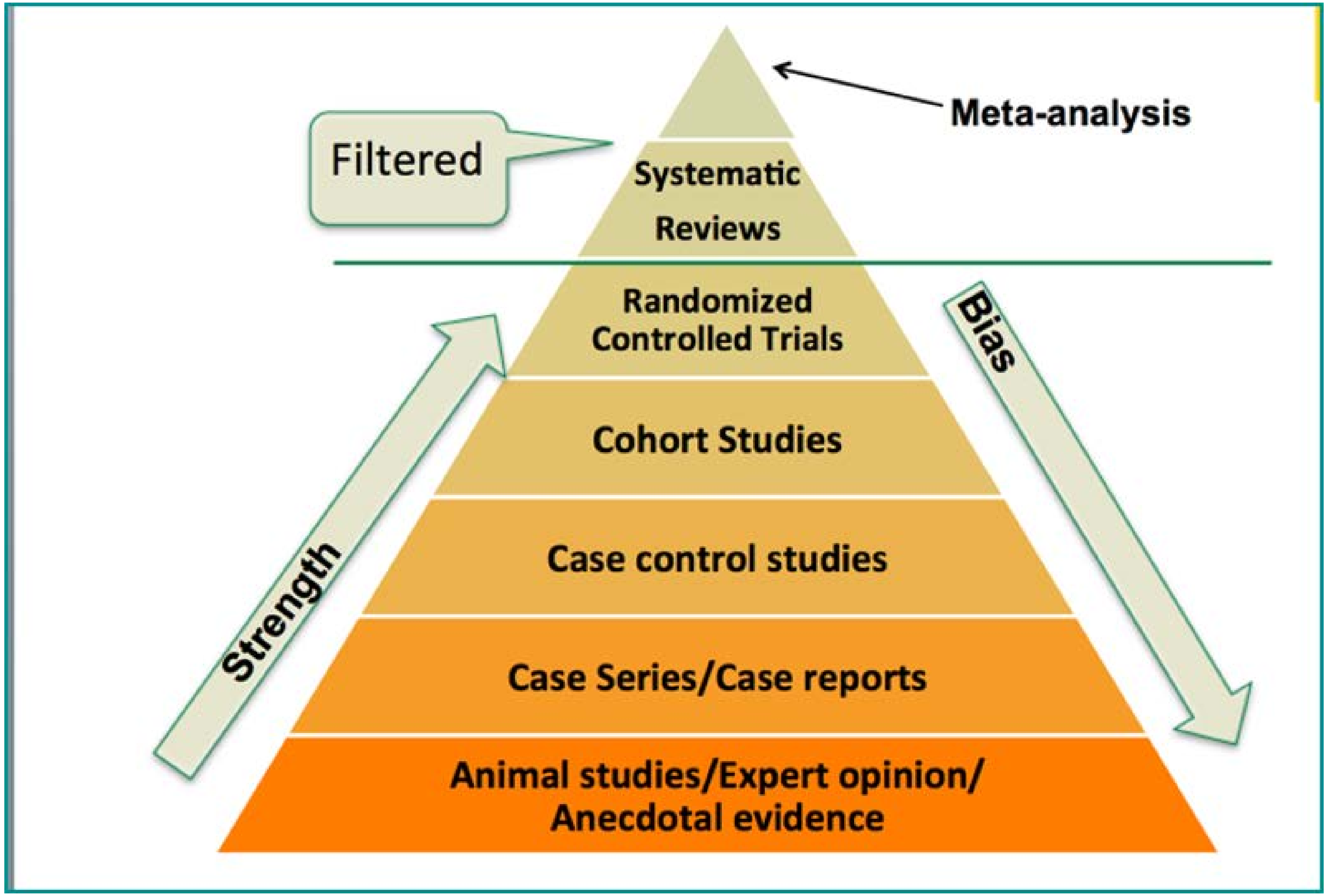
- Helpful in telling about prognosis and natural history of disease.
  - "Real world" use, medication adherence, detect signals about benefits and risks.
- Not useful when answering questions about cause and effect.
- Highly prone to bias, hypothesis-generating only.
- Estimated chance of observational studies (for therapies) being correct as low as 20%.

Experimental studies:

- *Only method* to establish true cause and effect of therapeutic interventions.
- Involves random assignment of participants into groups (control and treatment).
- When evaluating literature for therapeutic efficacy/safety, avoid observational studies.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Determining Interval Validity... **Study Design** (2 of 2)



http://neoreviews.aappublications.org/content/neoreviews/7/9/e474/F5.large.jpg

Internal Validity – the Study Details: Bias

- There are 4 reasons that can explain the relationship between what is studied (intervention) and the results from the study (outcomes):
  - **Bias**
  - **Confounding**
  - **Chance**
  - **Cause and effect (aka "truth")**

- A study finding without bias/confounding and not due to chance is said to have "internal validity."

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Bias…

- Definition(s):
  - Anything in the study that leads us away from the truth (other than chance).
  - Any difference between study groups other than what is being studied (intervention) is automatically a bias.
  - Systematic errors that encourage one outcome over another.

- Bias in studies tends to favor the intervention.

- Everyone involved in research should be assumed to be biased.
  - Industry funding.
  - Academic stature.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Types of study biases…

The same type of biases often have differing names and differing levels of importance.

- Attrition bias
- Classification bias
- Performance bias
- Publication bias
- Recall bias
- Reporting/research bias
- Selection/sampling bias

"There's a flaw in your experimental design. All the mice are scorpios."

How do we (attempt to) Mitigate bias in a trial?

Certain trials characteristics are recommended:

- **Randomization** (or “allocation concealment and sequencing”)
  - Prevents selection bias; ensures each patient has equal chance of receiving either treatment; allocation concealment necessary – sealed envelopes.
- **Blinding**
  - Purpose: prevent bias associated with patients’ and researchers’ expectations.
  - Single-, double- and triple-blinding (e.g., outcome assessors).
  - Inadequate blinding shown to distort trial results by ~ 70%.
- **Follow-up** (of missing patient data)
  - Missing data (protocol deviations, drop-outs, side effects) can mislead results.
  - Intention-To-Treat design can avoid attrition bias.

Gluud L. Am J Epidemiol 2006;163(6); 493-501

Intention-To-Treat (ITT) vs. Per-Protocol…

Intention-To-Treat:

- Comparison of groups that includes all patients as originally allocated after randomization.
- Recommended method in superiority trials to avoid any bias.

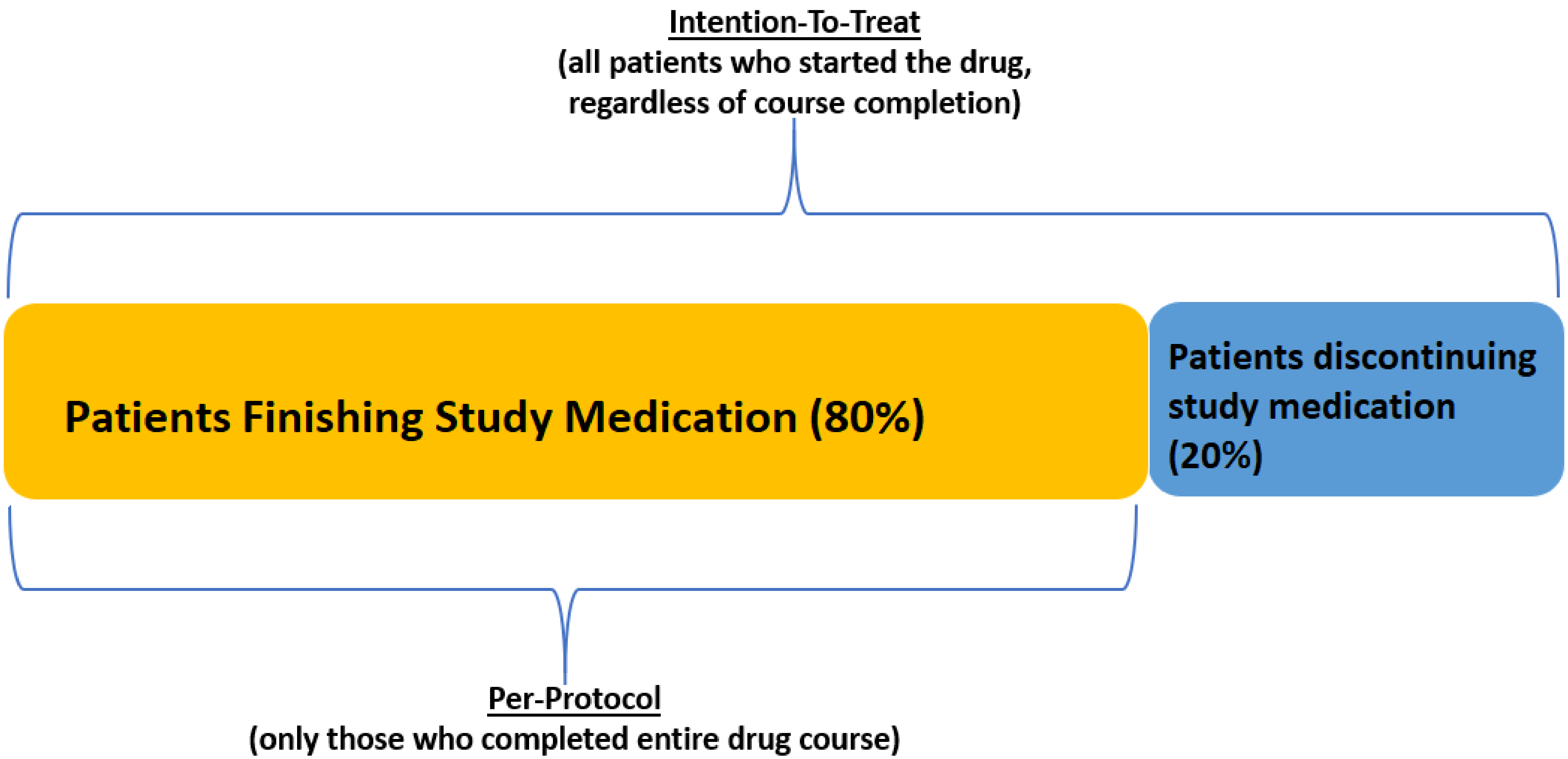
Per-Protocol:

- Comparison of groups that includes only patients who completed the treatment originally allocated.
- If done alone, this analysis leads to bias.

- Easy method for determining if study is ITT
  - Number of patients randomized = number of patients analyzed.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Intention-To-Treat vs. Per-Protocol designs



Intention-To-Treat
(all patients who started the drug, regardless of course completion)

Patients Finishing Study Medication (80%)

Patients discontinuing study medication (20%)

Per-Protocol
(only those who completed entire drug course)

Internal Validity – the Study Details: Confounding

- There are 4 reasons that can explain the relationship between what is studied (intervention) and the results from the study (outcomes):
  - **Bias**
  - **Confounding**
  - **Chance**
  - **Cause and effect (aka “truth”)**

- A study finding without bias/confounding and not due to chance is said to have “internal validity.”

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Confounding… Defined as:

Any variable within a study which, by potentially increasing variance and introducing bias, distorts the study results.

- Not technically a bias, but often referred to as one due to varying definitions of bias.
- Example:
  - If a new antidepressant is known (or believed) to decrease the risk of suicide, many prescribers will put their highest risk patients on the new antidepressant, leaving stable patients on older antidepressants. On review of their databases, investigators will note higher rates of suicide associated with the new antidepressant.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.
Boston University School of Public Health. MPH Online Learning Modules. Residual Confounding, Confounding by Indication, & Reverse Causality. Available at http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_confounding-em/BS704-EP713_Confounding-EM4.html

Confounding…

- Confounding "by indication":
  - When unblinded clinicians tailor interventions to meet the needs of specific patients (age, condition, severity of illness), thereby creating a selection bias.

- Common in observational (non-experimental) studies of drugs.

- Effective randomization and blinding will prevent this.

- Review baseline patient demographics to ensure equality.
  - Are use of non-study medications allowed in the trial?

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.
Boston University School of Public Health. MPH Online Learning Modules. Residual Confounding, Confounding by Indication, & Reverse Causality. Available at http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_confounding-em/BS704-EP713_Confounding-EM4.html

Internal Validity – the Study Details: Chance

- There are 4 reasons that can explain the relationship between what is studied (intervention) and the results from the study (outcomes):
  - **Bias**
  - **Confounding**
  - **Chance**
  - **Cause and effect (aka "truth")**

- A study finding without bias/confounding and not due to chance is said to have "internal validity."

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Chance (aka Random Error, Variation, etc.)

Defined as:

Observed outcomes *not due to intervention or bias*, rather, findings are a random accident.

- **Things that increase risk of chance findings:**
  - Small sample size
    - Smaller studies (<100 participants) are more prone to chance
  - Outcomes that are not pre-determined ("*a priori*")
  - Analyzing subgroups that are not *a priori*
  - Analyzing interim analyses of trial results
- To address for "chance," we <u>use tools</u> to verify if there is statistical significance of the findings:
  - *P-value*
  - *Confidence Intervals (CI)*

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p

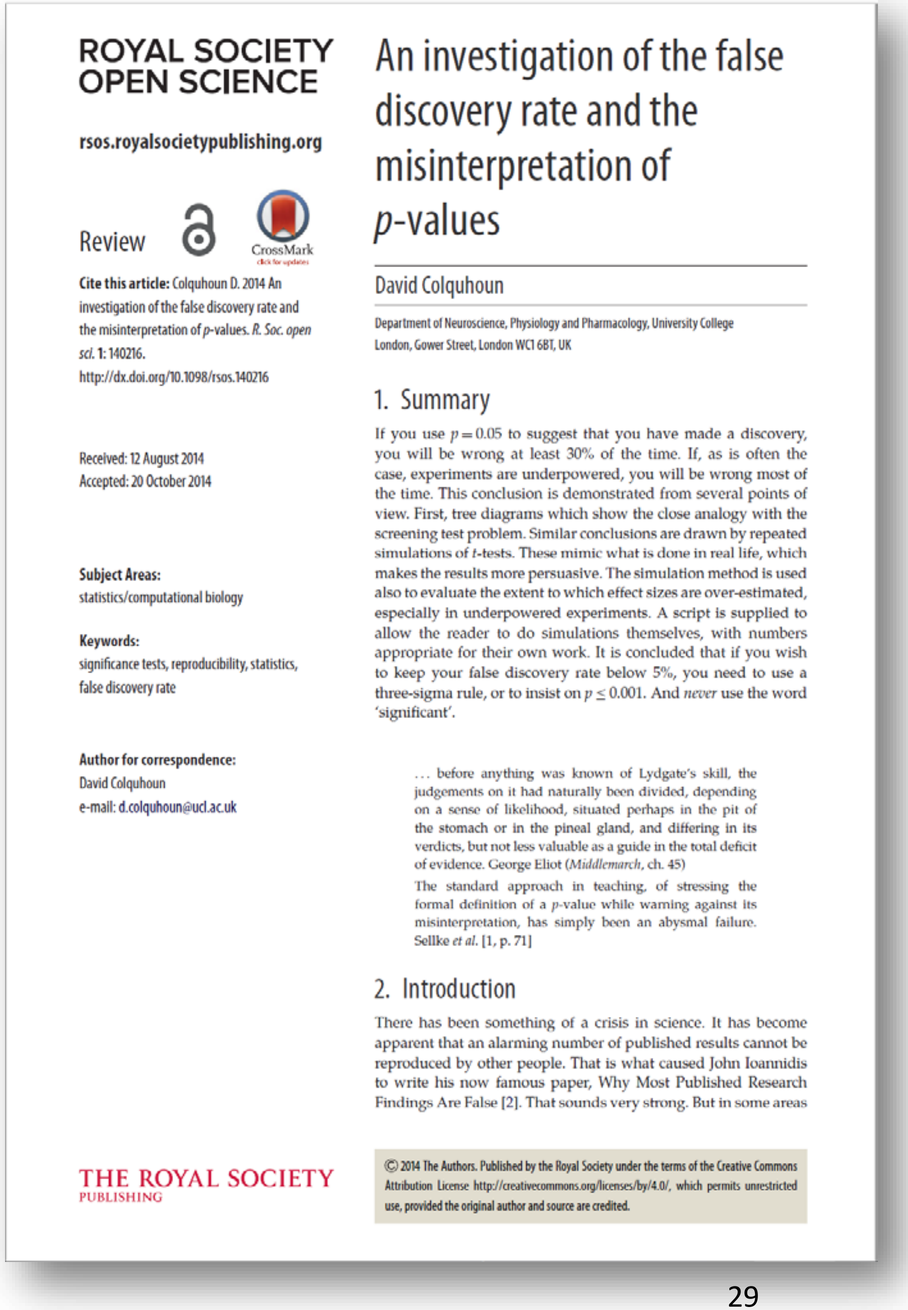
P-value(s)… (1 of 2)

- Common cutoff for determining “significance” of an outcome/finding.
    - p<0.05 (or <5%) chance of being random finding; completely arbitrary.
- Can be used as an indicator of the potential for chance effects.
- Assumes all treatments are randomized, thus cannot be used in observational studies.
- Most useful when no true difference exists between groups.
    - Less helpful than typically thought.

P-value(s)… (2 of 2)

- "If you use p=0.05 as a criterion for claiming that you have discovered an effect, you will make a fool of yourself at least 30% of the time."
- "If you want to avoid making a fool of yourself very often, do not regard anything greater than p<0.001 as a demonstration that you have discovered something."
- Delfini suggests that:

  Review of *confirmatory studies and patterns* (of similar outcomes) are potentially better methods to address the likelihood that the study results are due to chance or not.



ROYAL SOCIETY
OPEN SCIENCE

rsos.royalsocietypublishing.org

Review

Cite this article: Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of *p*-values. *R. Soc. open sci.* **1**: 140216.
http://dx.doi.org/10.1098/rsos.140216

Received: 12 August 2014
Accepted: 20 October 2014

Subject Areas:
statistics/computational biology

Keywords:
significance tests, reproducibility, statistics, false discovery rate

Author for correspondence:
David Colquhoun
e-mail: d.colquhoun@ucl.ac.uk

An investigation of the false discovery rate and the misinterpretation of *p*-values

David Colquhoun

Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1 6BT, UK

1. Summary

If you use $p = 0.05$ to suggest that you have made a discovery, you will be wrong at least 30% of the time. If, as is often the case, experiments are underpowered, you will be wrong most of the time. This conclusion is demonstrated from several points of view. First, tree diagrams which show the close analogy with the screening test problem. Similar conclusions are drawn by repeated simulations of t-tests. These mimic what is done in real life, which makes the results more persuasive. The simulation method is used also to evaluate the extent to which effect sizes are over-estimated, especially in underpowered experiments. A script is supplied to allow the reader to do simulations themselves, with numbers appropriate for their own work. It is concluded that if you wish to keep your false discovery rate below 5%, you need to use a three-sigma rule, or to insist on $p \leq 0.001$. And *never* use the word 'significant'.

> … before anything was known of Lydgate's skill, the judgements on it had naturally been divided, depending on a sense of likelihood, situated perhaps in the pit of the stomach or in the pineal gland, and differing in its verdicts, but not less valuable as a guide in the total deficit of evidence. George Eliot (*Middlemarch*, ch. 45)

> The standard approach in teaching, of stressing the formal definition of a p-value while warning against its misinterpretation, has simply been an abysmal failure. Sellke *et al.* [1, p. 71]

2. Introduction

There has been something of a crisis in science. It has become apparent that an alarming number of published results cannot be reproduced by other people. That is what caused John Ioannidis to write his now famous paper, Why Most Published Research Findings Are False [2]. That sounds very strong. But in some areas

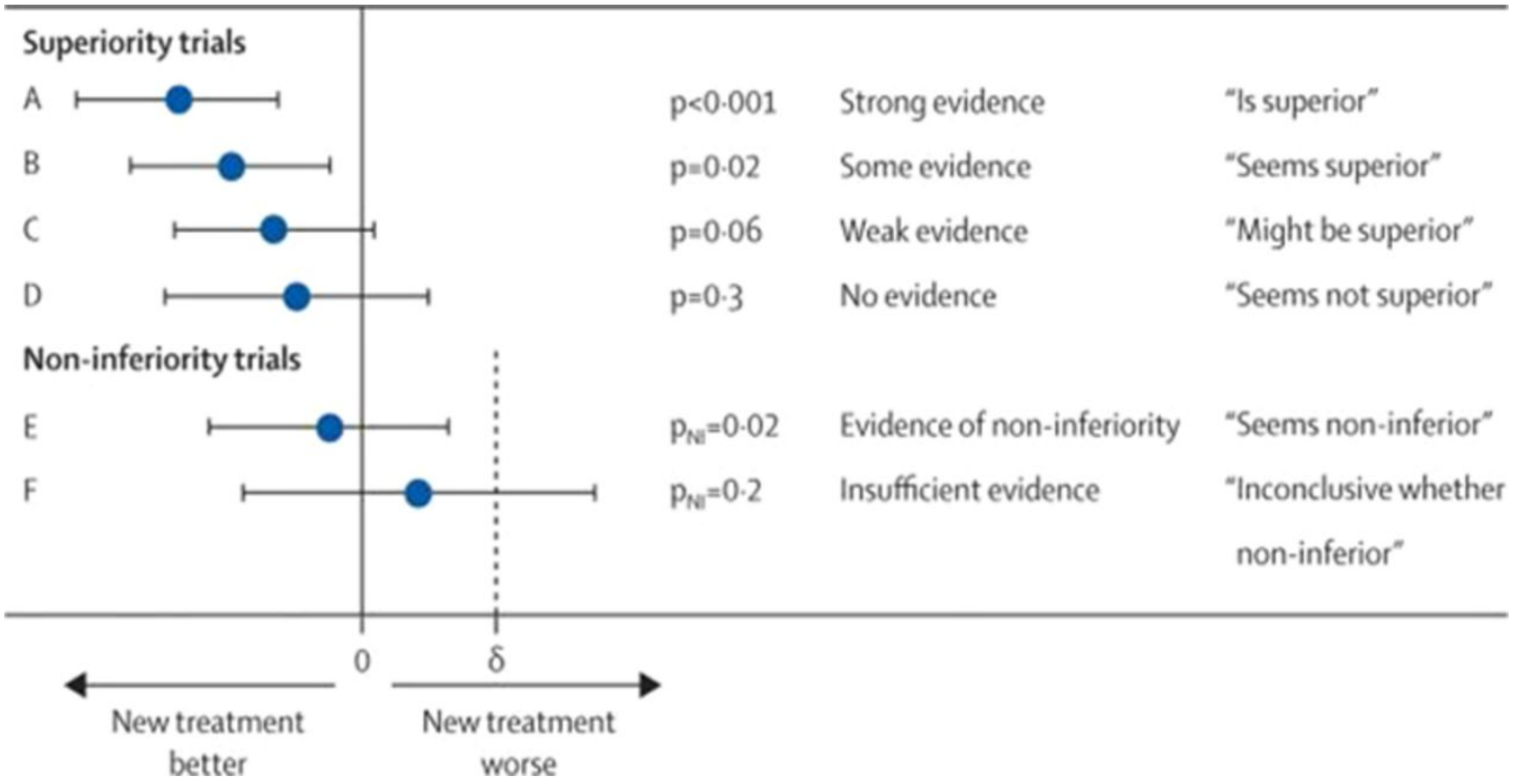
THE ROYAL SOCIETY
PUBLISHING

Confidence Intervals (CI)… (1 of 2)

- CIs are more helpful than *P*-values in evaluating study findings.
- Range of possible results that are as statistically plausible as the actual result found in the study.
- 95% CI implies a **5% chance** the **true value lies outside** the CI range.
- Narrow CIs provide greater confidence in the result (versus wide CIs).

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

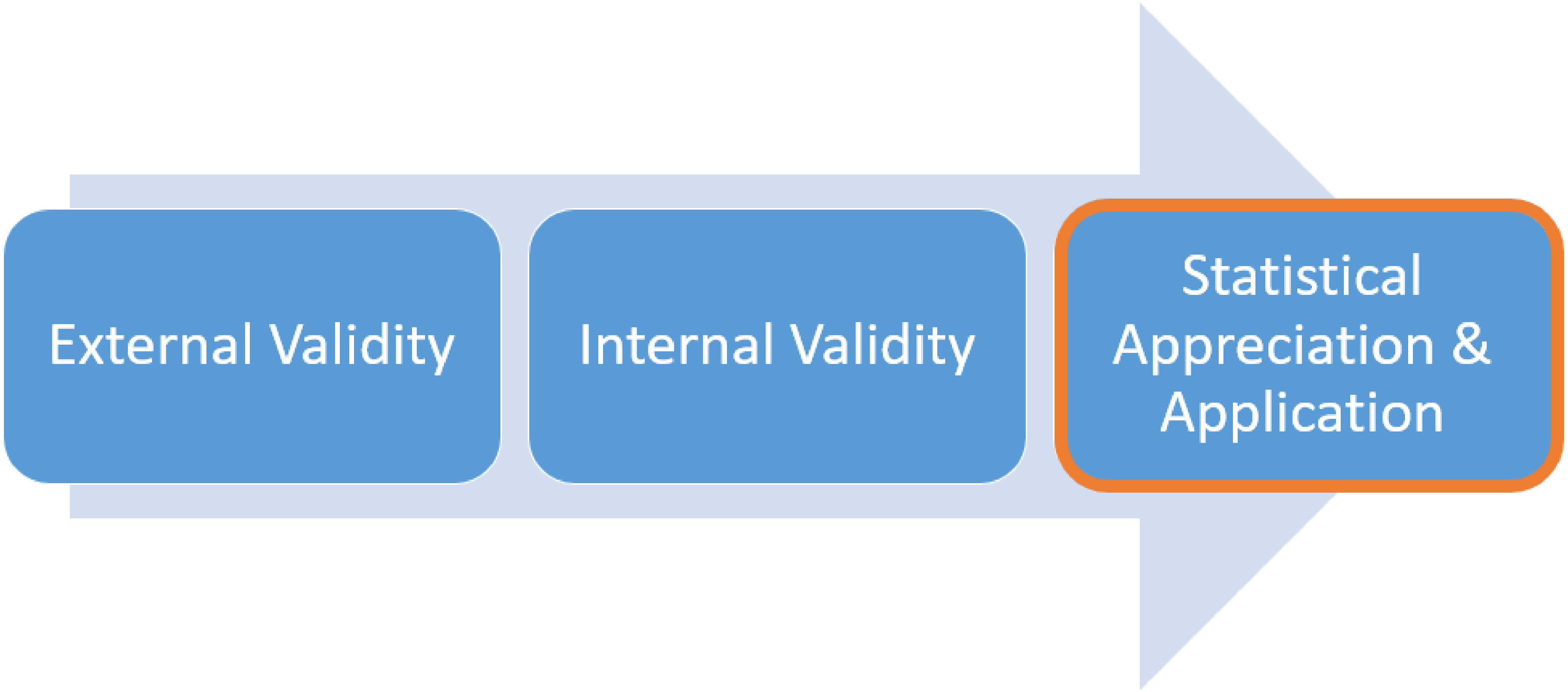
Confidence Intervals (CI)… (2 of 2)



Superiority trials			
A	p<0·001	Strong evidence	"Is superior"
B	p=0·02	Some evidence	"Seems superior"
C	p=0·06	Weak evidence	"Might be superior"
D	p=0·3	No evidence	"Seems not superior"
Non-inferiority trials			
E	p_{NI}=0·02	Evidence of non-inferiority	"Seems non-inferior"
F	p_{NI}=0·2	Insufficient evidence	"Inconclusive whether non-inferior"

http://stroke.ahajournals.org/content/strokeaha/46/8/e184/F2.large.jpg

Medical Literature Appraisal – the Process
Statistical Appreciation & Application



External Validity

Internal Validity

Statistical Appreciation & Application

Appraising the Study Results… **Effect Size**

- Outcome Measure vs. Effect size (or magnitude of difference):
  - Helps determine statistical vs. clinical significance

- 2 types of measures (of effect size) used in clinical studies:
  1. Probability
  2. Odds
- Measures of **Probability**:
  - Absolute Risk (AR)
  - Absolute Risk Reduction (ARR) / Absolute Risk Increase (ARI)
  - Number Needed to Treat (NNT) / Number Needed to Harm (NNH)
  - Relative Risk (RR) aka Risk Ratio / Relative Risk Reduction (RRR)

- Measures of **Odds:**
  - Odds Ratio (OR)
  - Hazard Ratio (HR)

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Appraising the Study Results… **Probability**

- **Absolute Risk (AR):**
  - Calculated risk of an event occurring in one comparison group.

  ***Example:** 2 groups of patients in a study have a bad outcome at different rates:
    - Control group: 15 out of 100 patients (15%) experience a bad outcome.
    - Study group: 10 out of 100 patients (10%) experience a bad outcome.

- **Absolute Risk Reduction (ARR):**
  - Difference (simple subtraction) of event rates between 2 groups.
  - ARR in this case is: **5%** [15%-10%].

What does this mean?

- 5% more people who take the study drug will avoid a bad outcome (vs. those in control group).

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Number Needed to Treat (NNT)… (1 of 2)

- Number of patients that need to be treated in order to have impact on one person.
- Reciprocal of the ARR (NNT = 1 ÷ ARR).

*__Example__: 2 groups of patients in a study have a bad outcome at different rates:

- Control group: 15 out of 100 patients (15%) experience a bad outcome.
- Study group: 10 out of 100 patients (10%) experience a bad outcome.

- ARR is: **5%** (15%-10%)
- NNT is: 1 ÷ 5% (or 0.05) = **20**

What does this mean?

- *For every 20 patients who took the study drug, 1 more patient would benefit (avoid the bad outcome) versus those in the control group, over the study duration.*

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

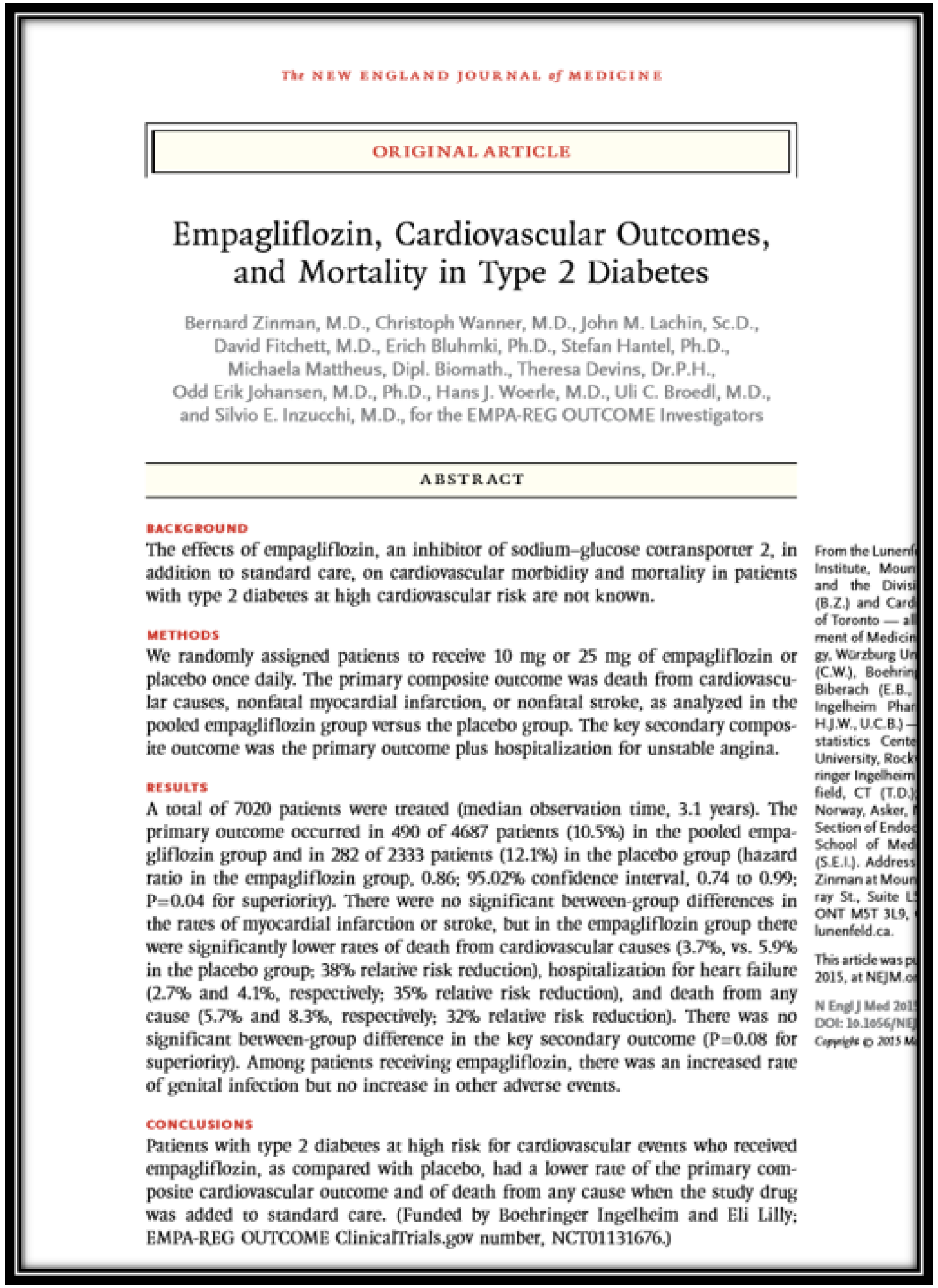
Number Needed to Treat (NNT)… (2 of 2)

- **Advantages**
  - Useful summary of trial results.
  - Useful to inform decision-making about individual patients and treatment options.
  - Relatively easy to calculate.

- **Disadvantages/Limitations**
  - NNT is based on “most probable” value in a normally distributed population.
    - Does not take into account an individual patient’s baseline risk
  - Clinical meaning is subject to interpretation
    - EXAMPLE: <u>NNT = 100 over 5 years</u> to avoid one clinical event might be seen by some as a health benefit, whereas others will consider the benefit as only moderate or even slight.
  - Time frame of given study is important; benefit of treatment is usually not linear over time
    - For example, if a treatment was conducted over a mean of 4 years, its NNT should be expressed with the same time component *(e.g., 12 patients need to be treated over about 4 years...).*

https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/nnts

Absolute Risk…



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes

Bernard Zinman, M.D., Christoph Wanner, M.D., John M. Lachin, Sc.D., David Fitchett, M.D., Erich Bluhmki, Ph.D., Stefan Hantel, Ph.D., Michaela Mattheus, Dipl. Biomath., Theresa Devins, Dr.P.H., Odd Erik Johansen, M.D., Ph.D., Hans J. Woerle, M.D., Uli C. Broedl, M.D., and Silvio E. Inzucchi, M.D., for the EMPA-REG OUTCOME Investigators

ABSTRACT

BACKGROUND

The effects of empagliflozin, an inhibitor of sodium–glucose cotransporter 2, in addition to standard care, on cardiovascular morbidity and mortality in patients with type 2 diabetes at high cardiovascular risk are not known.

METHODS

We randomly assigned patients to receive 10 mg or 25 mg of empagliflozin or placebo once daily. The primary composite outcome was death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke, as analyzed in the pooled empagliflozin group versus the placebo group. The key secondary composite outcome was the primary outcome plus hospitalization for unstable angina.

RESULTS

A total of 7020 patients were treated (median observation time, 3.1 years). The primary outcome occurred in 490 of 4687 patients (10.5%) in the pooled empagliflozin group and in 282 of 2333 patients (12.1%) in the placebo group (hazard ratio in the empagliflozin group, 0.86; 95.02% confidence interval, 0.74 to 0.99; P=0.04 for superiority). There were no significant between-group differences in the rates of myocardial infarction or stroke, but in the empagliflozin group there were significantly lower rates of death from cardiovascular causes (3.7%, vs. 5.9% in the placebo group; 38% relative risk reduction), hospitalization for heart failure (2.7% and 4.1%, respectively; 35% relative risk reduction), and death from any cause (5.7% and 8.3%, respectively; 32% relative risk reduction). There was no significant between-group difference in the key secondary outcome (P=0.08 for superiority). Among patients receiving empagliflozin, there was an increased rate of genital infection but no increase in other adverse events.

CONCLUSIONS

Patients with type 2 diabetes at high risk for cardiovascular events who received empagliflozin, as compared with placebo, had a lower rate of the primary composite cardiovascular outcome and of death from any cause when the study drug was added to standard care. (Funded by Boehringer Ingelheim and Eli Lilly; EMPA-REG OUTCOME ClinicalTrials.gov number, NCT01131676.)

From the Lunenf
Institute, Moun
and the Divisi
(B.Z.) and Card
of Toronto — al
ment of Medicin
gy, Würzburg Un
(C.W.), Boehrin
Biberach (E.B.,
Ingelheim Phar
H.J.W., U.C.B.) —
statistics Cente
University, Rock
ringer Ingelheim
field, CT (T.D.)
Norway, Asker,
Section of Endoc
School of Med
(S.E.I.). Address
Zinman at Moun
ray St., Suite L
ONT M5T 3L9,
lunenfeld.ca.

This article was pu
2015, at NEJM.o

N Engl J Med 201
DOI: 10.1056/NE
Copyright © 2015 M

- **Absolute Risk** (study) = 10.5%
- **Absolute Risk** (control) = 12.1%

  - ARR = [12.1%-10.5%] = 1.6%
  - **NNT** = (1 ÷ ARR) = (1 ÷ 0.016) = 63

- So… 63 patients need to be treated with empagliflozin *(for 3.1 years)* to avoid the primary outcome in 1 patient.

- *Is this good? What if the primary outcome was ER admissions?*

Relative Risk (RR)…

Estimate of risk of an event when compared (or relative) to >1 group.

***Example:** 2 groups of patients in a study have a bad outcome at different rates:

- Control group: 15 out of 100 patients (15%) experience a bad outcome.
- Study group: 10 out of 100 patients (10%) experience a bad outcome.

- Risk for control = 15%
- Risk for study = 10%
- **Relative Risk** is: 10% ÷ 15% = **0.67** (or 67%)

- *RR of <1.0 represents a decrease in risk than comparison group; >1.0 means an increase in risk.*

What does this mean?

- *Patients in the study group have a reduced risk of 67% (vs. those in the control group).*

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Relative Risk Reduction (RRR)…

Difference in event rates between 2 groups, expressed as a proportion of the event rate in the untreated group.

- Calculation: 1-RR

***Example:** 2 groups of patients in a study have a bad outcome at different rates:

- Control group: 15 out of 100 patients (15%) experience a bad outcome.
- Study group: 10 out of 100 patients (10%) experience a bad outcome.

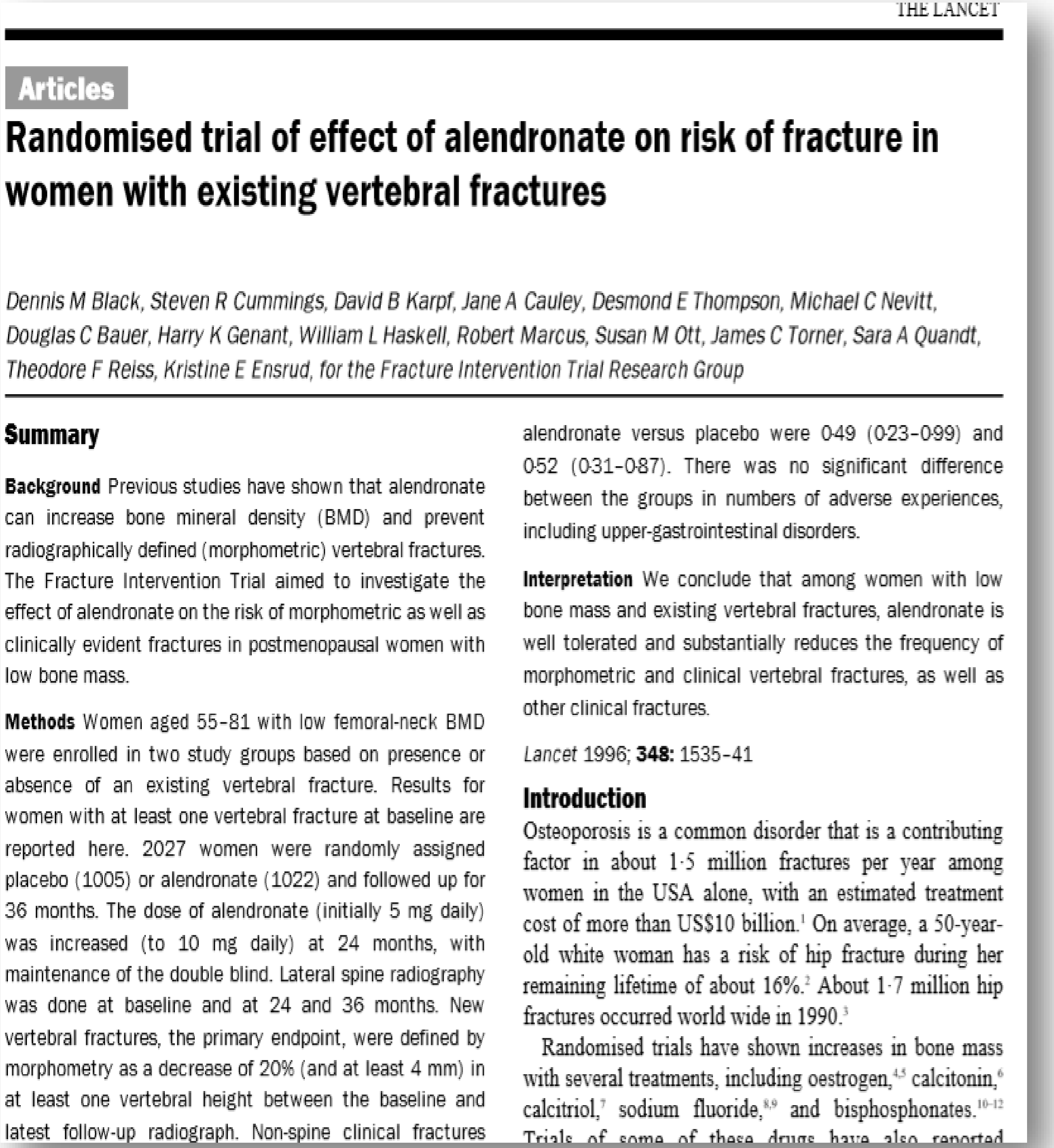
- Risk (control) = 15%
- Risk (study) = 10%
- RR for this example is: 10% ÷ 15% = 67%
- **Relative Risk Reduction** = (1-RR) or (1.0 - 0.67) = 0.33 or **33%**

What does this mean?

- *Patients in the study group had a relative 33% reduction in risk (vs. those in the control group).*

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

RRR: Example...



Articles

Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures

Dennis M Black, Steven R Cummings, David B Karpf, Jane A Cauley, Desmond E Thompson, Michael C Nevitt, Douglas C Bauer, Harry K Genant, William L Haskell, Robert Marcus, Susan M Ott, James C Torner, Sara A Quandt, Theodore F Reiss, Kristine E Ensrud, for the Fracture Intervention Trial Research Group

Summary

Background Previous studies have shown that alendronate can increase bone mineral density (BMD) and prevent radiographically defined (morphometric) vertebral fractures. The Fracture Intervention Trial aimed to investigate the effect of alendronate on the risk of morphometric as well as clinically evident fractures in postmenopausal women with low bone mass.

Methods Women aged 55–81 with low femoral-neck BMD were enrolled in two study groups based on presence or absence of an existing vertebral fracture. Results for women with at least one vertebral fracture at baseline are reported here. 2027 women were randomly assigned placebo (1005) or alendronate (1022) and followed up for 36 months. The dose of alendronate (initially 5 mg daily) was increased (to 10 mg daily) at 24 months, with maintenance of the double blind. Lateral spine radiography was done at baseline and at 24 and 36 months. New vertebral fractures, the primary endpoint, were defined by morphometry as a decrease of 20% (and at least 4 mm) in at least one vertebral height between the baseline and latest follow-up radiograph. Non-spine clinical fractures alendronate versus placebo were 0·49 (0·23–0·99) and 0·52 (0·31–0·87). There was no significant difference between the groups in numbers of adverse experiences, including upper-gastrointestinal disorders.

Interpretation We conclude that among women with low bone mass and existing vertebral fractures, alendronate is well tolerated and substantially reduces the frequency of morphometric and clinical vertebral fractures, as well as other clinical fractures.

Lancet 1996; **348:** 1535–41

Introduction

Osteoporosis is a common disorder that is a contributing factor in about 1·5 million fractures per year among women in the USA alone, with an estimated treatment cost of more than US$10 billion.[1] On average, a 50-year-old white woman has a risk of hip fracture during her remaining lifetime of about 16%.[2] About 1·7 million hip fractures occurred world wide in 1990.[3]

Randomised trials have shown increases in bone mass with several treatments, including oestrogen,[4,5] calcitonin,[6] calcitriol,[7] sodium fluoride,[8,9] and bisphosphonates.[10–12] Trials of some of these drugs have also reported

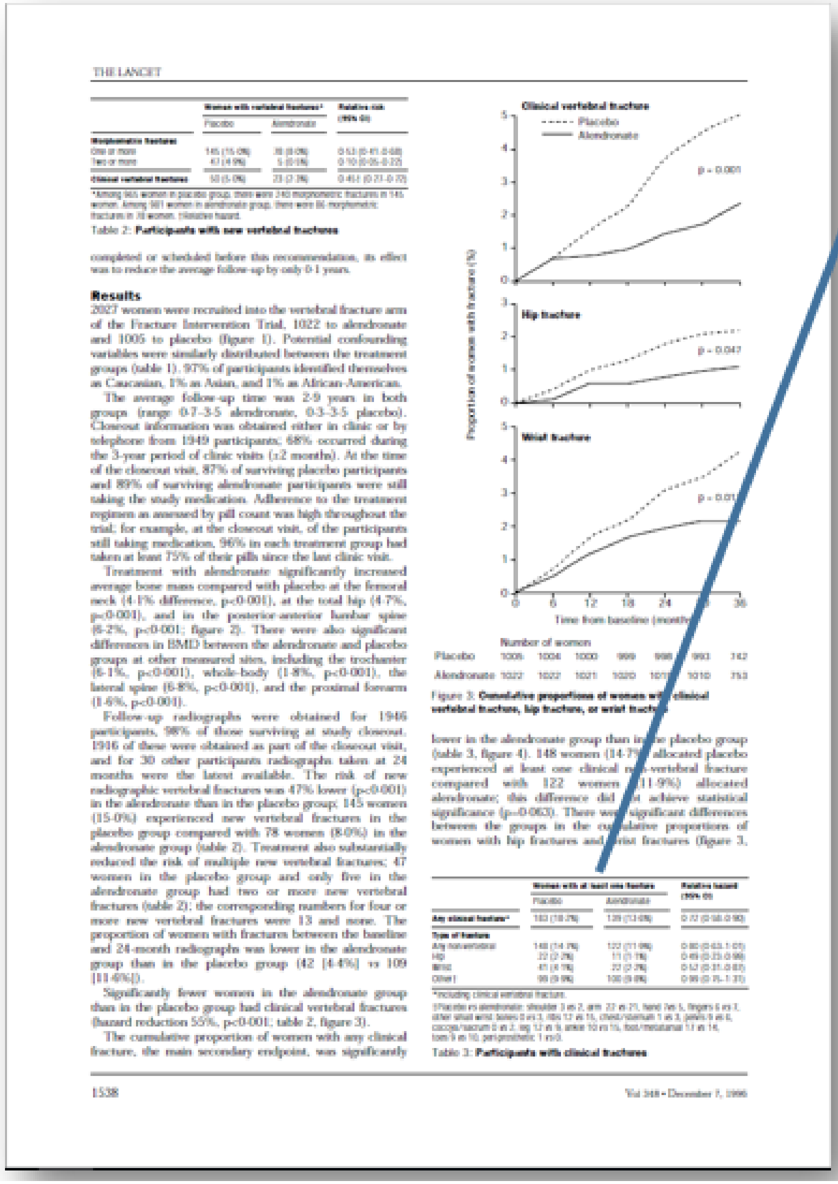
EXAMPLE - Relative Risk Reduction
Lancet 1996; 348: 1535-1541.

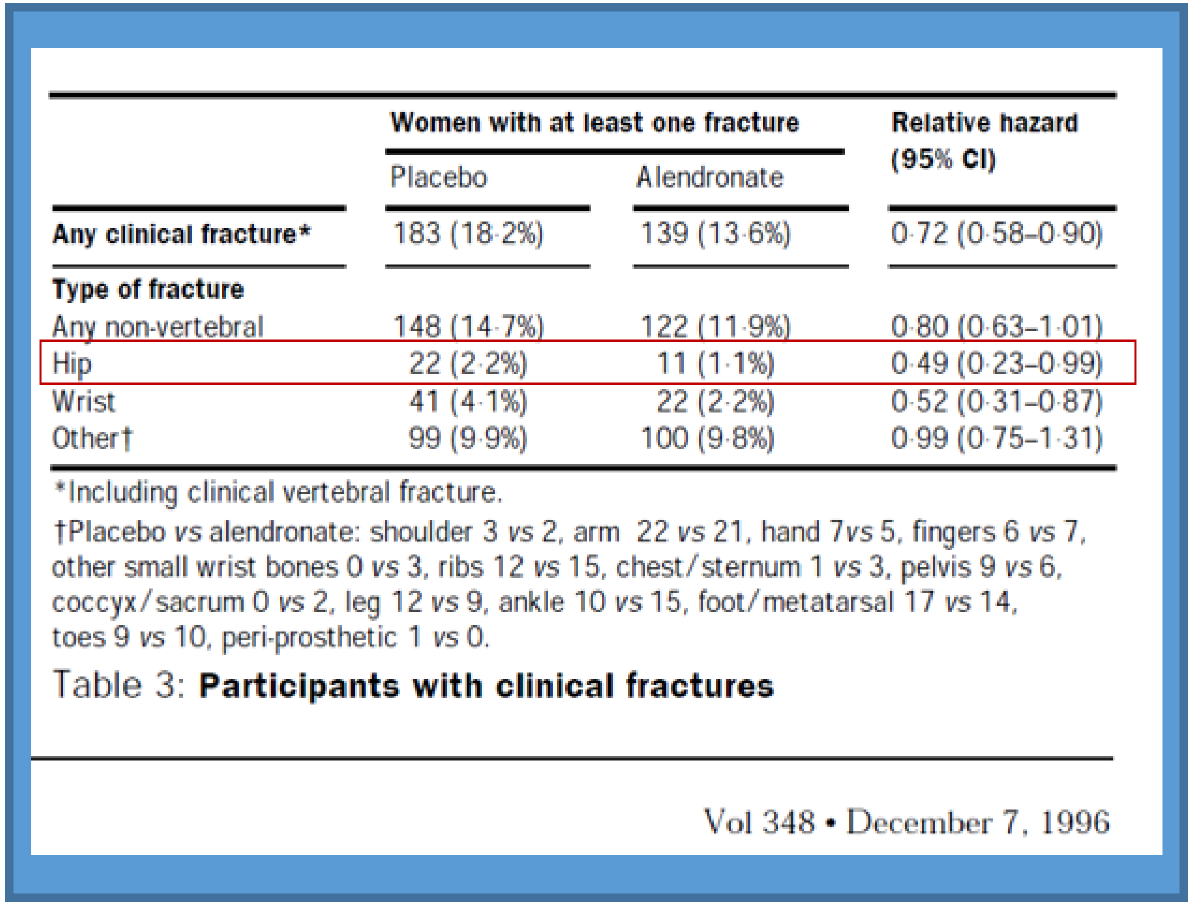
Primary endpoint:

- *New vertebral fractures,* defined by morphometry as a decrease of 20% (and at least 4 mm) in at least one vertebral height between the baseline and the latest follow-up radiography.

- *"Fosamax reduces hip fractures by 50%."*

RRR Example from Alendronate Study





	Women with at least one fracture		Relative hazard (95% CI)
	Placebo	Alendronate	
Any clinical fracture*	183 (18·2%)	139 (13·6%)	0·72 (0·58–0·90)
Type of fracture			
Any non-vertebral	148 (14·7%)	122 (11·9%)	0·80 (0·63–1·01)
Hip	22 (2·2%)	11 (1·1%)	0·49 (0·23–0·99)
Wrist	41 (4·1%)	22 (2·2%)	0·52 (0·31–0·87)
Other†	99 (9·9%)	100 (9·8%)	0·99 (0·75–1·31)

*Including clinical vertebral fracture.

†Placebo *vs* alendronate: shoulder 3 *vs* 2, arm 22 *vs* 21, hand 7*vs* 5, fingers 6 *vs* 7, other small wrist bones 0 *vs* 3, ribs 12 *vs* 15, chest/sternum 1 *vs* 3, pelvis 9 *vs* 6, coccyx/sacrum 0 *vs* 2, leg 12 *vs* 9, ankle 10 *vs* 15, foot/metatarsal 17 *vs* 14, toes 9 *vs* 10, peri-prosthetic 1 *vs* 0.

Table 3: **Participants with clinical fractures**

Vol 348 • December 7, 1996

- RR = 1.1% ÷ 2.2% = 0.50
- RRR = 1 – RR = (1 – 0.50) = 0.50 or 50%
- **ARR = 2.2% – 1.1% = 1.1%**
- NNT = 1 ÷ ARR or 1 ÷ 0.011 = 91

Odds Ratio (OR)…

Odds represent likelihood of event occurring vs. not occurring:

- Similar to probability, especially if event rate (incidence) is low (e.g., 10%).
- Tends to overestimate risk as the incidence increases (RR does not).

OR used in prospective or retrospective studies; RR only in prospective studies.

***Example:** (case-control study)

Control group: 20 out of 100 patients die
Study group: 10 out of 100 patients die

- Odds of death in control group = 20/80 (25%)
- Odds of death in study group = 10/90 (11%)
- Odds ratio = 0.25 ÷ 0.11 = **2.27**

What does this mean?

- *The odds of dying in the control group are >2 times that in the study group*

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Hazard Ratio (HR)…

- Used in time-to-event studies (survival/death) where rates of a hazard are determined and applied to a hazard curve or slope.
- Approximates the relative risk in intervention group vs. control group in a Kaplan-Meier curve or other time-to-event model.

Calculated similarly to ORs:

- Chance of an event occurring in treatment arm
- Chance of the event occurring in the control arm

= Hazard Rate (slope of the survival curve)

Example:
If the HR is 2, a patient who has not yet experienced an event has twice the chance of experiencing the event at the next point in time.

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

Appraising the Study Results:
Author’s Conclusions

Some experts suggest to avoid reading the study “Discussion” altogether; focus on methodology and results & draw your own conclusions.

- Conclusions are generally opinions; offer speculation and conjecture.
- Must assume some degree of bias (rooting for the intervention).

Strite S, Stuart M. Basics for evaluating medical research studies. 1st edition. United States: Delfini Group Publishing; 2013. 112 p.

In Summary… Review Roadmap

1. **Is the study applicable to your patients and practice?**
   - *Inclusion/exclusion criteria, practice setting, meaningful outcomes?*
2. **Is the study an observation or an experiment?**
3. **Can you identify bias(es) or confounders in the study?**
   - *Prospective? Randomized? Controlled? Blinding? Significant Attrition?*
4. **Are the results significant? (statistically and/or clinically)**
   - *95% CI expectations met? Do AR, ARR and NNT indicate benefit?*
5. **Is this similar to other reported findings?**
   - *Do "real world evidence"/post-marketing reports support this?*

Paul Glasziou, MD

Professor of Evidence-Based Medicine, University of Oxford

- *“The search engine is now as essential as the stethoscope.”*

- *“…a 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system.”*

Glasziou P, Burls A, Gilbert R. Evidence based medicine and the medical curriculum. BMJ 2008; 337:a1253.