



An Introduction to Meta-Analysis

Sandy Hale & Mike Strube



Two Goals and the “Level Problem”

Meta-Analysis and Other Forms of Analysis

Primary analysis: The analysis of data from a single study to test the hypotheses originally formulated.

Meta-Analysis and Other Forms of Analysis

Secondary analysis: The re-analysis of data from a single study to test new hypotheses or to apply more appropriate statistical procedures to test the original hypotheses.

Meta-Analysis and Other Forms of Analysis

Meta-analysis: The application of statistical procedures to examine tests of a common hypothesis from more than one study.

“The analysis of analyses”

A Basic Definition

Meta-analysis attempts to apply to a collection of studies the same methodological rigor and statistical precision ordinarily found in primary research.

In a meta-analysis, the collection of studies test the same conceptual hypothesis, but may do so using a wide variety of methods, measures, sample, and settings.

The challenge that meta-analysis answers is to provide a way to combine the seemingly disparate studies to provide a convincing overall test of the hypothesis and to explore its moderators.

A Brief History

Although commonly viewed as a relatively modern advancement, the basic elements of meta-analysis can be traced back to R. A. Fisher, who developed an early quantitative procedure for combining the probabilities from multiple hypothesis tests.

A Brief History

A common problem in many laboratories is that multiple studies are conducted on the same hypothesis and some way of “combining” the studies is sought. Often each study provides weak evidence but there may appear to be some consistency in findings across studies. A way of capturing that consistency is desired.

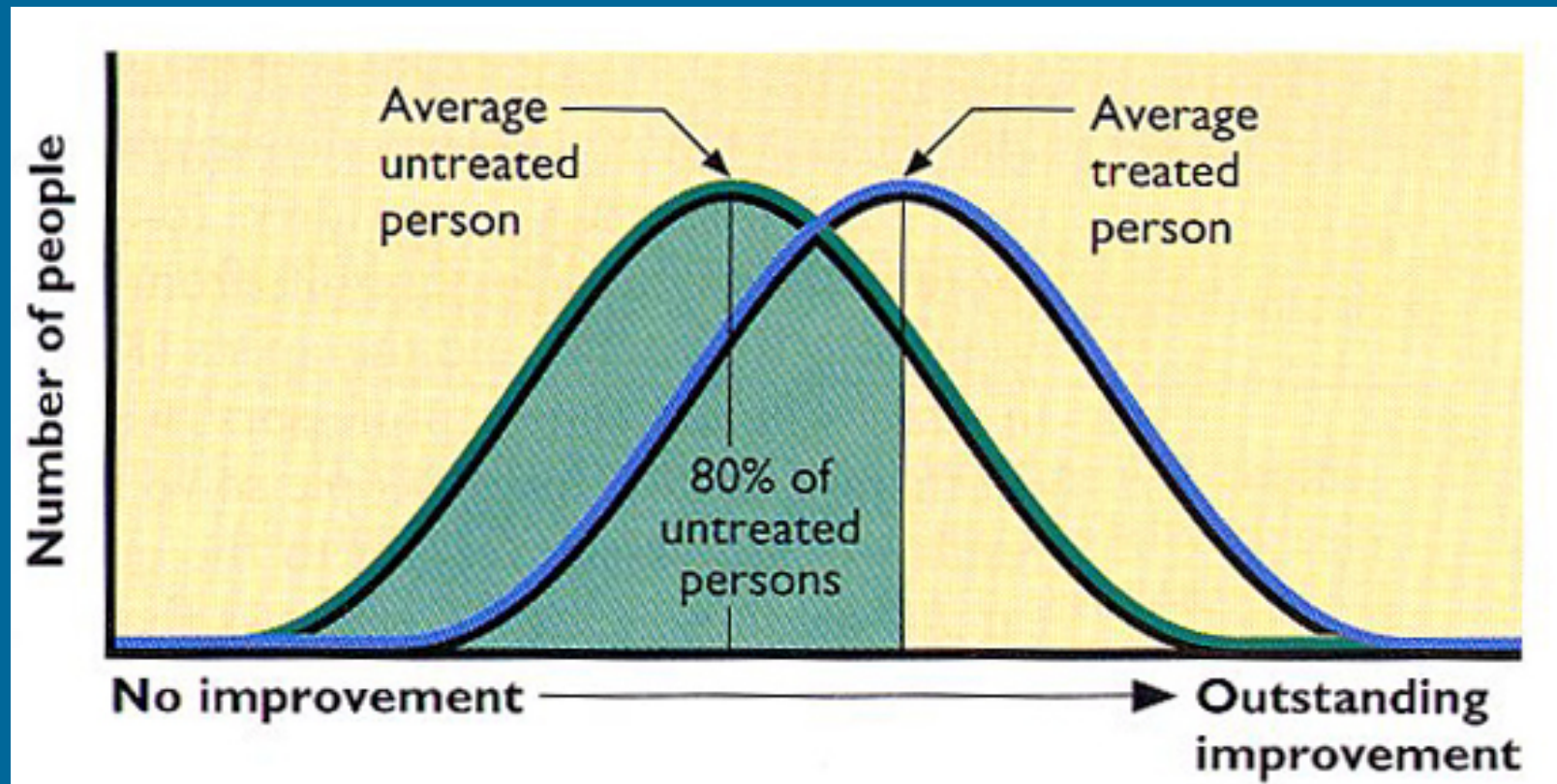
A Brief History

An equally common problem is that hundreds of studies accumulate in a research area. Here too there are likely to be many weak studies and considerable inconsistency in study outcomes. A precise method for combining the studies would be desirable. An additional problem is that the studies may be different in a variety of ways.

A Brief History

Modern methods of meta-analysis were developed to solve this second kind of problem. The meta-analytic “revolution” began in the late 1970s. The psychotherapy outcome analysis published by Smith and Glass (1977) brought the method to the attention of many psychologists, highlighting its potential advantages.

Smith and Glass (1977) examined over 300 therapy outcome studies and concluded that, overall, therapy was quite effective, placing the average treated person better off than 80% of untreated people.



An Example

Hypothesis: Frustration produces aggression

Study 1: Sample = college students
 Setting = psychology laboratory
 IV = easy versus impossible anagrams
 DV = rating of the experimenter's
 competence

An Example

Hypothesis: Frustration produces aggression

Study 2: Sample = college students
 Setting = psychology laboratory
 IV = helpful versus unhelpful confederate
 DV = shock level chosen to punish the
 confederate during a learning task

An Example

Hypothesis: Frustration produces aggression

Study 3: Sample = movie goers
 Setting = natural setting
 IV = number of people ahead in line when a
 show sells out
 DV = observed nonverbal and verbal
 behavioral indicators of anger

An Example

Hypothesis: Frustration produces aggression

Several years go by . . .

An Example

Hypothesis: Frustration produces aggression

Study 100: Sample = Third grade boys
 Setting = Phys. Ed. class
 IV = order in which a child is picked to be
 on a dodge-ball team
 DV = number of aggressive throws (above
 the waist)

Does Frustration Cause Aggression?

The challenge:

How can several studies, testing the same hypothesis, but using different methods, measures, samples, and settings be combined?

The solution:

In a meta-analysis, each study outcome is “standardized” to allow their combination and comparison. The seemingly troublesome differences (e.g., settings, measures) are coded and become moderators that are incorporated into the analysis.

In a meta-analysis, the unit of analysis is the individual study, akin to an individual participant in primary research.

Each study provides a statistical test of the conceptual hypothesis. The outcome of that statistical test (e.g., statistical significance, effect size) provides a standardized outcome, akin to the dependent measure in primary research.

Each study can also be described in many ways (e.g., type of sample, method of measurement, location) that can be used as predictors in the meta-analysis, akin to the independent variables in primary research.

Primary Research (Study 1)
(Laboratory Setting, College Men, Self-reported Anger
Following Easy versus Impossible Anagrams)

Unit of Analysis	Dependent Variable	Independent Variable
Participant 1	Anger Rating = 4	Treatment Condition
Participant 2	Anger Rating = 6	Treatment Condition
Participant 3	Anger Rating = 2	Control Condition
Participant 4	Anger Rating = 3	Control Condition
Participant 5	Anger Rating = 3	Treatment Condition
Participant 6	Anger Rating = 5	Treatment Condition
Participant 7	Anger Rating = 1	Control Condition
Participant 8	Anger Rating = 2	Control Condition
Etc.		
Etc.		

In this single study, the t-test comparing treatment and control might show support for the hypothesis that frustration produces anger, $t(28) = 1.70$, $p = .10$. But, the sample size is small, the significance level marginal, and so some doubt exists about the legitimacy of the inference.

By itself, the study is not very informative. If many such studies exist, however, they can be aggregated in a meta-analysis, and their combined results used to make a stronger inference about the hypothesis.

Aggregation simply requires that we convert each study outcome to a standard metric. This can be done in two basic ways: *statistical significance* and *effect size*. Both provide a “metric-free” measure that allows combination across different kinds of outcomes.

The kind of effect size used distinguishes the major types of meta-analysis: d (Cohen; Glass), g (Hedges), r (Rosenthal; Hunter & Schmidt),

Meta-Analysis

Unit of
Analysis

Dependent
Variable

Independent
Variables

Study 1

$d = 0.23$

Males, self-report, laboratory

Study 2

$d = 0.55$

Mixed, self-report, natural setting

Study 3

$d = 0.02$

Females, behavior, laboratory

Study 4

$d = -0.23$

Females, behavior, natural setting

Study 5

$d = 1.54$

Males, self-report, natural setting

Study 6

$d = 0.10$

Mixed, behavior, laboratory

Study 7

$d = 0.05$

Males, behavior, laboratory

Study 8

$d = -0.33$

Females, self-report, laboratory

Etc.

Etc.

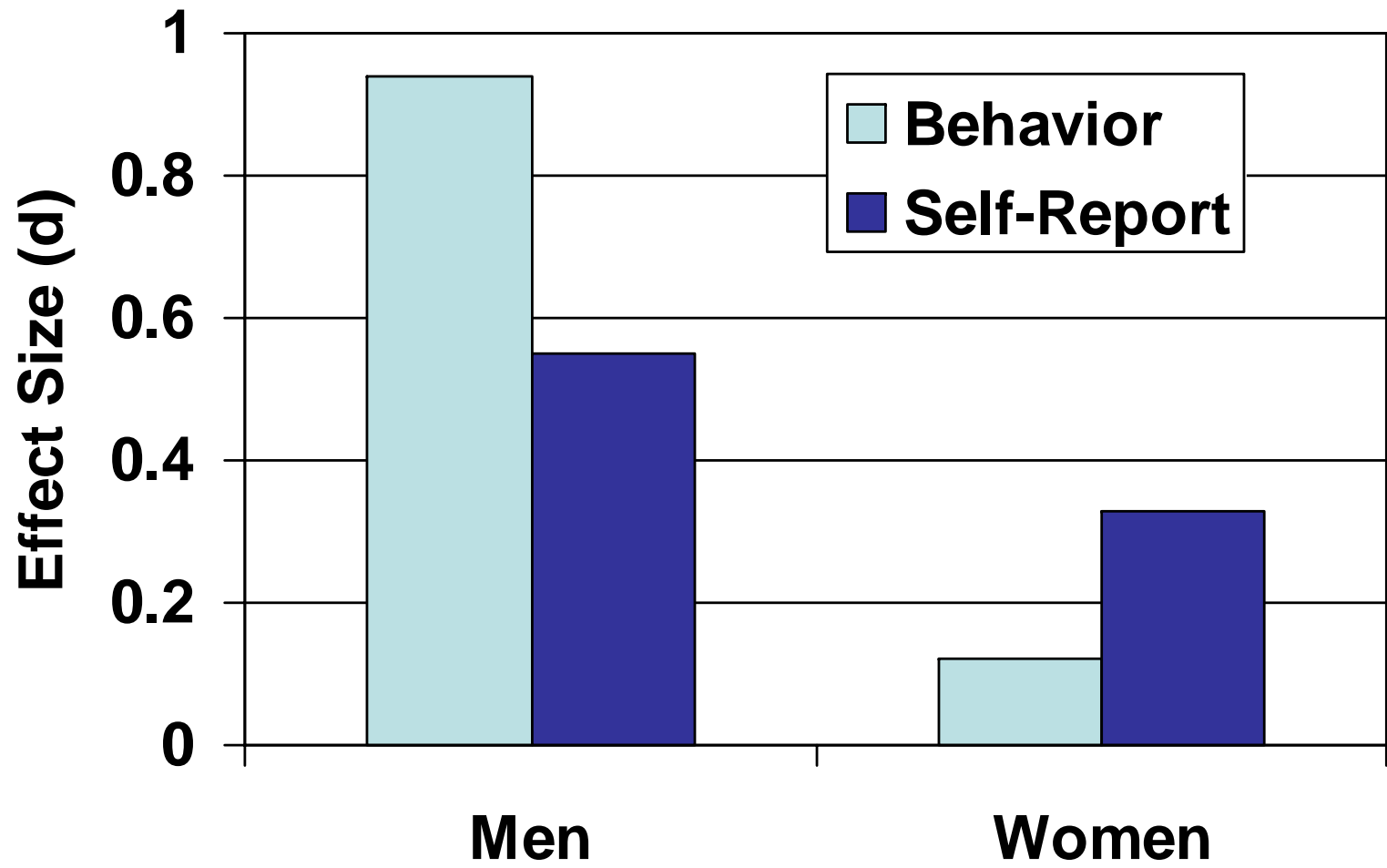
Simply aggregating across the studies provides an overall test of the common hypothesis: *“Do groups exposed to some manipulation of frustration exhibit some form of aggression more than groups not exposed to frustration?”*

The mean effect size gives an indication of the strength of the relation. A confidence interval around that mean provides an appropriate level of caution.

A meta-analysis can go further than simple aggregation. There is likely to be considerable variability in the size of effects across the many studies. This variability may be random, but, it may also be due to important moderator variables. The many ways that the studies vary can be coded and used as predictors of effect size to test these moderator variable hypotheses.

A significant moderator variable would indicate that the nature of the basic relation being studied (e.g., frustration-aggression), depends on some feature of the measures, methods, sample, or settings.

The results might show, for example, that the frustration-aggression relation is stronger for men than women and is especially likely to emerge for behavioral measures compared to self-reports.



What can meta-analysis accomplish?

1. Classic forms of validity

- statistical conclusion validity
- internal validity
- construct validity
- external validity

2. Description of study characteristics: What has been done and found to be particularly effective or ineffective?

3. Power

4. What “holes” need to be filled?

	Men		Women	
	Self-Report	Behavior	Self-Report	Behavior
Laboratory	Study 1, 2, 4, 7, 9, 10, 15		Study 3, 5, 8	
Natural Setting		Study 6, 11	Study 12, 13	Study 14

Caveat Emptor . . .

1. Meta-analysis is best thought of as a quasi-experimental design. Ordinarily the “independent variables” in a meta-analysis are study characteristics over which the meta-analyst has no control.
2. The conclusions are only as good as the methods used to select, code, and analyze the data.

3. The number of moderators that can be tested is limited by the number of studies included and the natural crossing of variables that exists.

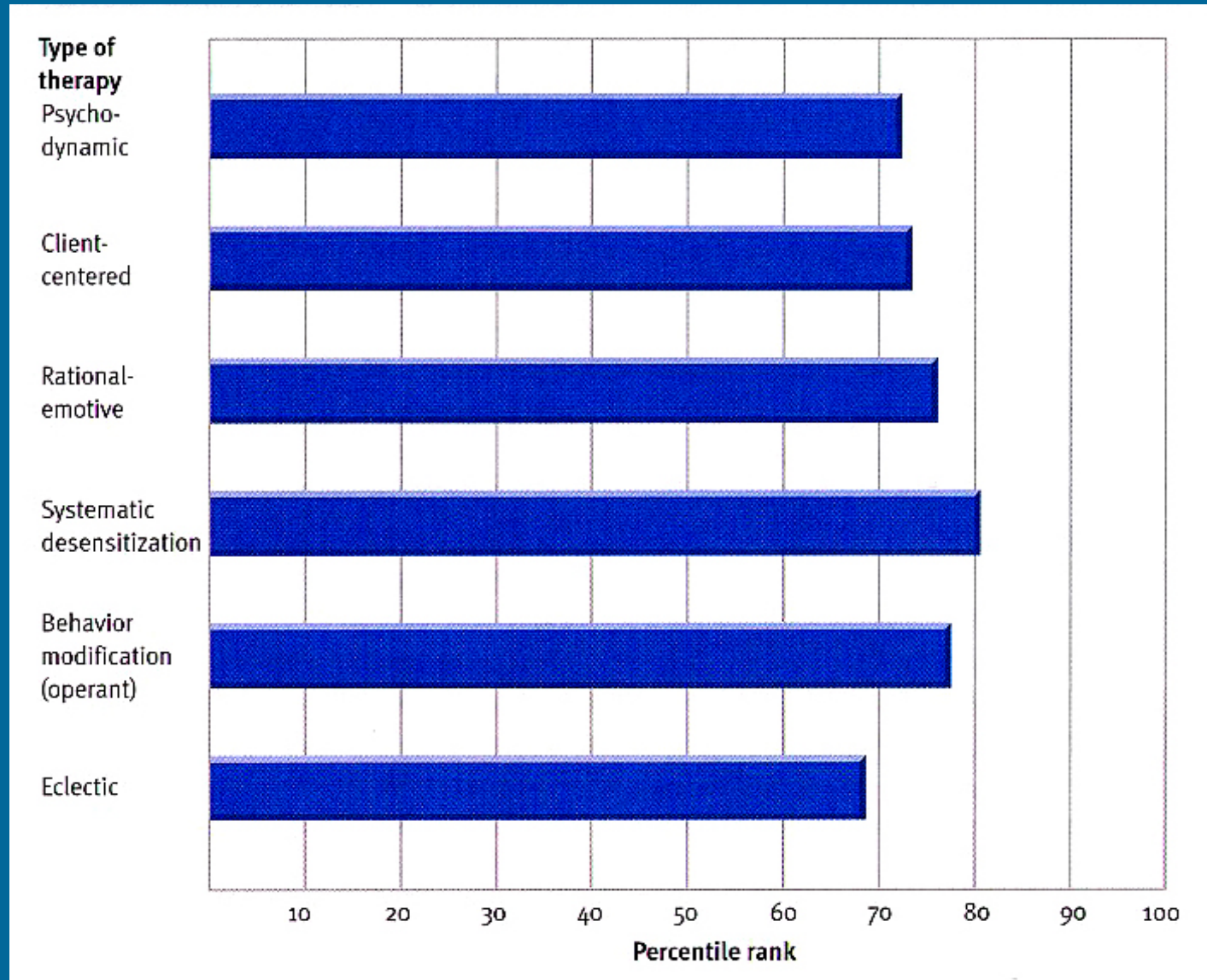
	Men		Women	
	Self-Report	Behavior	Self-Report	Behavior
Laboratory	Study 1, 2, 4, 7, 9, 10, 15		Study 3, 5, 8	
Natural Setting		Study 6, 11	Study 12, 13	Study 14



4. The file drawer problem.

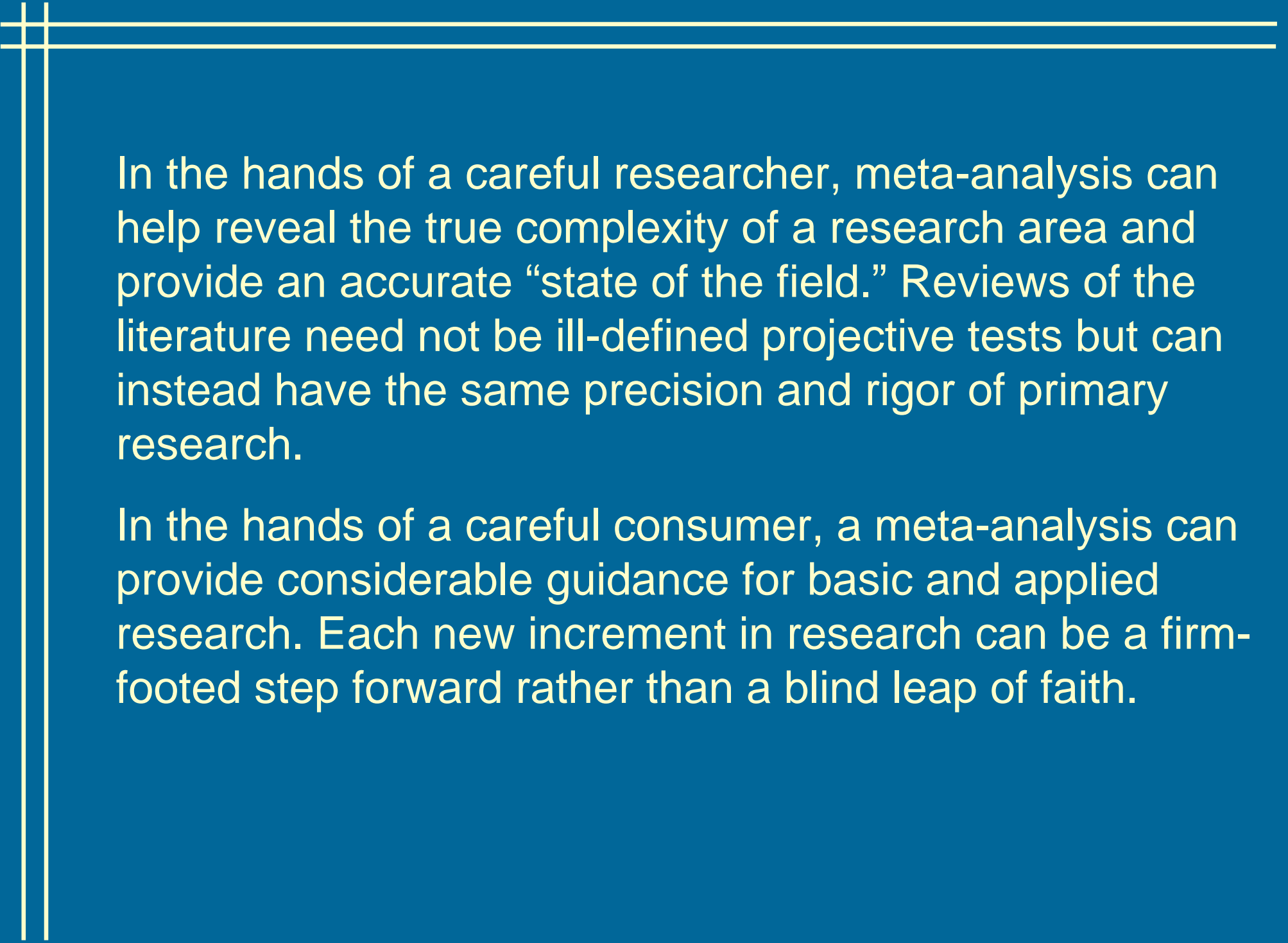
5. Multiple outcomes from a single study are not independent.

6. Review level conclusions may not correspond to study level conclusions.



6. Review level conclusions may not correspond to study level conclusions.

	Treatment A	Treatment B	Treatment C	Treatment D
Diagnosis W	Studies 1-20	?	?	?
Diagnosis X	?	Studies 21-40	?	?
Diagnosis Y	?	?	Studies 41-60	?
Diagnosis Z	?	?	?	Studies 61-80



In the hands of a careful researcher, meta-analysis can help reveal the true complexity of a research area and provide an accurate “state of the field.” Reviews of the literature need not be ill-defined projective tests but can instead have the same precision and rigor of primary research.

In the hands of a careful consumer, a meta-analysis can provide considerable guidance for basic and applied research. Each new increment in research can be a firm-footed step forward rather than a blind leap of faith.



But wait . . . there's more . . .



Diversity in Meta-Analysis: Brinley Plots & Related Approaches

- ▶ Two Meta-Analyses:

 - Sliwinski & Hall (1998)

 - Myerson, Adams, Hale, & Jenkins (in press)

- ▶ Purpose of Sliwinski & Hall (1998):

 - Testing the General Slowing Hypothesis

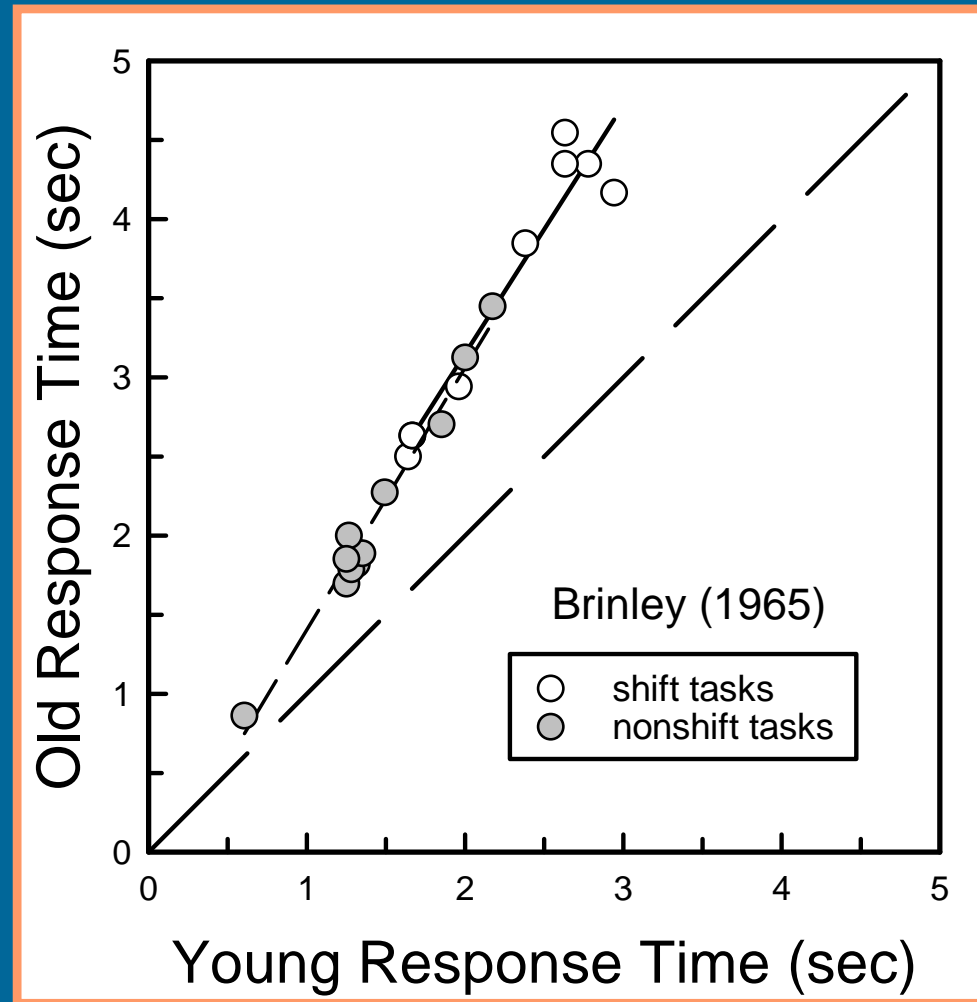
 - New Technique: Hierarchical Regression

- ▶ Purpose of Myerson et al. (in press):

 - Testing the Relative Speed Hypothesis

 - Systematic Comparison of Two Speed Measures

In the beginning...



Sliwinski & Hall (1998)

Defining a Data Set

Hierarchical Regression Technique: Study-level Slopes

⇒ Complexity Manipulation Requirement: 3 Tasks

Inclusion Criteria:

Ages of two groups: (~~30~~ 30) and (~~60~~ 60)

Error rates comparable for two age groups

Appropriate complexity manipulation

RT data reported as a function of complexity

Sliwinski & Hall (1998)

Data Set

Mental Rotation

Berg et al. (1982)
Cerella et al. (1981)
Dror & Kosslyn (1995)
Hale et al. (1991)
Hale et al. (1995)
Hertzog et al. (1993)

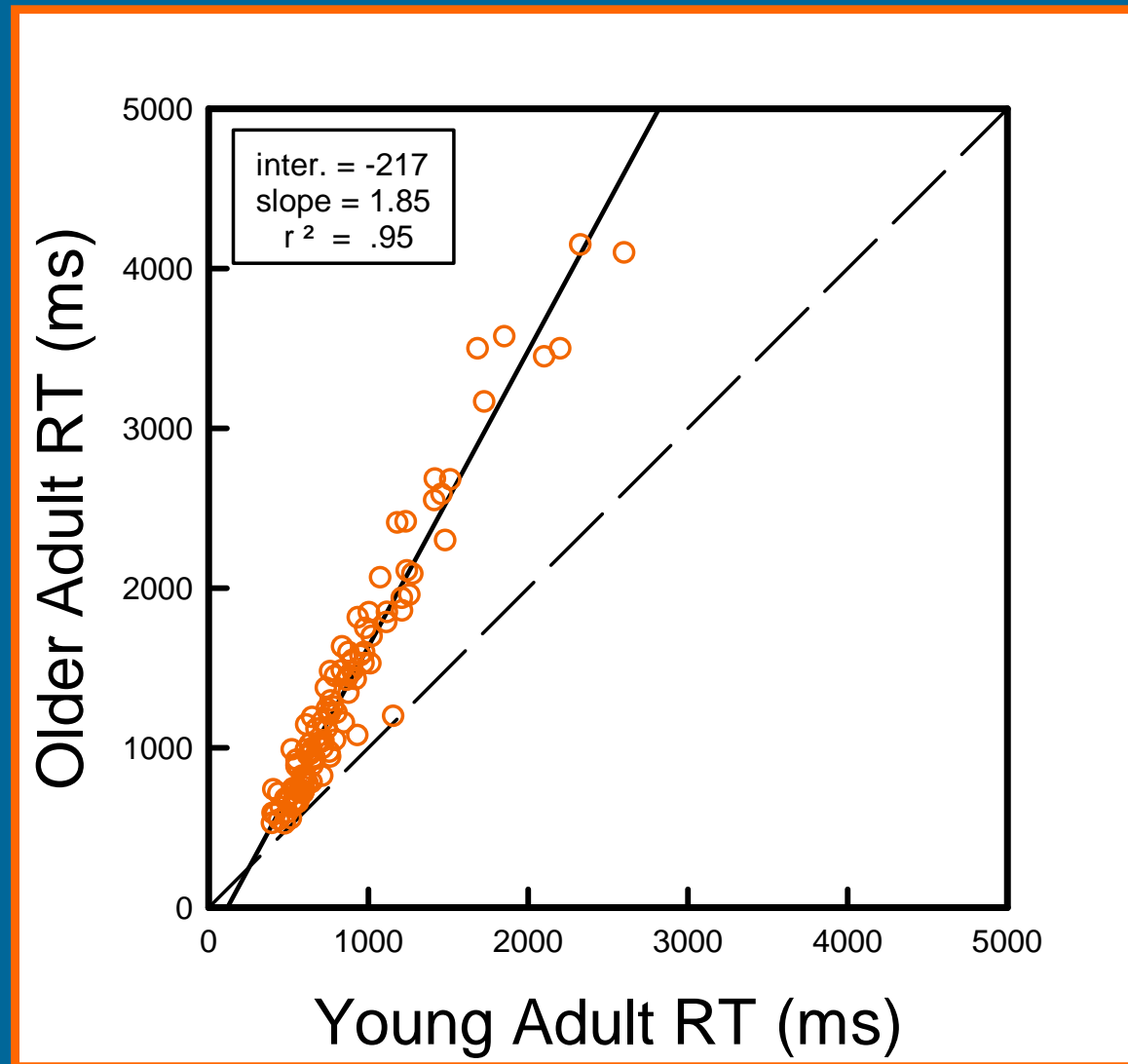
Visual Search

Foster et al. (1995)
Hale et al. (1995)
Madden (1986)
Plude et al. (1983)
Plude & Hoyer (1986)
Plude & D-Roosevelt (1989)
Zacks & Zacks (1993)

Memory Scanning

Cerella et al. (1986)
Coyne et al. (1986)
Fisk et al. (1990)
Madden (1982)
Menich & Baron (1990)
Puglisi (1986)
Salthouse & Somberg (1982)
Salthouse (1994)
Strayer & Kramer (1994)

Sliwinski & Hall (1998)



Sliwinski & Hall (1998)

Results and Conclusions

Within-experiment slopes varied as a function of task:

Average slowing for Mental Rotation: All estimates: 95% CI > 1.0

Average slowing for Visual Search: All estimates: 95% CI > 1.0

Average slowing for Memory Scanning: Some estimates: 95% CI < 1.0

Slowing equal for Mental Rotation and Visual Search

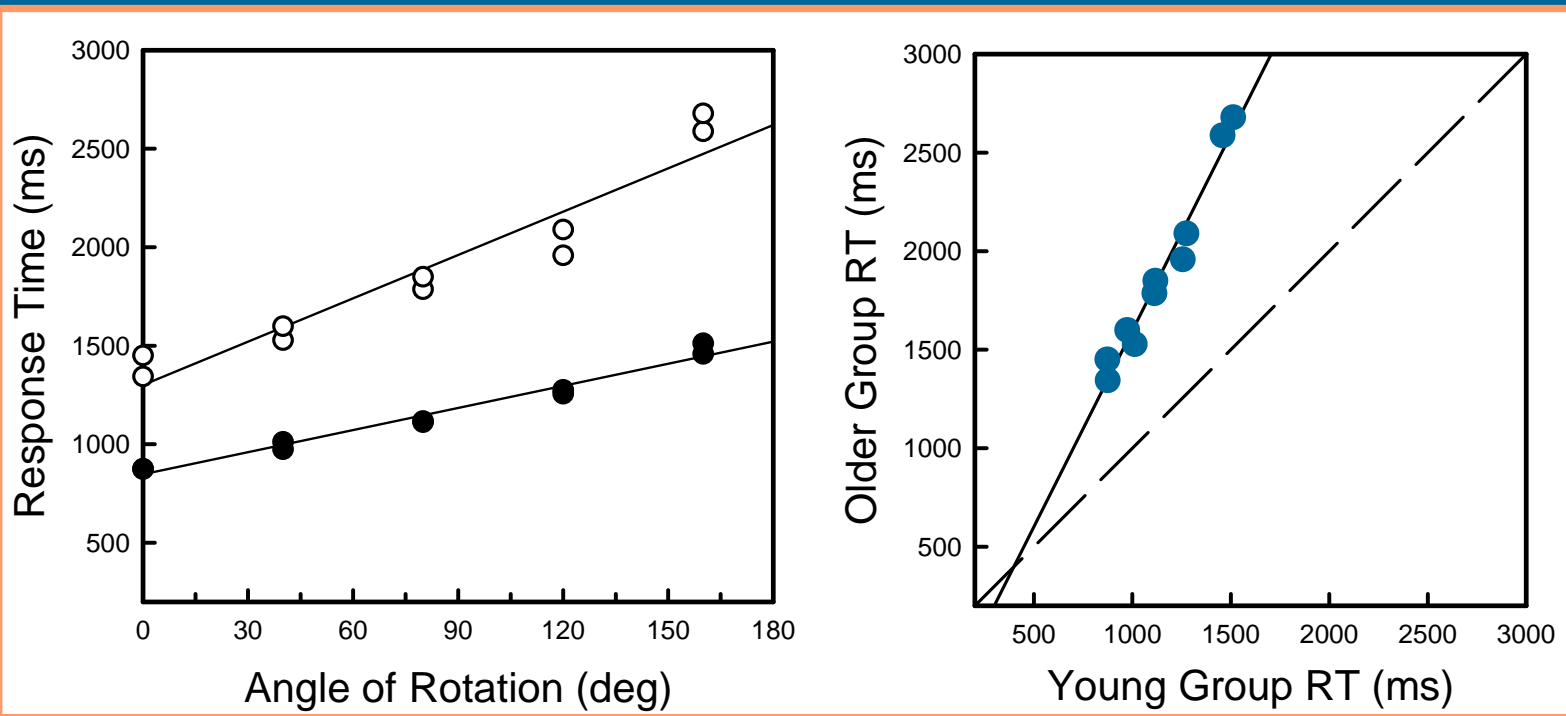
General Slowing **not** General: Only Evident for 2 Tasks

Hierarchical Regression Technique useful:

Revealed what linear fit to combined data could not provide

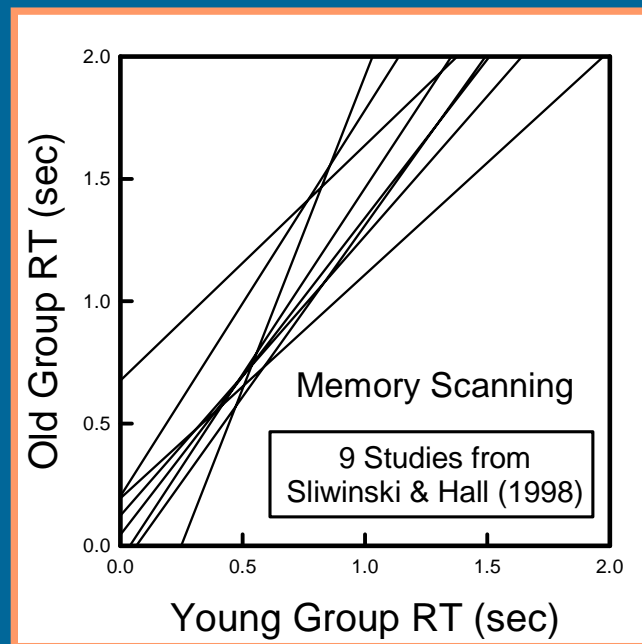
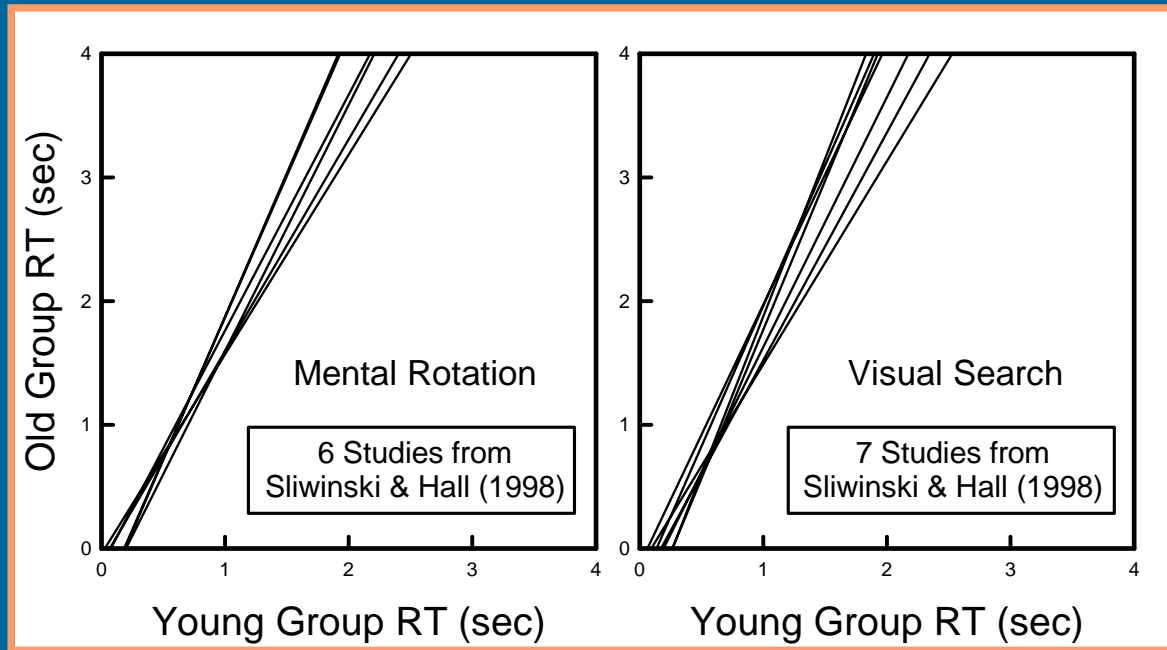
Myerson et al. (in press)

Cerella et al. (1981): Mental Rotation Study

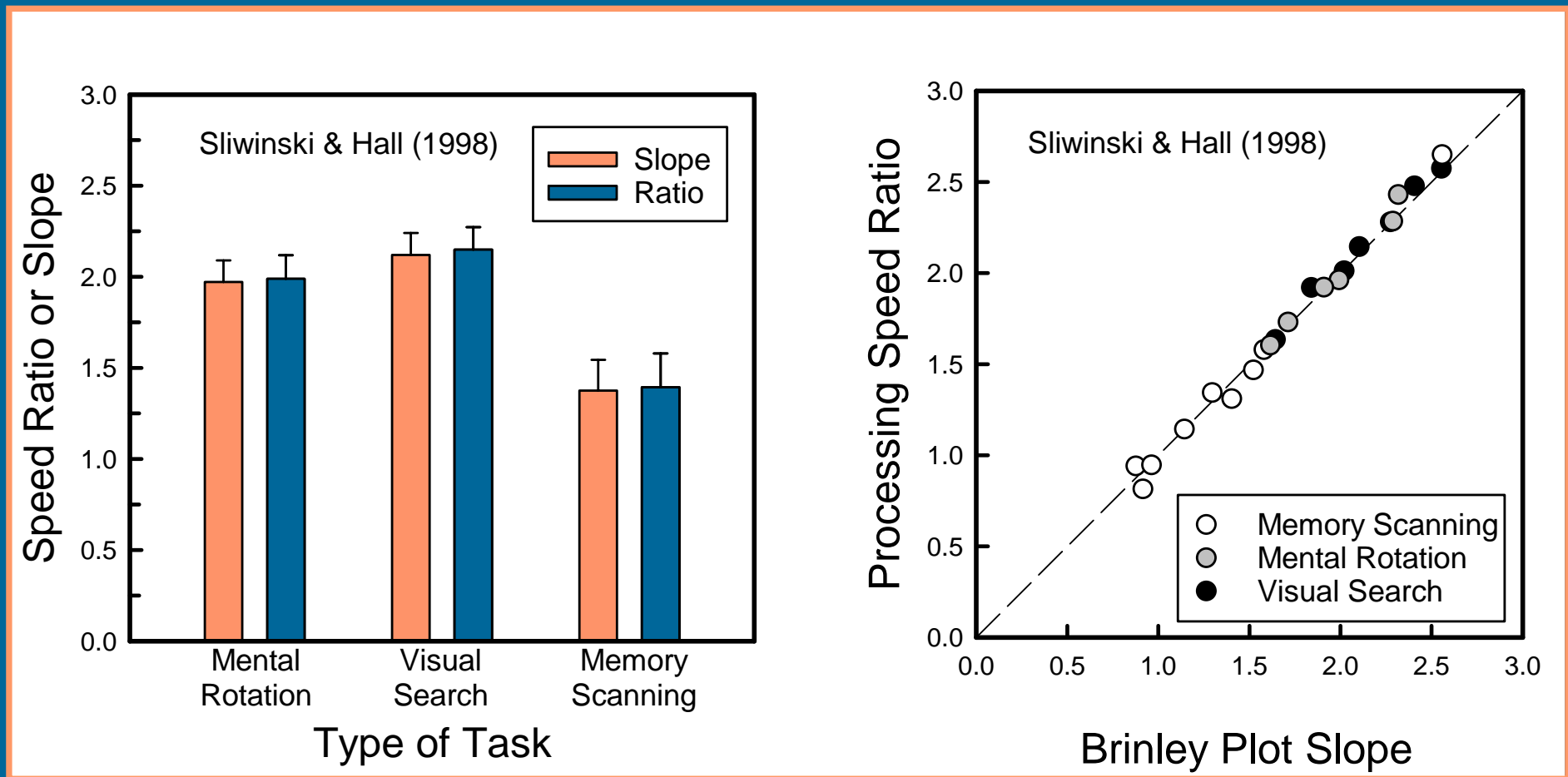


<u>Rate of Rotation</u>	<u>Young</u>	<u>Old</u>	<u>Speed Ratio</u>	<u>Brinley Slope</u>
Slope (ms/deg)	3.74	7.33	1.96	1.99

Myerson et al. (in press)

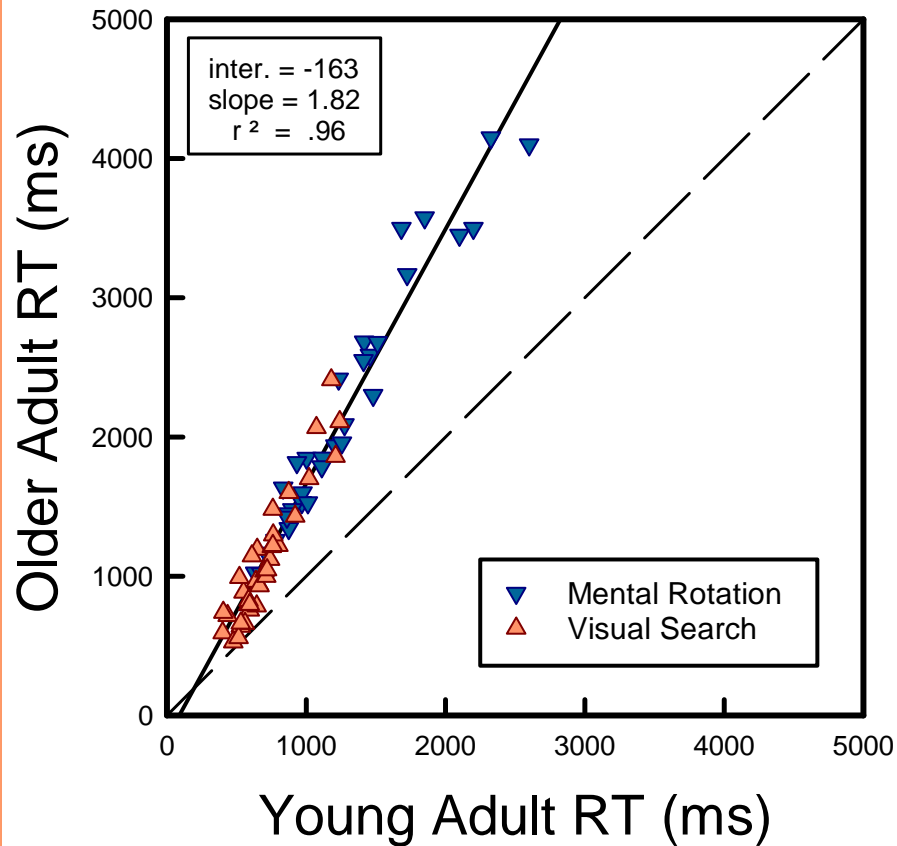


Myerson et al. (in press)

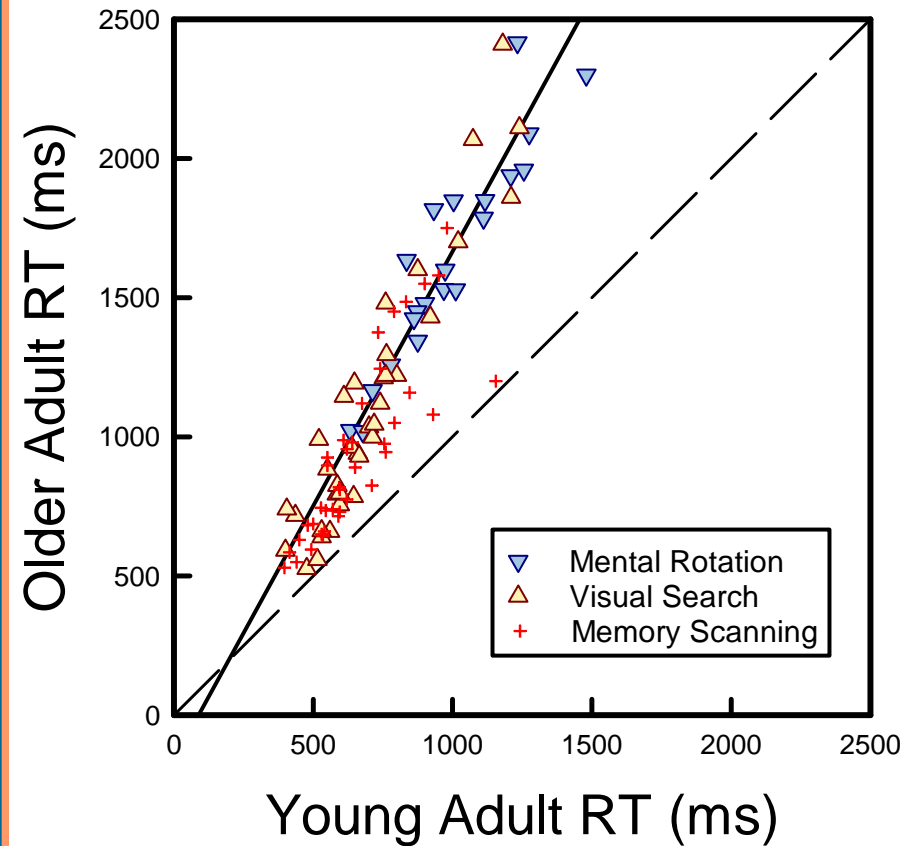


Myerson et al. (in press)

Full Range: 2 Tasks



Restricted Range: 3 Tasks



Myerson et al. (in press) Conclusions

Estimates of General Slowing Differ by Task:

Mental Rotation and Visual Search 2.0

Memory Scanning 1.25

For Each Task: Speed Ratios equivalent to Brinley Slopes

Explanation for Task Differences:

Domain Differences

Poor Reliability of Memory Scanning Slopes

Technique Reveals: Support for Relative Speed Hypothesis

Meta-Message

“Good data are always separable, with respect to their scientific importance, from the purposes for which they were obtained.”

Murray Sidman (1960)
Tactics of Scientific Research

Meta-Meta-Message

Good databases are always separable, with respect to their scientific importance, from the purposes for which they were originally assembled.