

Clinical Practice, Clinical Research, and Evidence-Based Practice:

Some Places We Might Like To Go

Randall R. Robey
University of Virginia

From the outset, it is important to note that the ideas expressed in this address emerge from a confluence of three important works. That is, the basics notions described below flow from, and build upon, three original works:

Thompson (2002)

Dollaghan (2007)

Man-Son-Hing, et al. (2002)

The integration of these works leads naturally to evidence-based practice and, most importantly for our purposes, clinical research for informing clinical practice. We begin then with a brief review of our understandings of evidence-based practice (EBP) and then move to a series of possible future states in which clinical research informs clinical practice in meaningful ways. At all points, the premise is that practitioners and researchers mutually benefit and the realized values accrue to the well-being of individuals with communication disorders.

A. Our starting point

1. We know that evidence-based practice is about individuals with communication disorders perceiving important and meaningful changes in their lives brought about by our clinical interventions.
2. Furthermore, we know that EBP is about a single and certain clinician identifying a single and certain clinical decision that she would like to improve in her clinical practice. As a result,
 - ... EBP is client specific
 - ... EBP is clinician specific
 - ... EBP is clinical-action specific
 - ... EBP is focused learning; it is the result of a choice for self; it's a process for filling a self-determined gap in one's clinical knowledge or skill.
3. We know that EBP is the integration of ...
 - a. Current best evidence
Valid and relevant clinical research
 - b. Clinical expertise
Clinical skills and experience

- c. Client values
 - i. Client preferences
 - ii. Client concerns
 - iii. Client expectations
- d. Client circumstances

Individual clinical picture and circumstances

That is, in consultation with a client, an evidence-based practitioner integrates best evidence through her clinical expertise, to decide upon an option for intervention on a case-by-case basis (adapted from Straus, et al., 2005).

4. We are familiar with the five steps of evidence-based practice. (adapted from Straus, et al., 2005)
 - a. Identify a need for clinical information to improve the quality of clinical decisions and express that need as an answerable clinical question particularizing four dimensions...
 - P populations
 - I interventions
 - C comparative interventions
 - O outcomes
 - b. Find and access all the literature for relevant and best-available clinical information
 - c. Critically appraise that evidence for appropriateness and validity for a client and a practice
 - d. Present each client with highest-quality choices and, together, determine a course of care.
 - e. Self-evaluate performance as an evidence-based practitioner.
5. We understand that EBP is a means of managing uncertainty
 - a. There a very few things we know for certain.
 - b. In the absence of absolutes, the best we can do is make decisions with as clear an understanding of the uncertainties as is possible.
 - c. EBP is a process for making informed and rational decisions in an uncertain world.
 - d. Example:

One diagnostic protocol has sensitivity of .80, specificity of .80, and a LR+ of 6.0.

A competing protocol has sensitivity of .90, specificity of .90, and LR+ of 18.0

Is either test so flawed that it is rendered useless?

Is either test guaranteed to 'get it right' every time?

Within that context, in using which test do we best manage the uncertainty?

And so we have come to understand that EBP is not about uncovering the one correct thing to do; it's about doing one or another of the best things under the circumstances in which a choice must be made.

6. Lastly, we understand that EBP is a dynamic continuing process.

What was a best intervention choice for the last client is not necessarily the best choice for the next client.

We evolve and adapt all aspects of the EBP process as new evidence becomes available.

EBP is a dynamic integration of ever evolving clinical expertise and external evidence in day-to-day practice.

With that brief review of the basics completed, let's turn to a few possibilities for optimizing our future.

B. Possible destination #1

1. Let's say two things begin happening and continue happening into the foreseeable future.
 - a. Scientists consistently report primary outcomes obtained through group-comparison research designs in an EBP friendly fashion.

Here and elsewhere, that means reporting estimates of effect size for primary outcomes. For this case in particular, it means reporting estimates of effect size with (non)central confidence intervals about them.
 - b. Other scientists consistently produce meta-analyses of those primary studies
2. A little background

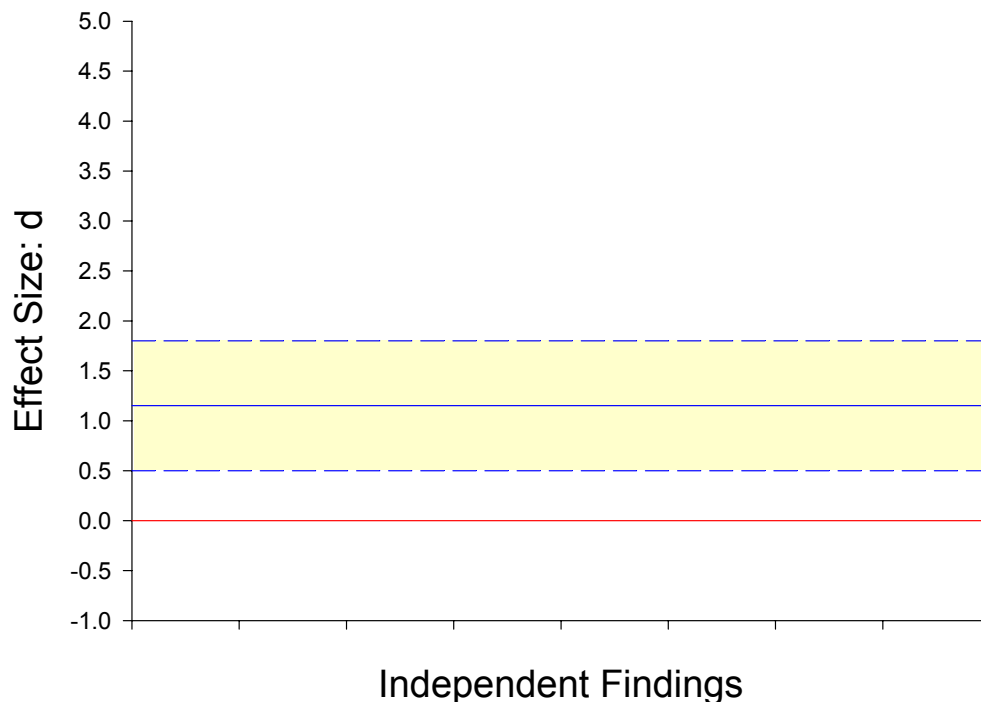
An effect size is a metric for indexing departures from the null hypothesis.

 - a. It indexes how untenable is the null hypothesis.
 - b. In general, when the null hypothesis is true, effect size is zero.
 - c. The range of d is $-\infty$ to $+\infty$.
3. Now, let's suppose that a meta-analysis of treatments for aphasia in the acute period of recovery produces an average effect size of $d = 1.15$ with a .95 confidence interval (CI) of ± 0.65 .

Furthermore, suppose that everyone agrees that ...

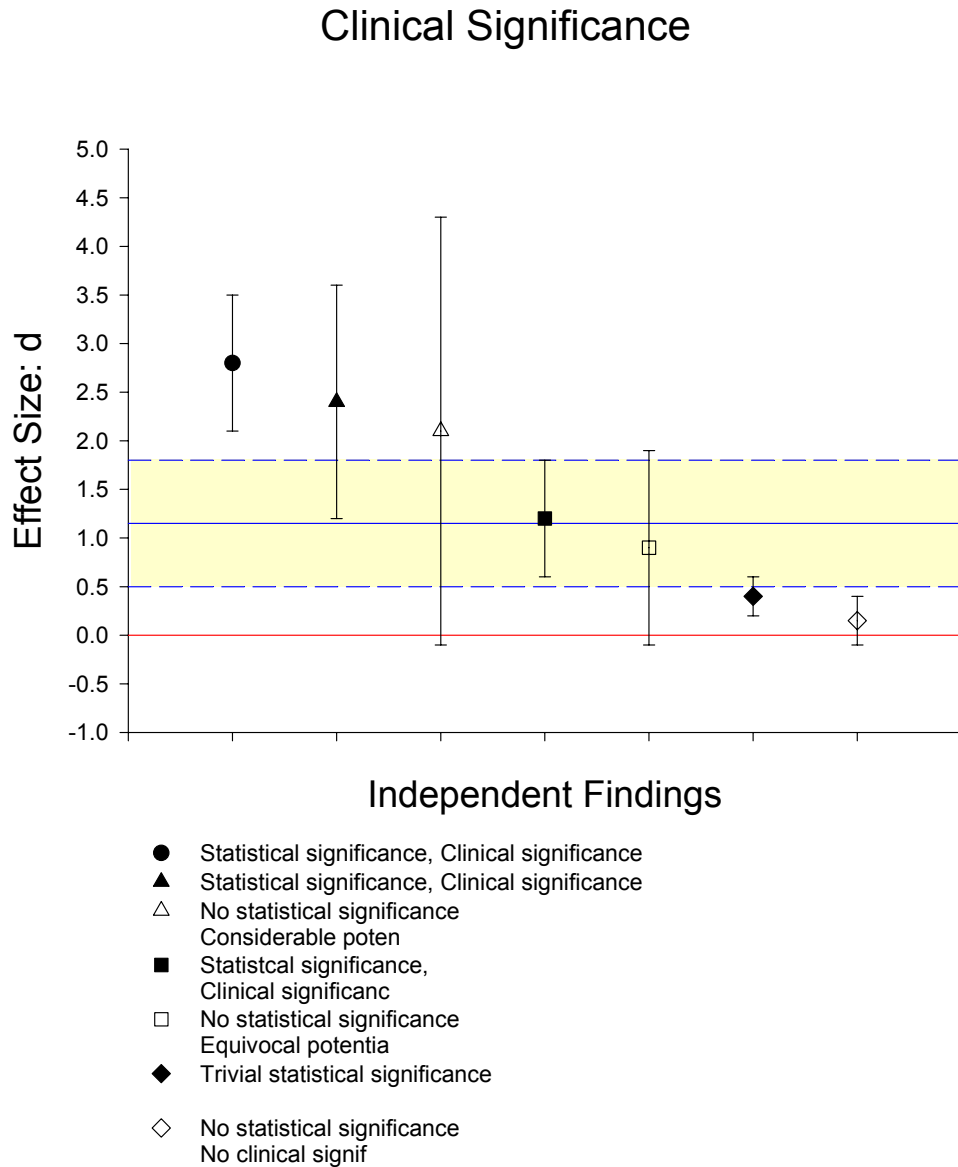
- a. A small effect size for this literature is $d = 0.50$ (the lower limit of the CI)
 - b. A medium effect size for this literature is $d = 1.15$ (the mean)
 - c. A large effect size for this literature is $d = 1.80$ (the upper limit of the CI)
4. If we can project that far, it is a small step to depict status quo re. extant research within the vertical space of a two dimensional plot. In the following figure, the solid red line represents $d=0.00$. A confidence interval encompassing this criterion implies an absence of statistical significance. The solid blue line represents the average effect size in the salient literature. The interrupted blue lines represent the upper and lower bounds of the CI about that mean. The space shaded in yellow then represents a range of status quo for outcomes brought about through treatments for aphasia. That is, the yellow space is the range of what can be called current clinically important outcomes based on relevant research to date.

Clinical Significance



5. Accepting that premise allows a consumer of a new clinical report of an aphasia treatment to assess the clinical importance of the finding with relative ease.

Consider each of the CI's in the following figure as coming from a new study. A clinically significant finding is one that exceeds status quo ante and so motivates a change in clinical practice (the first two cases on the left).



4. Now, let's take the example a step further. The meta-analysis described above already exists (Robey, 1998). Let's say that a brand new primary study reports a test of a new treatment for aphasia.

As clinicians, we look at the new effect size along with its CI following the example above. The clinical significance of the new protocol clear to us and all concerned.

Is it within our means to realize that possible future?

5. Let's consider a different example.

Let's say that a new technology is designed to achieve the same level of change as an existing technology but at a substantially faster rate and substantially reduced cost.

In this case, $d = 0.00$ is an important outcome in a head-to-head competition.

That is, the new technology achieves the same outcome as the standard but in less time and at less cost. The lesson, of course, is that the magnitude of effect size that is clinically important in any one clinical context is solely determined by whatever constitutes a meaningful departure from status quo ante in that clinical circumstance.

C. Possible destination #2

1. Let's say two things begin happening and continue happening into the foreseeable future.

a. Scientists consistently report primary outcomes in an EBP friendly fashion.

In this example, we focus on estimates of effect size with central confidence intervals about them – for primary outcomes obtained through single-subject research designs.

b. Other scientists consistently produce meta-analyses of those primary studies

2. A little background

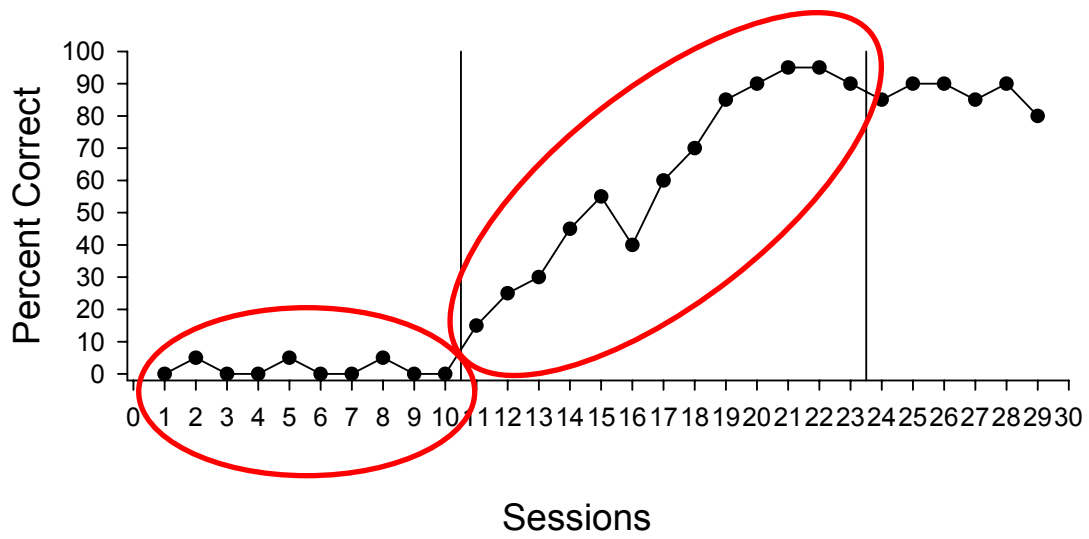
a. Change in slope between the A₁ and B periods

i. Efficiency of an intervention

ii. Rate of change

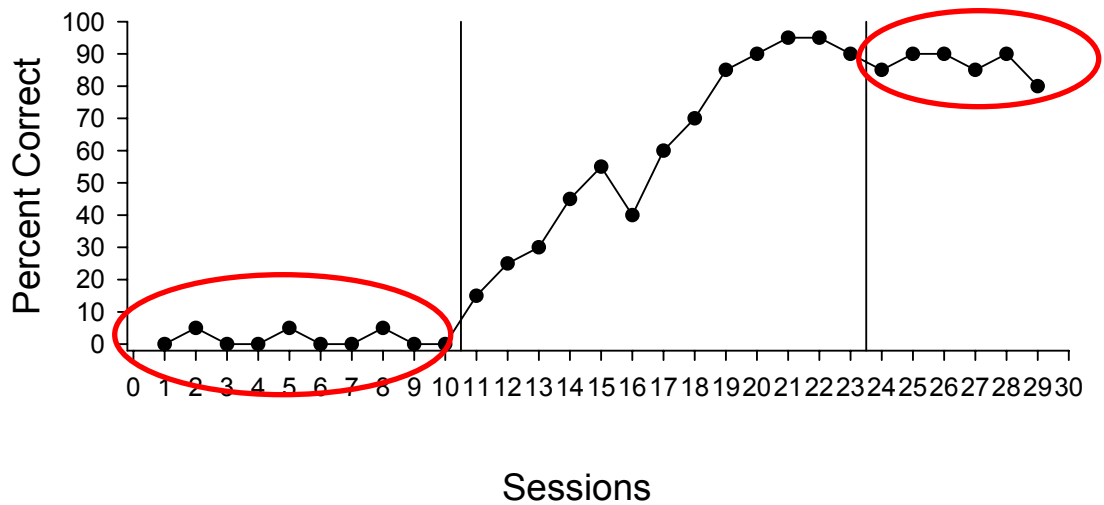
iii. Profile of change

iv. Learning curve



b. Change in level between the A₁ and A₂ periods

- i. Magnitude of change
- ii. Efficacy of an intervention
- iii. Effectiveness of an intervention



- c. Estimates of effect size for magnitude of change

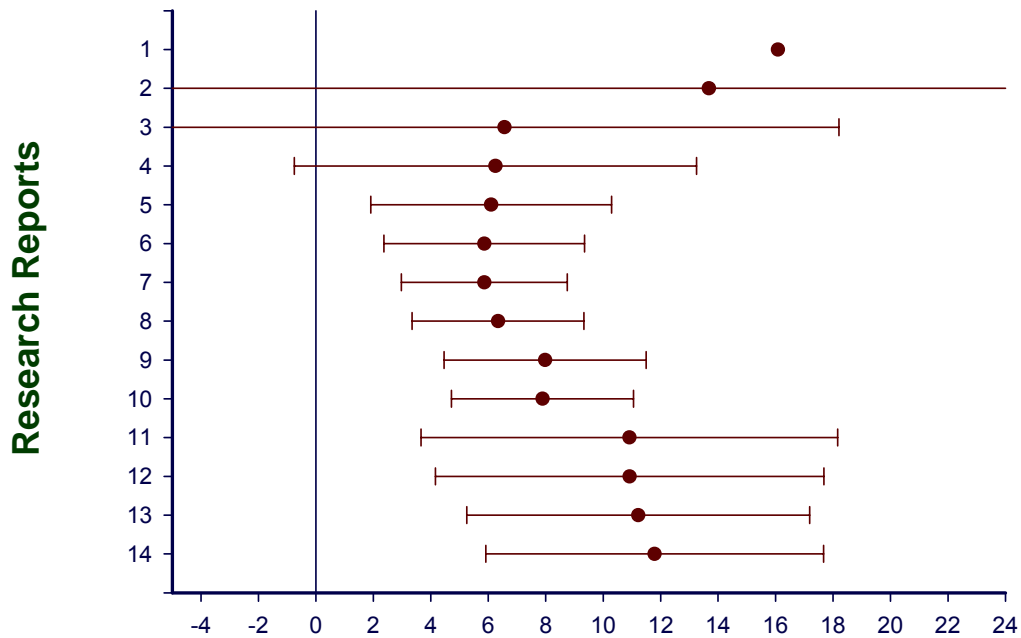
$$d = \frac{\bar{X}_{A_2} - \bar{X}_{A_1}}{S_{A_1}}$$

where A₁ is the baseline period and A₂ is the maintenance or withdrawal period.
 (see Beeson & Robey, 2006, for more detail)

- 3. For an example, let's say a researcher produces a meta-analysis of single-subject studies of treatments for aphasia aimed at improving syntax.

Single-Subject Direct Treatment Effects

Outcome: Syntax



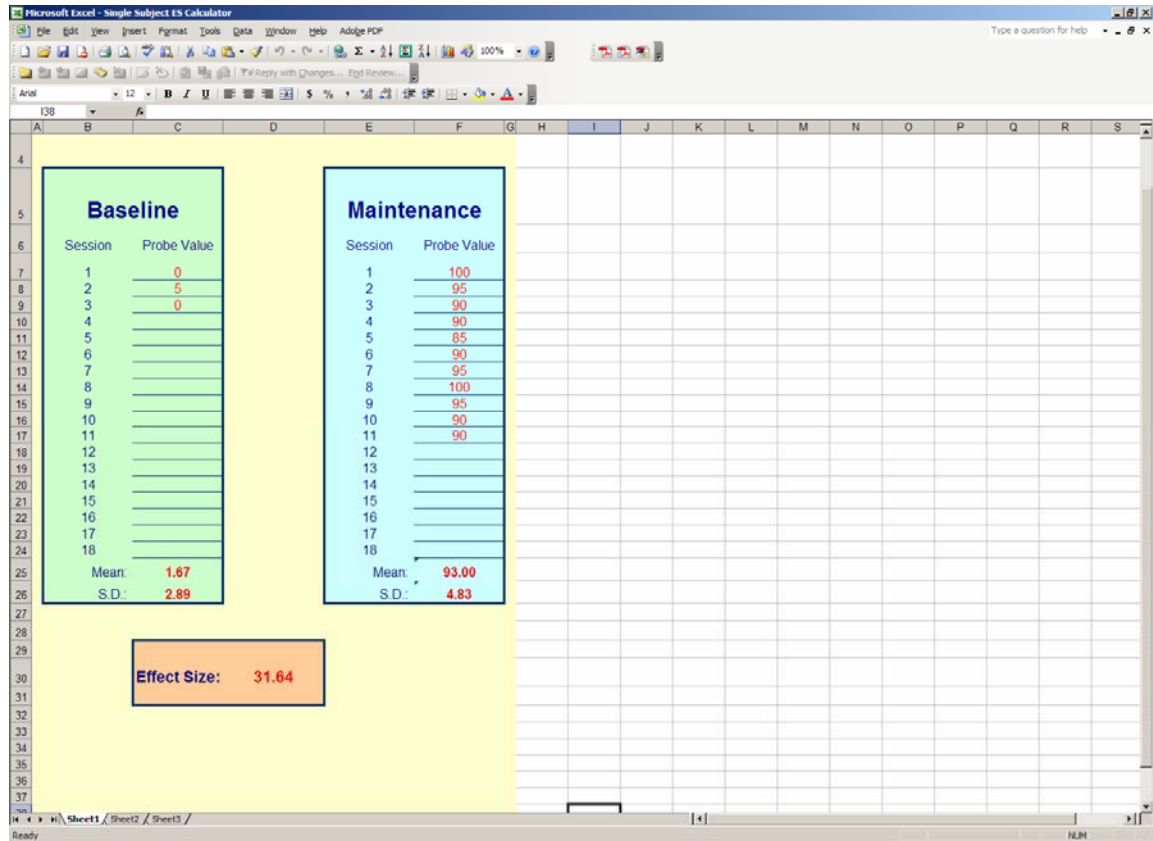
Average of Effect Size with .95 CI (Progressive Cumulative Average)

- 4. Now, how might we index small, medium, and large treatment effects in this literature?

How large an effect size would a new treatment protocol need to produce in order to be considered clinically important?

Could a clinician administering that protocol, calculate the very same estimate of effect size for a direct comparison to research-based outcomes?

Consider a simple Excel® sheet can calculate d and remove the need for any math.



Is it within our means to realize that possible future?

D. Possible destination #3

1. Let's say two things begin happening and continue happening into the foreseeable future.
 - a. Scientists consistently report primary outcomes of diagnostic studies in an EBP friendly fashion.

In this example, we are interested in estimates of effect size with central confidence intervals about them – for primary outcomes obtained through 2X2 classification studies (i.e., LR+ and LR-).

- b. Other scientists consistently produce meta-analyses of those primary studies
2. A little background (OK, a lot of background)

- a. Outcomes for a diagnostic protocol

- i. A negative result (NegRes; “-”) is a range of scores suggesting the absence of disorder

- No disorder
- Status: normal or typical
- Pass

This outcome does not mean that the disorder is not present in truth.

An outcome of this type is taken as evidence of the disorder is not present.

The obvious question is: How (un)certain am I in concluding that the evidence is indeed correct?

- ii. A positive result (PosRes; “+”) is a range of scores suggesting the presence of disorder

- Status: abnormal
- Status: disorder
- Fail

This outcome does not mean that the disorder is present in truth.

The obvious question is: How (un)certain am I in concluding that the evidence is indeed correct?

b. Sensitivity

i. Sensitivity is the rate of correct PosRes-s.

ii. When sensitivity ≈ 1.00 , we have a test property known as a SnNOut
Sensitive test yielding a NegRes rules Out disorder

When a test is a SnNOut, a NegRes almost certainly means the person is not presenting with the disorder

c. Specificity

i. Specificity is the rate of correct NegRes-s.

ii. When specificity ≈ 1.00 : we have a test property known as a SpPIn
Specific test yielding a PosRes rules In disorder

When a test is a SpPIn, a PosRes almost certainly means the person is presenting with the disorder.

d. Likelihood ratio positive

LR+ answers the question: How influential should a PosRes be for ruling in the presence of a disorder?

We want this value to be as large as possible.

i. Sensitivity / (1 - Specificity)

ii. The degree of confidence that a person who scores in the positive (affected or disordered) range truly has the target disorder (Dollaghan, 2007).

iii. Scaling LR+

> 10	Large and often conclusive increase in the likelihood of disorder
5 - 10	Moderate increase in the likelihood of disorder
2 - 5	Small increase in the likelihood of disorder
1 - 2	Minimal increase in the likelihood of disorder
1	No change in the likelihood of disorder

e. Likelihood ratio negative

LR- answers the question: How influential should a NegRes be for *ruling out* the presence of a disorder?

We want this value to be as small as possible.

- i. $(1 - \text{Sensitivity}) / \text{Specificity}$
- ii. The degree of confidence that a person scoring in the negative (normal) range is truly free of the target disorder (Dollaghan, 2007).

- iii. Scaling LR+

1	No change in the likelihood of disorder
0.5 - 1.0	Minimal decrease in the likelihood of disorder
0.2 - 0.5	Small decrease in the likelihood of disorder
0.1 - 0.2	Moderate decrease in the likelihood of disorder
< 0.1	Large and often conclusive decrease in the likelihood of disorder

3. For the purpose of an example, let's think about a clinician reading two articles.

Each article assesses a different diagnostic battery for differentiating persons presenting with early-stage dementia of the Alzheimer's type from persons present with primary aging.

The first battery produces a LR+ of 8.00 with a CI of ± 1.10 and a LR- of 0.15 with a CI of ± 0.10 .

The second battery produces a LR+ of 6.00 with a CI of ± 1.80 and a LR- of 0.30 with a CI of ± 0.15 .

All other things being equal, which test does she begin using in her practice?

Importantly, the math can be reduced considerably. A simple Excel® sheet can calculate these values.

The screenshot shows an Excel spreadsheet titled "Microsoft Excel - Sensitivity & Specificity Calculator". The spreadsheet is divided into two main sections. The top section is a 2x2 contingency table for "Disorder Status by Reference Test Gold Standard". The bottom section is a list of calculated metrics with their respective values.

		Disorder Status by Reference Test Gold Standard		
		Present	Absent	
Experimental Test	Positive (Fail)	105	21	126
	Negative (Pass)	69	673	742
	TOTALS	174	694	868

	Probability	Odds
Sensitivity	0.60	
Specificity	0.97	
PPV	0.83	
NPV	0.91	
LR+	0.95	19.94
LR-	0.29	0.41
Prevalence	0.20	
PreTest	0.20	0.25
PostTest +	0.83	5.00
PostTest -	0.09	0.10

Even better: This web page does a much better job of the calculation in that it provides confidence intervals about sensitivity, specificity, LR+, and LR-; it just doesn't get any better than that.

<http://faculty.vassar.edu/lowry/clin1.html>

Clinical Calculator 1. From an Observed Sample: Estimates of Population Prevalence, Sensitivity, Specificity, Predictive Values, and likelihood Ratios

Given a sample of subjects cross-classified according to whether a certain condition is present or absent, and according to whether a test designed to indicate the presence of that condition proves positive or negative, this page will calculate the estimated population midpoints and 95% confidence intervals for

- prevalence of the condition;
- test sensitivity (conditional probability that the test will be positive if the condition is present);
- test specificity (conditional probability that the test will be negative if the condition is absent);
- predictive values of the test (probabilities for true positive, true negative, false positive, and false negative); and
- positive and negative likelihood ratios.

To proceed, enter the observed frequencies for each of the four cross-classifications into the designated cells, then click the «Calculate» button. To perform a new analysis with a new set of data, click the «Reset» button.

	Condition		Totals
	Absent	Present	
Test Positive	23	102	125
Test Negative	617	15	632
Totals	640	117	757

Reset Calculate

	Estimated Value	95% Confidence Interval	
		Lower Limit	Upper Limit
Prevalence	0.154557	0.129944	0.182748
Sensitivity	0.871795	0.794276	0.924066
Specificity	0.964063	0.945729	0.976564
For any particular test result, the probability that it will be:			
Positive	0.165125	0.139764	0.193952
Negative	0.834875	0.806048	0.860236
For any particular positive test result, the probability that it is:			
True Positive	0.816	0.734595	0.87748

Is it within our means to realize this possible future?

E. Possible destination #4

We get over 'levels of evidence.' (drawn heavily from Dollaghan (2007) with adaptations.)

Some research designs are more scientifically rigorous than others.

In general, more rigorous designs produce higher quality evidence.

However, it is very important to note that a design does not by itself define the quality of the scientific evidence it produces.

Let's examine some properties of research designs.

1. Non-experimental vs. Experimental Research

a. Non-experimental

- i. Experimenter observes a situation that is unfolding in nature, but does not control or change experimental components.

b. Experimental

- i. Experimenter constructs a research design for testing a hypothesis.
- ii. Theory/experimental components are determined and operationalized
- iii. Experimenter controls antecedents to data collection

2. Uncontrolled versus Controlled Research

a. Uncontrolled

Observations are made of a single group (perhaps a single person) experiencing the experimental condition

b. Controlled

Observations are made of one or more groups experiencing the experimental condition -- and a group observed without experiencing the experimental condition.

The control-group data serves as a reference against which the experimental-group data is compared.

3. Retrospective versus Prospective Research

a. Retrospective

A researcher examines existing data. These are observations that have been recorded under non-experimental (natural) conditions.

b. Prospective

A research constructs an experiment and collects data under controlled conditions.

4. Random assignment versus Non-Random Assignment

a. Non Random Assignment

Participants are assigned group membership by convenience or some other method other than random assignment

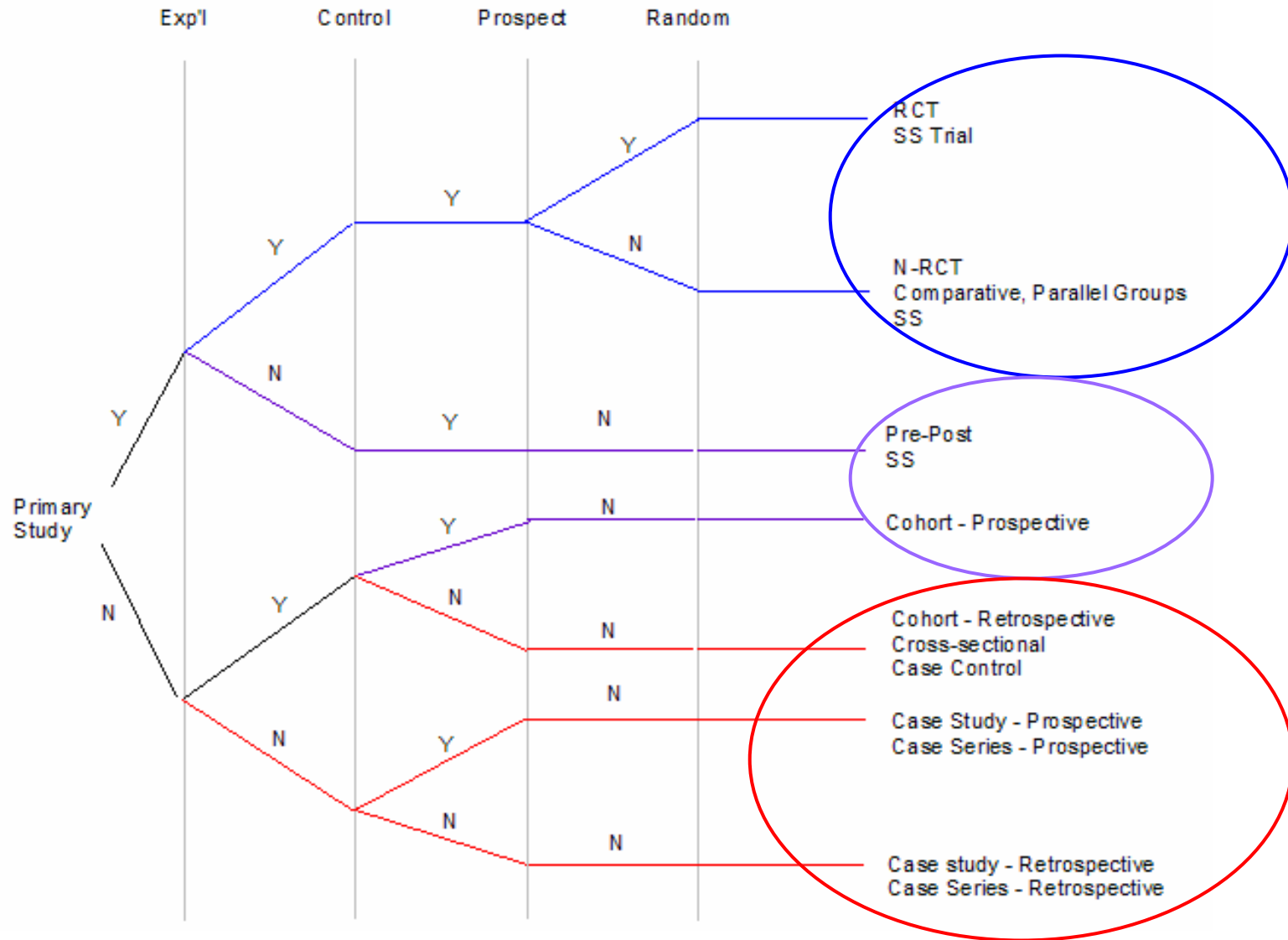
b. Random Assignment

For each participant, group assignment is strictly a matter of true randomness.

If we apply those dichotomous groupings serially to primary studies in a flow-chart like fashion, a hierarchy of studies results with designs having more built-in rigor at the top. Only a stringent application of quality indicators can tell us if that rigor is indeed present and just how valid the study really is.

Notice that three strata become obvious. The top tier of studies are various forms of clinical trials. The middle tier consists of research designs mostly used in effectiveness research. The lower tier of studies comprises designs mostly used in pre-trial research.

With some thought, it becomes clear that the quality indicators for each tier are different. Evaluating a study in one tier (e.g., effectiveness research) using quality indicators for another tier (e.g., clinical trials or efficacy research) makes no sense and is indeed counterproductive. What we need is a system for identifying the purpose of a study and then assess the quality with which that purpose is achieved.



5. Q: How do we know when a primary study is a clinical trial?
 A: If we can place a check in every un-shaded box in the following form.

Identifying a Clinical Trial

Study ID: _____

Defining Quality	Yes	No
The intervention protocol (IP) is clearly established <i>a priori</i>		
This is a paper about establishing/developing the theoretical basis for a prospective IP		
This is a paper about inventing, uncovering, or developing the IP		
The outcome measure(s) is(are) set and understood		
This is a measurement or psychometric study		
The clinical population is clearly defined		
This study assesses the magnitude of efficacy or safety (or both) of the intervention protocol for the first time, or in confirmation, or in some new dimension of efficacy or safety (e.g., direct treatment effect at the impairment level, generalization treatment effect at the activity-limitation level, QoL, adverse effects, harm, specificity, LR+, post-test probabilities)		
The study comprises an experimental sample and a control sample obtained through one or more service delivery sites		
The experimental and control conditions are administered through specially trained and qualified staff.		
The observed effect is attributed to the protocol and is generalized to the clinical population through statistical inference		
The study reports a group research design (e.g., cross over, parallel groups) or a single-subject research design (min: min. ABA with multiple baselines across token type or treatment effect) replications across participants		

6. Q: How do we know when a primary study reports effectiveness research?
 A: If we can place a check in every un-shaded box in the following form.

Identifying Effectiveness Research

Study ID: _____

Defining Quality	Yes	No
The intervention protocol (IP) is clearly established <i>a priori</i>		
This is a paper about establishing/developing the theoretical basis for a prospective IP		
This is a paper about inventing, uncovering, or developing the IP		
The outcome measure(s) is(are) set and understood		
This is a measurement or psychometric study		
The clinical population is clearly defined		
This study assesses the magnitude of effectiveness or safety (or both) of the IP in some dimension (e.g., direct treatment effect at the impairment level, generalization treatment effect at the activity-limitation level, QoL, adverse effects, harm, specificity, LR+, post-test probabilities)		
The study includes a rather large experimental sample		
The IP is administered as part of regular services provided through a clinical practice		
The observed effect is attributed to the protocol and is generalized to the clinical population through statistical inference		
The study reports a pre-post research design (elaborations are possible) or a single-subject research design (min: min. ABA with multiple baselines across token type or treatment effect) replications across participants		

7. Q: How do we know if a primary study is an instance of pre-trial research?
A: When neither form can be completed with a check in every un-shaded box.
8. Now, projecting forward, is it a reasonable thing to consider research designs as they relate to their intended purpose and then assess their validity in meeting that purpose?
Is it within our means to realize that possible future?

F. Conclusion

1. Clinical practice can inform clinical research.
2. Clinical research can be optimized for informing clinical practice.
3. The processes, and the products coming from them, are imaginable and within our means.

References

- Beeson, P. M., & Robey, R. R. (2006). Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. *Neuropsychology Review*, 16, 161-169.
- Dollaghan, C. A. (2007). *The Handbook for Evidence-Based Practice in Communication Disorders*. Baltimore, MD: Paul H. Brookes Publishing Co.,
- Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar F, Mahon J, Chan K, Wells G. (2002). Determining the clinical importance of study results: A review. *Journal of General Internal Medicine*, ;17, 469-476.
- Robey, R. R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia, *Journal of Speech, Language, and Hearing Research*, 41, 172-187.
- Straus, S. E., Haynes, R. B., Richardson, W. S., & Glasziou, P. (2005). *Evidence Based Medicine*. London: Elsevier Health Sciences
- Thompson, B. (2002). *What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes*. *Educational Researcher*, 31, 25-32.