

Questions to Ask When Reviewing Empirical Studies

Rationale

Hypothesis clearly stated and operationalized. Is it falsifiable?
Congruent with prior studies
Should more basic questions be answered first
Literature review focused, comprehensive (including in press, presentations), balanced and does it interpret results too severely or too leniently
Alternate explanations considered
Novelty, significance, magnitude, of likely impact

Participants

How selected/recruited
Generalizability: selection bias, groups over/underrepresented
Inclusion criteria too strict/lenient
What is group of interest
What are S expectancies
Flowchart of S attrition (see CONSORT)
Sample size adequacy – is estimated effect size realistic, base rates realistic
Inclusion criteria (too strict/lenient), group of interest?

Design/Procedures

Need intro summary of design?
Control groups: placebo, no drug, historical, multiple baseline, usual care, optimal care, standard care
Positive control group
Historical, epidemiological base rates
What is control group or base rates
Other control/comparison groups possible
Can it be conceptualized factorially
Able to do within-Ss
Adequacy of test – esp if obtain negative results
How does test map onto hypothesis – are constructs well-operationalized
Measure possible confounds/moderators, mediators
Time, order or repeated testing effects
Biases if not random assignment
Groups differ on variables other than variable of interest
Regression to the mean possible explanation
Diffusion, demoralization possible
Is intervention well-operationalized
Alternative explanations

Human Ss

Safety, especially for subpopulations
Confidentiality

Ethics of control group
S concerns
Steps to minimize risks
Alternate treatments
Is control consistent with usual care
Debriefing used

Interventions

Adequate dose, duration and timing of intervention
Use of multiple doses
Adequate training and monitoring of therapists
Practicality
Both conditions same emphasis/quality
Monitor compliance
Blindness maintained
Sufficiently well described

Measures

Designated major outcome
If multiple outcomes are they expected to be convergent; how to handle inconsistent results across outcomes
How well operationalized
Actual behavior>self-report of behavior>subjective reports>intentions, attributions
What is not being measured
How well do measures map onto construct
Process/mechanism measures, adequacy of proxy measures
Blinding of assessments
Test-retest and interrater reliability, predictive validity, sensitivity
Use measures hypothesized not to change as specificity test
What is likely effect of demand bias and political correctness
Measure functional status
Are questions ambiguous
Measure possible confounds stability of measures

Results

Clear, a priori criterion for success
Stats tied to hypotheses
Magnitude of effect- statistical vs clinical significance
Anticipated dropout rate- reasons for dropouts
How much missing data and how handled
Carryover effects,
Type of scale – continuous, ordinal, categorical, nominal
Chop up continuous to ordinal/categorical and lose power?
Distribution for stats
Covariates included
Show raw data

Is nominal made to seem ordinal
Sufficient variability for correlational analyses
Linear vs curvilinear vs threshold effects
Adverse events – clinically significant, cause dropouts
Amount of overlap of groups
Intent-to-tx, when designated as S

Interpretation

Cross sectional association, prospective prediction, exp manipulation vs claims of causality
Relate to others work
Is it actually a conceptual replication of prior work?
Qualifiers needed
Alternate explanations
Significance for explaining vs intervening Magnitude/clinical significance
Replicated by others
Convergent validity
Consistency across Ss
External validity
Sufficient vs necessary
Consistent with other data
Adequate test

Significance

Does it Increase understanding
Will it Improve public health/clinical outcomes
How big of an impact will it have
Magnitude of problem
Is it innovative