Single-Subject Research Design

Recommendations for Levels of Evidence and Quality Rating

Who?
- Lynne Romeiser Logan PT, MA, PCS
  - Long time member of Outcomes committee of AACPDM. Co-author of 2 systematic reviews on ITB
- Robbin Hickman PT, DSc, PCS
  - Member of Outcomes committee
- Susan Harris PhD, PT, FAPTA, FCAHS
  - Author of several studies using SSRD
- Carolyn Heriza EdD, PT, FAPTA
  - Author of several studies using SSRD

Objectives for Today
- Identify the 6 types of SSRD.
- Assign levels of evidence to each type of SSRD.
- Rate the quality of 1 example of SSRD.
- Understand the relevance of SSRD within the context of a systematic review.

Outline for Today
- 10 minutes: Background
  - Lynne Romeiser Logan
  - Course Purpose/Outline
  - Single Subject Research Design (SSRD) definitions and context in clinical research
  - Purpose of levels of evidence and quality questions for SSRD
- 30 minutes: Review of SSRD including levels of evidence for each type
  - Presenter: Susan Harris
- 20 minutes: Discussion of Quality Questions for SSRD
  - Review and discuss 2 examples of SSRD
  - Presenter: Robbin Hickman
- 40 minutes: Review SSRD paper in small groups (All)
  - Review essential components
  - Assign level of evidence
  - Assign quality score
- 10 minutes: Summary and Questions

What?
- Levels of evidence for SSRD (analogous to Sackett’s levels for group design research)
- Quality/rigor analysis to judge internal validity of SSRD (analogous to Jadad and van Tulder criteria)

Why?
- Aid in critical review of these studies for:
  - Application to practice
  - Inclusion in systematic reviews of evidence
  - Improvement of quality and clarity of reporting and design of SSRD

References:
1. Centre for Evidence-Based Medicine, Oxford University. www.cebm.net
What SSRD is NOT

- Case Report
  - Carefully reported example of new intervention, diagnosis, or combination of treatments
  - No experimental design
- SSRD
  - Tightly controlled experimental design
  - Must demonstrate stability of elements of interest before beginning intervention
  - May be randomized
  - Often includes more than a single subject

What SSRD IS

- Within subject methods
- Repeated measures designs
- Intrasubject replication designs
- N of 1 randomized controlled trial

What to look for in SSRD

- Baseline (A phases)
  - Outcome of interest is measured repeatedly until a stable state can be demonstrated
- Intervention phases (B, C, D phases)
  - Experimental treatment is applied and outcomes continue to be measured
- Phases are repeated in order to show that intervention has had an effect
  - ABAB for example

Quality Concerns

- Reliability
  - Accuracy
  - Reproducibility
  - Must be checked prior to and during study
  - Must be reported!
- Validity
  - Confidence that you measuring what you intend to measure
  - Test must actually measure outcome of interest
  - Face validity
    - logical
  - Construct validity
    - Compared to other similar measures

Internal/External Validity

- Internal Validity
  - Can demonstrate causal relationship between independent and dependent measures
  - Intervention caused change in measurement of outcomes
- External Validity
  - Extent to which results can be generalized beyond the subjects in the study
  - Similarity of subjects/setting/procedures to your population of interest
  - Number of subjects
Quality Questions/Rigor Criteria for SSRD

♦ Robbin A. Hickman, PT, DSc, PCS
Assistant Professor
University of Nevada Las Vegas

Objectives
- Describe & discuss quality questions
- Discuss scoring of quality questions
- Illustrate use with example

Referenced article

Group scale analogy
- Levels I – V assigned according to design
- Quality ratings complement design analysis
- Letter grade assigned according to rigor and study components

How Were Questions Determined & Analyzed?
- Reviewed published articles & texts
  - Questions & criteria used to evaluate group designs
  - SSRD methods
  - Questions & criteria used to evaluate SSRD

Question Design & Analysis
- Three versions of questions analyzed for inter-agreement analyses among 4 raters
  - Raters reviewed 3 published articles for each version
  - Post inter-rater reliability, versions 1 & 2 reviewed
    - Questions omitted
    - Questions refined
    - New questions added
    - Questions prioritized
**Versions of Questions**
- Version one
  - 18 questions
  - Inter-agreement analysis on methodological strength (weak, moderate, or strong) 67%
- Version two
  - 22 questions
  - Inter-agreement analysis on methodological strength (weak, moderate, or strong) 65%
- Version three
  - 14 questions
  - Inter-agreement analysis on methodological strength (weak, moderate, or strong) 75%

**SSRD Quality & Rigor Questions**
- Provide means of evaluating study parameters beyond design
- 14 questions
- Each scored yes/no
- Grading scheme:
  - 11-14 yes = Strong
  - 7-10 yes = Moderate
  - < 7 yes = Weak

**Quality Questions**
- Examine 5 study parameters; each receives subtotal score
  - Description of participants & settings
  - Independent variable
  - Dependent variable
  - Design
  - Analysis

**Scoring Quality Questions**
- Each question is scored Yes or No
  - Each Yes question assigned 1 point
  - Two questions are two-part questions, each part assigned 0.5 point

**Scoring Continued**
- Total # of yes questions added per section
- Total # of yes questions added for all sections

**Scoring Quality Questions**
- Perfect score=14
- Overall score ratings
  - Strong = 11-14 questions answered yes
  - Moderate = 7-10 questions answered yes
  - Weak = <7 questions answered yes
So, as we work through this example,

- If you believe the criteria have been met, please score a yes next to the question.
- If you don’t believe the criteria have been met, your score for that item will be zero.

### Description of Participants & Settings

1. Was/were the participant(s) sufficiently well described to allow comparison with other studies or with the reader’s own patient population?


#### Table 1: Characteristics of Individual

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Subject 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>15</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>White</td>
<td>White</td>
<td>White</td>
</tr>
<tr>
<td>Hemoglobin level</td>
<td>10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Type of pain</td>
<td>Severe</td>
<td>Mild</td>
<td>Mild</td>
</tr>
<tr>
<td>Ambulation age</td>
<td>3 years</td>
<td>4 years</td>
<td>3 years</td>
</tr>
<tr>
<td>Waddell’s activity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Walking</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Running</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sitting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

#### Independent Variable (IV)

2. Were the independent variables operationally defined to enable replication?
3. Were intervention conditions operationally defined to enable replication?

**IV** = *what is manipulated, controlled or observed by researcher (e.g., intervention of interest)*


- **IV** = See purpose paragraph just above methods section, p 571
- See also Methods section, subheading Procedure, p 571

### Dependent Variable (DV)

4. Were the dependent variables operationally defined as dependent measures?
5. Was inter-rater or intra-rater reliability of the dependent measures assessed prior to the study & during each phase of the study?

**DV** = *response variable or measured outcome; used to determine effects of IV*
6. Was the outcome assessor unaware of the phrase of the study (intervention vs. control) in which the participant was involved?

7. Was stability of the data demonstrated in baseline, i.e. lack of variability or a trend opposite to the direction one would expect following application of the intervention?

**Example: Franks, et al, 1991**

*See Methods, subheadings Instrumentation & Procedure pp. 572-3*

<table>
<thead>
<tr>
<th>DV</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading fluency</td>
<td>Operationally defined?</td>
</tr>
<tr>
<td>Visuomotor accuracy</td>
<td>Reliability in each phase? Examiner blinded?</td>
</tr>
<tr>
<td>Manual dexterity</td>
<td>Stability of data in baseline?</td>
</tr>
</tbody>
</table>

**Graphs of stability:** Fertel-Daly, et al, 2001

- **Unstable baseline data**
- **Stable data**

**Stable visuomotor data? p. 575**

**Stable manual dexterity scores? P. 576**
Design

8. Was the type of SSRD clearly & correctly stated e.g. AB, multiple baseline across subjects?

9. Were there an adequate number of data points in each phase for every subject (minimum of 5)?

10. Were the effects of the intervention replicated across 3 or more subjects?

Example: Multiple baseline

Dannemiller L. RMUoHP Dissertation. 2004

Alternating condition designs

Ottenbacher, p. 123 – Alternating treatment design

Analysis

11. Did the authors report visual analysis, e.g. level, trend, & variability?
12. Did the graphs for visual analysis follow standard conventions, e.g.,
- x & y axes labeled clearly & logically
- phases clearly labeled (A, B, etc) & delineated with vertical lines
- data paths separated between phases
- consistency of scales?

13. Did the authors report test of statistical analysis, e.g., celeration line approach, two-SD band method, C-statistic or other?

14. Were all criteria met for the tests of statistical used?

Analysis examples

Backman et al, 1997

Analysis: 2 standard deviation band


Example: Beauregard et al, 1998

Celeration line

Additional analysis info: Franks, et al p 573-575

Table 2. Descriptive Statistics for Each Phase of the Study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Reading latency</td>
<td>156.2</td>
<td>9.1</td>
<td>118.7</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>179.4</td>
<td>2.83</td>
<td>181.0</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>Vertical latency</td>
<td>222</td>
<td>2.7</td>
<td>228</td>
<td>3.4</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Reading latency</td>
<td>116.0</td>
<td>8.5</td>
<td>116.0</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>199.1</td>
<td>2.29</td>
<td>154.9</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Vertical latency</td>
<td>21.6</td>
<td>1.7</td>
<td>21.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Reading latency</td>
<td>114.4</td>
<td>14.4</td>
<td>114.2</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>176.5</td>
<td>3.08</td>
<td>191.1</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>Vertical latency</td>
<td>28.0</td>
<td>8.4</td>
<td>24.0</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Let's add up your scores

- Count up yes answers
- What's your score?
- What are our scores?
- Ranking?

Summary

- 14 questions designed to evaluate SSRD quality & rigor
- Yield rating:
  - Strong (11-14)
  - Moderate (7-10)
  - Weak (<7)

Summary Continued

- Areas examined:
  - Description of participants & setting
  - IV
  - DV
  - Design
  - Analysis

References

- Pearson Educational, Inc.
- Centre for Evidence-Based Medicine, Oxford University. www.cebm.net. Accessed on November 11, 2006
- Dannemiller L. RMUoHP Dissertation. 2004
**Single Subject Research Designs & Evidence Levels**

Carolyn B. Heriza, PT, EdD, FAPTA  
Professor, Doctor of Science, Pediatrics  
Rocky Mountain University of Health Professions  
&  
Susan R. Harris, PT, PhD, FAPTA, FCAHS  
Professor, Department of Physical Therapy, Faculty of Medicine  
University of British Columbia

**Objectives**

- Identify 6 different types of SSRD  
- Assign levels of evidence to each type of SSRD

**Single Subject Research Design (SSRD)**

Involves studying a single individual or individuals by taking repeated measures of one or more dependent variables & systematically applying & sometimes withdrawing or varying the independent variable.


**Major Components of SSRD**

- Use of frequent & repeated outcome measures  
- Sequential application & withdrawal or variation of intervention

**Basic Components of SSRD**

- **Phases**  
  - Baseline phase (A): initial period of observation in which the natural frequency of occurrence of the behavior obtained  
  - Intervention phase (B): once stable baseline obtained, intervention introduced & frequency of behavior continues to be measured  
- **Blind data collection**  
  - Strength of design greater when data collectors unaware of phase of intervention in which data collected  
- **Generalizability**  
  - Increased when SSRDs replicated across patients/clients by other therapists & in other settings

**Characteristics of SSRD**

- Target behaviors clearly specified & operationally defined  
- Methods of measurements precisely defined  
- Repeated measures of target behavior taken throughout each phase of study
Characteristics of SSRD

- One intervention strategy (independent variable) manipulated at a time
- Extraneous variables carefully controlled
- Methods used in data collection must be replicable & reliable

Types of SSRD

- Within series
  - AB, ABA, ABAB
- Between series
  - Alternating treating design (ATD)
- Combined series
  - Multiple baseline design (MBD)


Hypothetical Question

What is the effect of constraint-induced movement therapy on the frequency of right hand use?


- Simplest & least sophisticated analysis technique in which baseline (A) is established first followed by intervention phase (B)
- Least powerful SSRD as cannot be sure that a confounding variable did not cause the behavior change

Hypothetical Question

What effect does vibratory stimulation have on duration of head raising in the prone position?

Figure from Harris SR, 2005

- Baseline (A1) followed by introduction of intervention phase (B) & then return to baseline (A2)
- If intervention effective, desired behavior improves during B phase & gets worse when intervention withdrawn (A2)
- Disadvantage: terminates in a baseline phase

Hypothetical Question

What is the effect of muscle strengthening & endurance training on the number of lengths of parallel bars walked?

Figure from Harris SR, 2005

- Baseline (A1) followed by intervention (B1), intervention withdrawn (A2) & intervention reinstated (B2)
- Advantage: terminates in an intervention phase

Hypothetical Question

What is the effect of verbal/behavioral intervention in comparison to tapping the chin on the percentage of mouth closures?

Figure from Harris SR, 2005

- Two or more interventions rapidly alternated in random fashion & effects compared on the dependent variable within a single phase
- Most powerful when baseline condition provided prior to & following intervention phase
- Advantage: Alternations in intervention phase may occur over brief period in contrast to ABAB and multiple-baseline (MBD) designs
**Multiple Baseline Design**

- Can be across subjects, settings, or behaviors
- Can be either concurrent or non-concurrent with the former being the stronger of the two
- Random assignment most powerful

Advantage: Change in target behavior occurring during intervention for subject 1 (W1) compared to subject 1’s baseline & to baselines of other 2 subjects (B2, B3); change during intervention for subject 2 compared to subject’s 2 baseline (W4) & to baseline of subject 3 (B5); change during intervention for subject 3 compared to subject 3’s baseline (W6)


---

**MBD Across Behaviors**

The Question
Do changes in the environment and/or the task improve performance of functional tasks which are in transition?


---

**N-of-1 Randomized Controlled Trial**

- Essentially an SSRD
  - SSRD arose from research in behavioral sciences
  - N-of-1 trials arose from the medical literature
- Most examples of N-of-1 trials address effectiveness of drugs but could be applied to other types of discrete interventions that do not have a carry-over effect

Design of N-of-1 Trials

- As with SSRD, used to select optimal intervention on patient/client
- Trial of 2 different interventions or an intervention & placebo
- Interventions must be quick acting & effect must be reversible or cease when placebo introduced or intervention discontinued
- Patient/clinician & clinician/evaluator blind to interventions administered


---

**MBD Across Behaviors**

The Question
Do changes in the environment and/or the task improve performance of functional tasks which are in transition?


---

**Design of N-of-1 Trials**

The Question
Does a patient with chronic airflow limitations have better better QOL & exercise capacity when on theophylline than when on a placebo?


---

**Design of N-of-1 Trials**

Which medication, NSAIDs (treatment X) or paracetamol (treatment Y), provides greater pain relief in patients with osteoarthritis of the hip or knee?

Outcomes of N-of-1 Trials

- Requires measurable target behavior or symptom
- Examples of target behaviors: level of pain on VAS, # of episodes of specific system/day, or distance walked without SOB
- Generalizability enhanced through repetition

Levels of Evidence for Group Designs

- I Large randomized controlled trials with clear-cut results (and low risk of error)
- II Small randomized controlled trials with uncertain results (and moderate to high risk of error)
- III Non-randomized concurrent cohort comparisons

- IV Non-randomized historical cohort comparisons
- V Case series without controls

Proposed Level of Evidence for Single Subject Research Designs: Level I

- Designs
  - Randomized controlled N-of-1 RCT, ATD, MBD with clear-cut results
- Generalizability
  - ATD replicated across 3 or more subjects
  - MBD consists of minimum of 3 subjects, behaviors, settings
- Causal inference
  - Provides

Proposed Level of Evidence for Single Subject Research Designs: Level II

- Designs
  - Non-randomized, controlled, concurrent MBD with clear cut results
- Generalizability
  - Minimum or 3 subjects, behaviors, or settings
- Causal inference
  - Limited

Proposed Levels of Evidence for Single Subject Research Designs: Level III

• Designs
  – Non-randomized, non-concurrent, controlled MBD with clear-cut results
• Generalizability
  – Minimum of 3 subjects, behaviors, or settings
• Causal inference
  – Limited


Proposed Levels of Evidence for Single Subject Research Designs: Level IV

• Designs
  – Non-randomized controlled experimental designs with at least 3 phases (ABA, ABAB, BAB, etc.) with clear-cut results
• Generalizability
  – Replicated across 3 or more different subjects
• Causal inference
  – Only hints at


Proposed Levels of Evidence for Single Subject Research Designs: Level V

• Designs
  – Non-randomized controlled AB single subject research design with clear-cut results
• Generalizability
  – Replicated across 3 or more different subjects
• Causal inference
  – Only suggestive allowing for testing of ideas


Summary
Level of Evidence: ______

<table>
<thead>
<tr>
<th>Level</th>
<th>Design Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>N-of-1 RCT, ATD, randomized multiple baseline (concurrent or non-concurrent)</td>
</tr>
<tr>
<td>II</td>
<td>Multiple baseline (concurrent across subjects)</td>
</tr>
<tr>
<td>III</td>
<td>Multiple baseline (non-concurrent across subjects)</td>
</tr>
<tr>
<td>IV</td>
<td>Replicated, controlled basic experimental designs with at least 3 phases (ABA, ABAB, BAB, etc.)</td>
</tr>
<tr>
<td>V</td>
<td>A-B design</td>
</tr>
</tbody>
</table>

Score each question as YES or NO:

DESCRIPTION OF PARTICIPANTS AND SETTINGS
1. Was/were the participant(s) sufficiently well described to allow comparison with other studies or with the reader's own patient population?

Count total number of “YES” responses: ______

INDEPENDENT VARIABLE
2. Were the independent variables operationally defined to enable replication?
3. Were intervention conditions operationally defined to enable replication?

Count total number of “YES” responses: ______

DEPENDENT VARIABLE
4. Were the dependent variables operationally defined as dependent measures?
5. Was inter-rater or intra-rater reliability of the dependent measures assessed prior to the study (0.5 point) and during each phase of the study (0.5 point)?
6. Was the outcome assessor unaware of the phase of the study (intervention vs. control) in which the participant was involved?
7. Was stability of the data demonstrated in baseline, i.e. lack of variability or a trend opposite to the direction one would expect following application of the intervention?

Count total number of “YES” responses: ______
DESIGN
8. Was the type of SSRD clearly (0.5 point) and correctly stated (0.5 point), e.g. AB, multiple baseline across subjects?
9. Were there an adequate number of data points in each phase for every subject (minimum of 5)?
10. Were the effects of the intervention replicated across 3 or more subjects?

Count total number of “YES” responses: _________

ANALYSIS
11. Did the authors report visual analysis, e.g. level, trend, and variability?
12. Did the graphs for visual analysis follow standard conventions, e.g., x and y-axes labeled clearly and logically, phases clearly labeled (A, B, etc.) and delineated with vertical lines, data paths separated between phases, consistency of scales?
13. Did the authors report tests of statistical analysis, e.g., celeration line approach, two-SD band method, C-statistic or other?
14. Were all criteria met for the tests of statistical analysis used?*

Count total number of “YES” responses: _________

Total score out of a possible 14: ______________

*To assist in answering questions #8 and #14, consult a SSRD text as below.


Overall Scores: Scoring of this quality rating is simply done by counting “yes” answers. Strong = 11-14, moderate = 7-10, weak = <7.
SINGLE SUBJECT RESEARCH REFERENCES

Susan R. Harris, Ph.D., P.T., Mark Wolery, Ph.D., Jean Deitz, Ph.D., O.T.R.,
Catherine Backman, M.S., OT(C), Carolyn Heriza EdD, PT.

Texts and Chapters on Single Subject Research Methodology:

Journal Articles on Single Subject Methodology:


Description of Specific Single-Subject Designs:

Analysis and Interpretation of Single Subject Designs:

Measurement and Reliability:


**Single Subject Research Examples in Rehabilitation:**


