Study Design, Power Analysis and Sample Size

Jerry Jarvik, MD MPH
Departments of Radiology, Neurological Surgery, Health Services, Comparative Effectiveness, Cost and Outcomes Research Center, University of Washington

Acknowledgements

- NIH: 1P30 AR072572-01; UH2 AT007766-01; UH3 AT007766
- PCORI: CE-12-11-4469

Disclosures (Jarvik)

- UpToDate
  - Section Editor
- Evidence-Based Neuroimaging Diagnosis and Treatment (Springer)
  - Co-Editor

Talk Outline

- Study design
- Power analysis
- Sample size

“Don’t” #1

Starting out, don’t try this without help

PROTO-PROFESSOR ALGARTH ZAG,
PIONEER IN FIRE RESEARCH.

Do #1

- Find somebody who knows what they are doing and work with them
  - –Epidemiologist
  - –Biostatistician
  - –Health services researcher
  - –Experienced researcher

“Do” #2

Read

“Designing Clinical Research” by Hulley, Cummings et al

Other Useful References

- Evidence Based Radiology: http://www.evidencebasedradiology.net/index.html
- JAMA: User's guide to medical literature
- Radiology: Statistical Concepts Series
- AJR: Fundamentals of Clinical Research for Radiologists
- Evidence Based Medicine- Sackett et al
- Center for Evidence Based Medicine: www.cebm.net
- EBM online: ebm.bmjjournals.com
- Blackmore and Medina- Evidence Based Medicine

Radiology

Statistical Concepts Series

John Eng, MD

Index terms:
Radiology and radiologists, research
Receiver operating characteristic (ROC) curve
Statistical analysis

Published online

Sample Size Estimation: A Glimpse beyond Simple Formulas[1]

Small increments in the complexity of clinical studies can readily take sample size estimation and statistical power analysis beyond the capabilities of simple math

Talk Outline

- Study design
- Power analysis
- Sample size

So Many Choices

Observational vs Experiment

Observational

- Case series
- Cross-sectional
- Case-control
- Cohort

Experimental

- Non-random allocation
- RCT

Timing

- Prospective
- Retrospective

Goal

- Descriptive
- Analytic

Goal

- Explanatory
- Pragmatic

Case Series

- Observational
- Retrospective
- Descriptive, not analytic
- Hypothesis generating
- No comparison group

Cross-Sectional

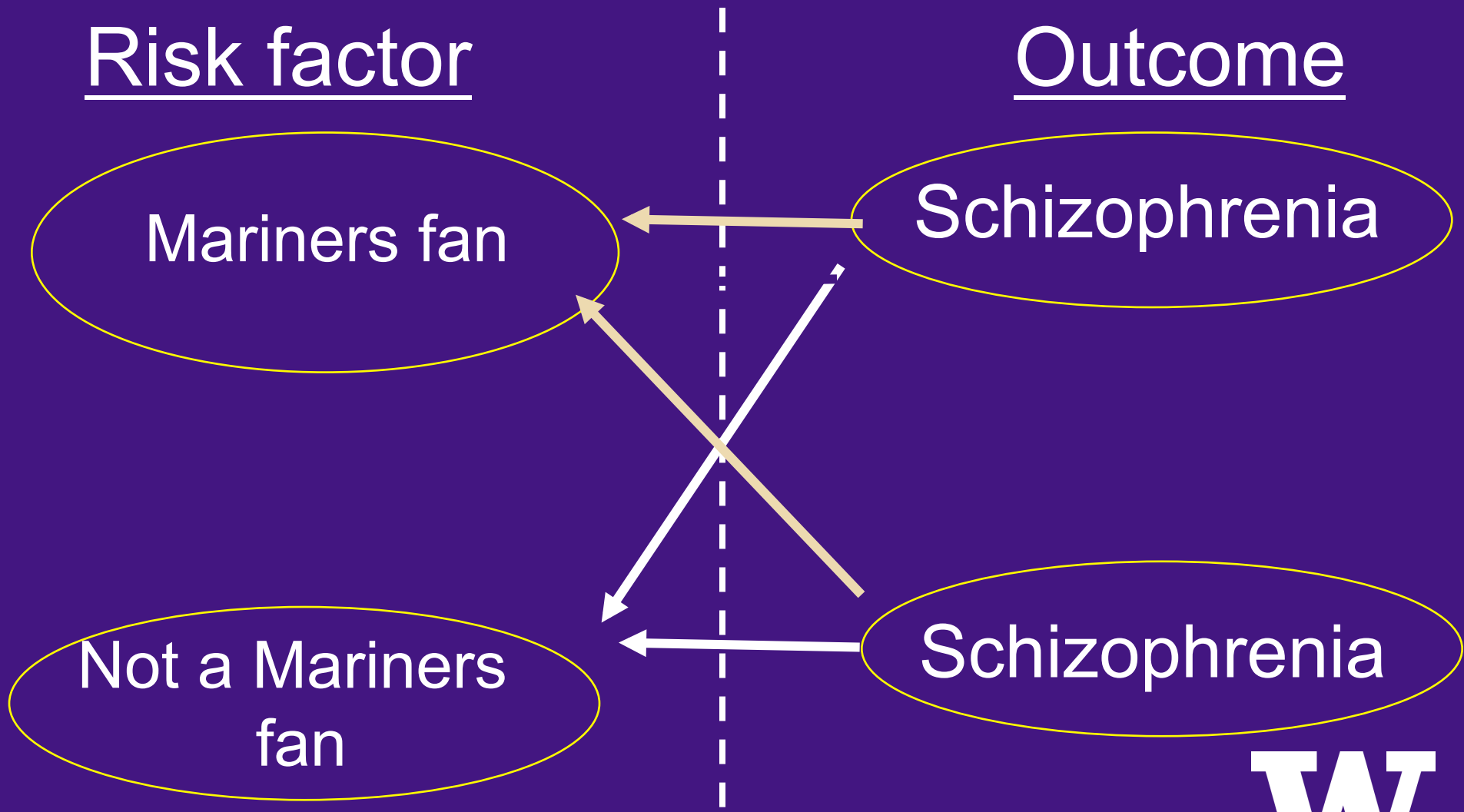
- Group examined at 1 point in time
- Observational
- Retrospective/prospective
- Usually descriptive
- Usually the design of diagnostic accuracy studies

Case-Control

- 2 groups defined by outcome
- Observational
- Retrospective- outcomes need to have occurred
- Descriptive or analytic
- Good for rare outcomes

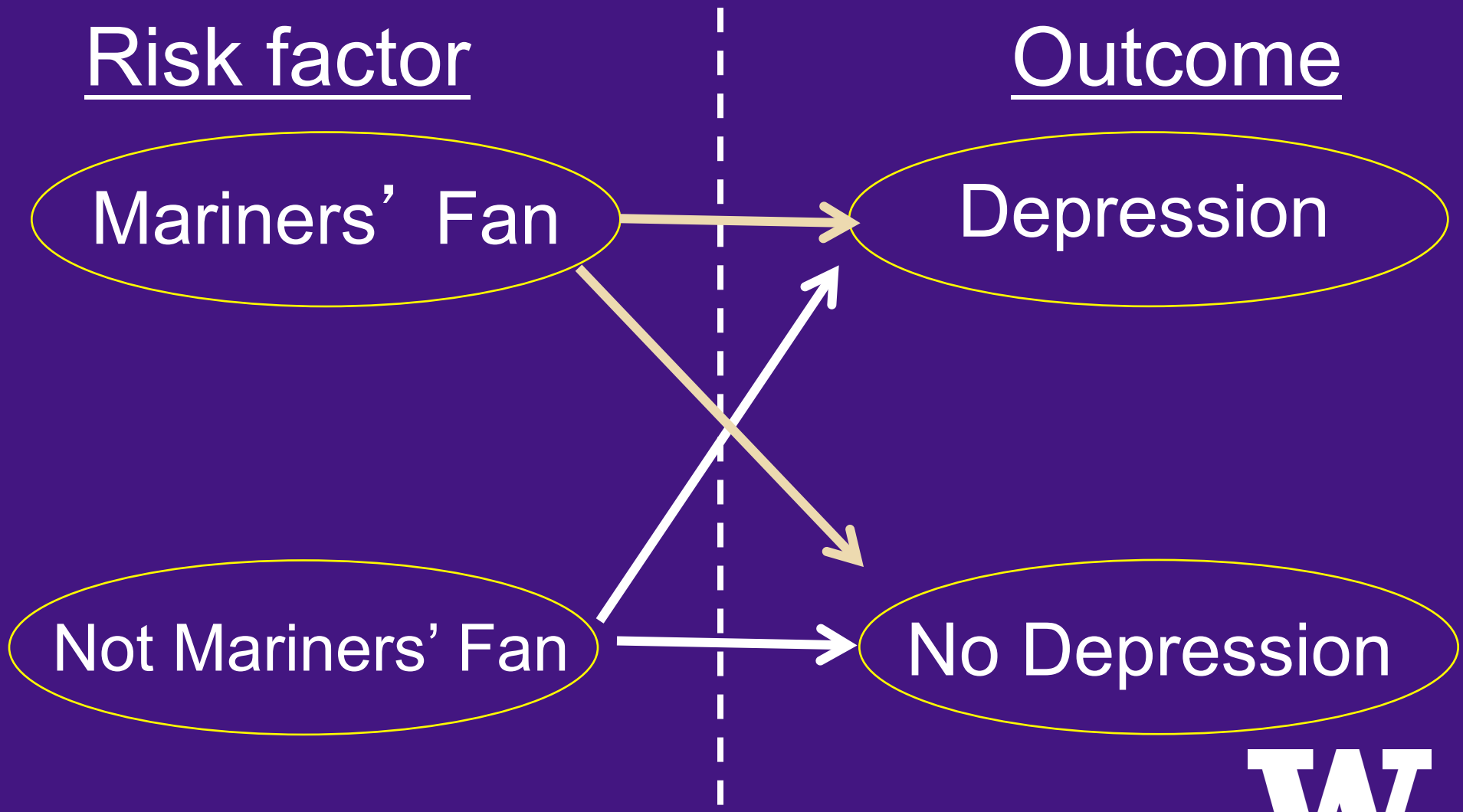
Study Question 1: Is being a Mariners fan associated with schizophrenia?

Case Control



Study Question 2: Is being a Mariners' fan associated with depression?

Cohort Study



RCT

- Experiment
- Groups created randomly
- Prospective
- Analytic
- Best design for eliminating bias

RCT

- Only design that controls for unknown biases
- Cohort and case-control designs can control for known biases

Impact of Design on Policy

- Many entities using GRADE (*Grading of Recommendations, Assessment, Development and Evaluation)*
- Developed by Guyatt et al.

GRADE: Quality of Evidence
Putting it All Together

Step 1 Starting grade based on study design	Step 2 Reduce grade	Step 3 Raise grade High-quality observational studies	Step 4 Final grade
RCT - High	***Study quality (risk of bias)*** Serious (-1) or very serious (-2) limitations ***Inconsistency*** Important inconsistency (-1) ***Indirectness*** Some (-1) or major (-2) uncertainty about directness ***Imprecision*** Imprecise or sparse data (-1) ***Publication bias*** High suspicion (-1)	***Large magnitude of effect*** Large effect (+1) RR > 2 or < 0.5, based on consistent evidence from two or more observational studies with no plausible confounders Very large effect (+2) RR> 5 or < 0.2, based on direct evidence with no major threats to validity ***Dose response gradient*** (+1) ***All plausible confounders would have reduced the effect*** (+1)	**High** Further research unlikely to change confidence in the estimate of effect
Observational – Low Quasi-RCT Cohort Case-control			**Moderate** Further research likely to have an important impact on confidence in the estimate of effect and may change the estimate
All others – Very Low Case reports Case series			**Low** Further research very likely to have an important impact on confidence in the estimate and may change the estimate
			Very Low Any estimate of effect is very uncertain

Step 1

Starting grade based on study design

RCT - High

Observational – Low

- Quasi-RCT
- Cohort
- Case-control

All others – Very Low

- Case reports
- Case series

GRADE and Dx Tests

- Cross sectional or cohort studies can provide high quality evidence of test accuracy
- However, test accuracy is a surrogate for patient-important outcomes, so such studies often provide low quality evidence about diagnostic tests

Pragmatic vs. Explanatory

- PCTs determine whether an intervention works in “real world” conditions and not to answer how or why

Pragmatic Trials

- **Patients**: selection reflects routine practice
- **Intervention**: useful in current practice
- **Comparator**: routine practice (vs. placebo)
- **Outcomes**: patient centered
- **Timing:** timescales important to stakeholders
- **Setting**: where everyday care happens, (community clinics, hospitals, and health systems) reflecting **routine clinical practice**

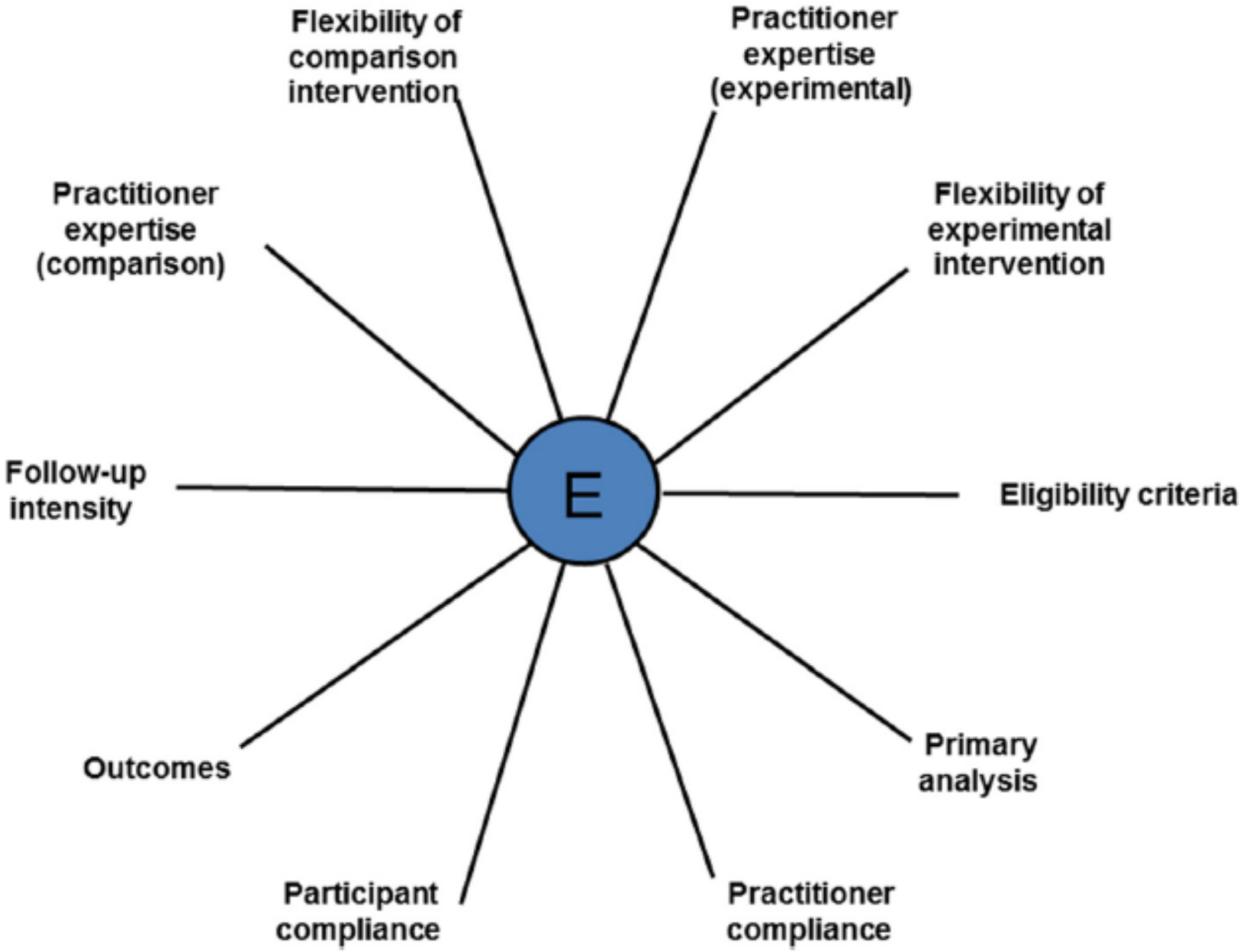
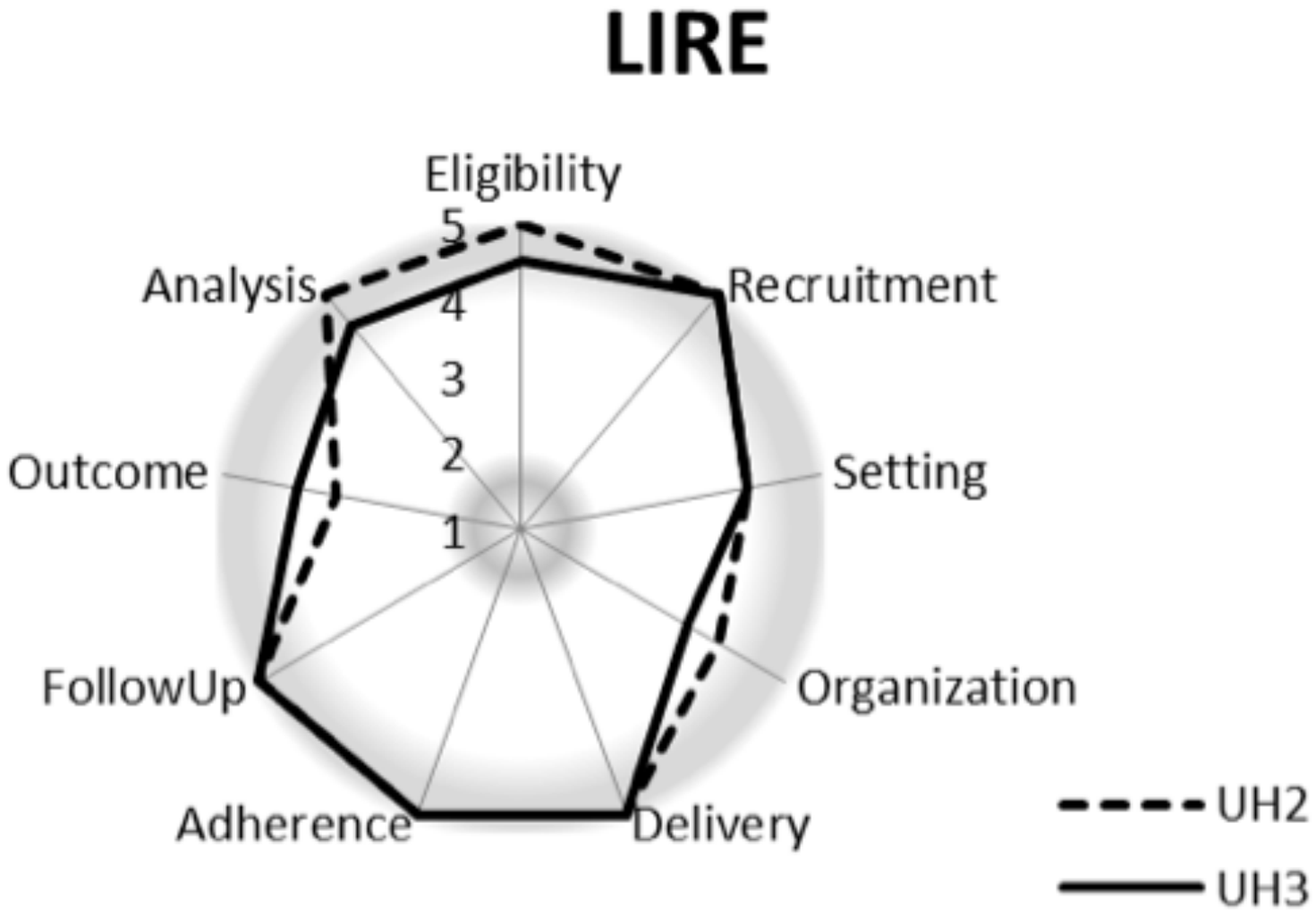


Figure 1 Pragmatic Explanatory Continuum Indicator Summary (PRECIS) [10].

PRECIS Diagram



Why Pragmatic Trials Are Important

Talk Outline

- Study design
- Power analysis
- Sample size

Sample Size & Power Comparing Means

$$N = 4\sigma^2(z_{crit+}z_{pwr})^2/D^2$$

N= total sample size

σ= standard deviation

z= constants given signif level and desired power

D= smallest meaningful difference between groups (effect size)

↑'s sample size needed or if sample fixed, ↓'s Power

- Smaller effect size
- Larger SD $N = 4\sigma^2(z_{crit} + z_{pwr})^2 / D^2$
- Need for greater significance

↑'s sample size needed or if sample fixed, ↓'s Power

- Smaller effect size
- Larger SD $N = 4\sigma^2(z_{crit+}z_{pwr})^2/D^2$
- Need for greater significance

↑'s sample size needed or if sample fixed, ↓'s Power

- Smaller effect size
- Larger SD
- Need for greater significance

$$N = 4\sigma^2(z_{crit} + z_{pwr})^2 / D^2$$

↑'s sample size needed or if sample fixed, ↓'s Power

- Smaller effect size
- Larger SD $N = 4\sigma^2(z_{crit} + z_{pwr})^2 / D^2$
- Need for greater significance

Data Driven Impact on Sample Size and Power

- Data variability: the more variable (↑SD), the larger sample size
- Clinically meaningful diff: the smaller this difference, the larger the sample size needed

$$N = 4\sigma^2(z_{crit} + z_{pwr})^2 / D^2$$

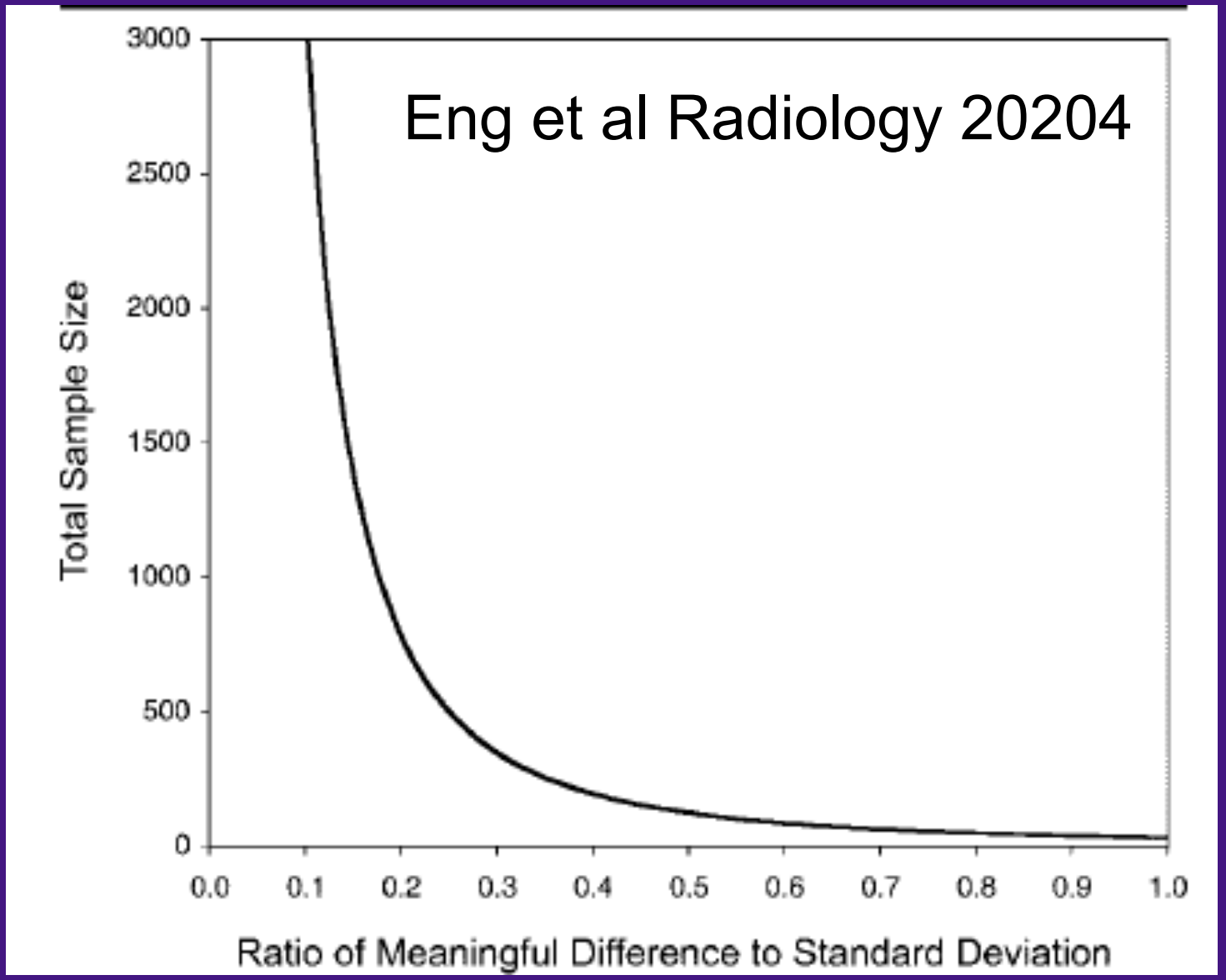
Effect Size Impact Example: Roland-Morris Disability Questionnaire (RDQ)

- 0-24 point back-related disability
- Clinically meaningful diff 2-4 pts
- Impact on sample size of 2 vs. 4 point diff between groups: $D^2=4$ vs. 16 (4 fold)

SD Impact Example: RDQ

- All older adults (BOLD): SD=6
- Adults w/osteoporotic fxs (INVEST): SD=4
- $4\sigma^2 = 144$ vs. 64

Ratio of Meaningful Diff to SD



Power and Sample Size

- Big variability= bad
- Big difference= good

Alternative to Formulas

- Simulations
- That's why God made biostatisticians…

Talk Outline

- Study design
- Power analysis
- Sample size- practical considerations

Sample Size

- Often this is fixed by budget considerations
- Craft question to fit within sample size
  - –Small variability
  - –Big difference
  - –Low drop-outs

Drop-Outs

- Potential for bias
- Hurts power
- “Reasonable” drop-out rate ~15%

Strategies for Minimizing Sample Size and Maximizing Power

- Continuous rather than dichot outcomes
- Paired rather than unpaired measurements
- Outcomes with less variability
- Questions with likely large diff between groups
- Outcome closer to 50% rather than 0% or 100%

Other Useful Resources

- JAMA: User's guide to medical literature
- Radiology: Statistical Concepts Series
- AJR: Fundamentals of Clinical Research for Radiologists
- Evidence Based Medicine- Sackett et al
- Center for Evidence Based Medicine: www.cebm.net
- Evidence Based Radiology: http://www.evidencebasedradiology.net/index.html
- EBM online: ebm.bmjjournals.com
- Blackmore and Medina- Evidence Based Radiology

Some GRADE References

- ***GRADE**: an emerging consensus on rating quality of evidence and strength of recommendations.* **Guyatt** GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; **GRADE** Working Group. **BMJ**. **2008** Apr 26;336(7650):924-6.
- *Grading quality of evidence and strength of recommendations for diagnostic tests and strategies.* Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, Williams JW Jr, Kunz R, Craig J, Montori VM, Bossuyt P, **Guyatt** GH; **GRADE** Working Group. **BMJ**. **2008** May 17;336(7653):1106-10


LAB INITIO
by Nick D Kim
THIS STUDY SUGGESTS THAT PESSIMISTS ARE A WHOLE LOT BETTER AT JUDGING THE TRUE STATE OF AFFAIRS THAN OPTIMISTS...
YEAH, BUT I BET THEY USED A SLOPPY EXPERIMENTAL PROTOCOL. - I DOUBT THAT THE CONCLUSIONS HAVE ANY VALIDITY...