SINGLE-CASE RESEARCH

And Small N Designs

Relevant History

- In last half of nineteenth century, researchers more often looked at individual behavior (idiographic approach)
- Founders of psychological research took this approach
  - Ebbinghaus: Studied experimental memory
  - Wundt: Studied self-perceptions of consciousness
  - Skinner: Developed operant conditioning techniques

Relevant History

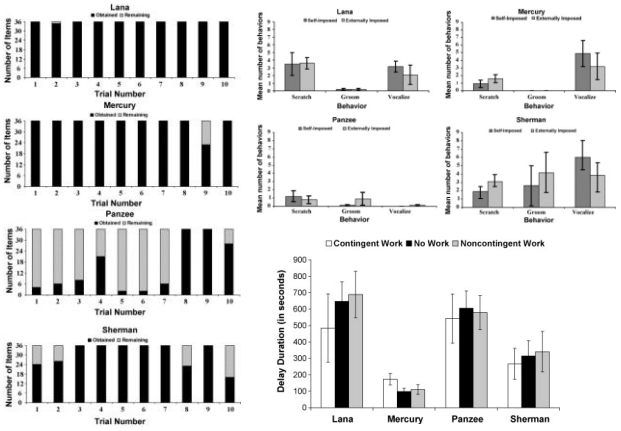
- By early twentieth century, focus changed
- Most contemporary research takes a group comparison approach
  - Exemplified by experimental and correlational research strategies
  - Looks at average behavior of groups (nomothetic perspective)
  - Aims to establish general principles and broad generalizations that apply across individuals

Relevant History

- However, single case research continues, especially in areas of
  - sensory and perceptual processes
  - clinical treatment research
  - comparative research
  - interest in individual differences
- Over time, methodology has improved
  - Researchers now emphasize control

Importance of Exceptions to Research Findings

- There are always exceptions to any particular finding!
- Behavioral science is probabilistic.
- Research findings uncover generalities and trends.
- Exceptions do not invalidate research findings, but should they be ignored?



Arguments for and Against Group Designs and Analyses

(1) Error Variance

- **Group design argument**
  - Averaging across participants provides a more accurate estimate of a variable's general effect
  - Group designs allow us to estimate the amount of error variance in our data
- **Single-case argument**
  - Error variance is partly created by averaging over participants in a group design (**interparticipant variance**)
  - Researchers using group designs ignore the "real" error (**intraparticipant)** variance within the participant

Arguments for and Against Group Designs and Analyses

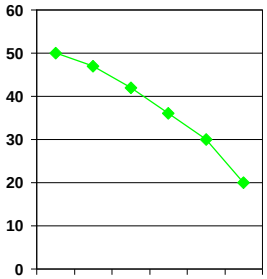
(2) Generalizability

- **Group design argument** – averaging the scores of several participants reduces the idiosyncratic responses of any one participant to show the general effect
- **Single-case argument** – averaging responses may not accurately describe any particular participant's responses

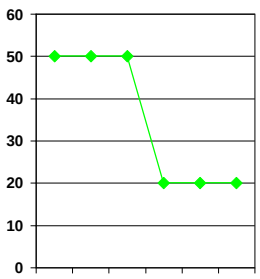
Subject	Training	Transfer
Pende	75	38
Chip	85	82
Kongo	80	90
Average	80	70

Example: Learning Curves

Result of Averaging Across Participants



An Individual Participant



Arguments for and Against Group Designs and Analyses

(3) Reliability

- **Group design argument** – reliability of findings is established by replicating studies
- **Single-case argument** – reliability of findings should be established via:
  - **Intraparticipant replication** – replicating the effects of the independent variable with a single participant
  - **Interparticipant replication** – seeing whether the effects obtained for one participant generalize to other participants in the same study

Arguments Against Group Designs and Analyses

Concerns

- about the ethics of withholding treatment from control groups
- that, for some diagnoses, too few participants are available for group comparison research
- that the individual becomes lost in the group average
- that group research rarely examines patterns of change over time

- Concerns led to renewed interest in single case research
- Contemporary single case research most often takes a behaviorist approach
  - Behavior therapy
  - Behavior modification
  - Applied behavior analysis
- Much of the research focuses on behavioral treatment of clinical disorders
- Approach also used in other subdisciplines (e.g., cognitive, developmental, organizational)

Single-Case Research

- Is often the only tool available for studying rare phenomena
- Can provide depth of understanding through its longitudinal approach
  - Especially if environmental, social, and historical contexts of behavior are considered
- Can identify cases that show limitations of general theories
- Can provide hypotheses for testing with other methodologies

Validity Problems

- Due to its longitudinal nature and lack of control, single-case research is especially vulnerable to:
  - history threats
  - maturation threats
- Clinical studies using extreme cases are vulnerable to statistical regression
- Problems can be addressed with careful planning

Measurement Criteria

Objectivity: High quality single-case research uses formal, objective measures of DV, including

- multiple DVs
  - If multiple DVs all change in predicted manner, it is less likely that result was due to chance or a confound

Measurement Criteria

Study quality also increases when there are

- multiple measures of each DV
  - Confidence is increased if multiple sources of evidence point to same conclusion
- frequent assessment of DVs
  - Assessments can be made before, during, and after an intervention
    - Change should be associated only with intervention
  - Helps rule out alternative explanations, such as maturation

Control Criteria

- Can create analog to experimental research in single-case research
  - The test case shows what happens when IV is present
  - The control case shows what happens in absence of IV
  - Comparing test and control case helps rule out threats to internal validity
    - May need more than one control case

Replication Criteria

- In single case research, replication cases should be as heterogeneous as possible
  - Demonstrates robustness of phenomenon
  - Failures to replicate can determine theory's boundary conditions
- If hypothesis is supported across heterogeneous cases, results are more generalizable

Impact Criteria

- In treatment-outcome research, the magnitude of the impact can indicate whether threats to internal validity are plausible
  - The greater the treatment impact, the less likely change is due to threats to history, maturation, and statistical regression
- Treatment is more likely to be cause of change if
  - a chronic rather than an acute problem is addressed
  - the treatment has an immediate rather than delayed impact
  - follow-up assessments show treatment continues to have an effect

Treatment Criteria

- Validity of intervention research improved when researcher has greater control over treatment
- Control is greater when treatment
  - is manipulated (versus observation of naturally occurring treatment)
  - onset can be controlled
  - is standardized
  - is implemented according to a set protocol

Evaluation Criteria for Selecting Cases to Study

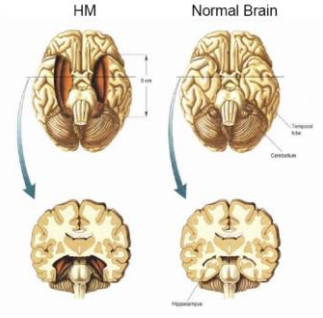
- Look for situations in which it is possible to manipulate the IV
  - If not possible, look for cases that best match your operational definition of the IV
- For replication, choose test cases as *different* as possible
- But choose *control cases* that are as similar to test cases as possible

Evaluation Criteria for Selecting Cases to Study

- Consider how much access you will have during data collection
- Better to have access
  - for continuous assessment
  - to multiple sources of information
  - for proper follow up

Two Types of Single-Case Studies

- Single-case experimental designs
- Case studies



Single-Case/Small N Experimental Designs

- Systematic procedure for testing changes in behavior
- Unit of analysis is the individual participant
  - More than one participant may be studied, but their responses are analyzed individually
  - Generally involve between 1-9 participants
- Difficult to analyze these data with inferential statistics such as t-tests and F-tests
- More flexible than traditional study
  - Require continuous assessment of participant
- Often used in clinical cases
  - Psychophysiological processes; effects of drugs
  - Behavior modification – techniques for changing problem behaviors based on operant conditioning

Measuring Targets of Intervention

- DV should be the target of the intervention
- Can be measured simultaneously or sequentially
- Measures of behavior are often categorized according to:
  1. Frequency = how often behavior occurs
  2. Duration = how long behavior lasts
  3. Interval = time between episodes
  4. Magnitude = intensity of behavioral event

Components of Small-N Designs

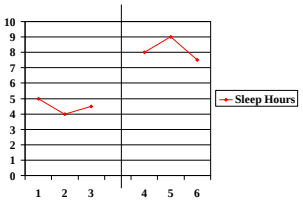
1. Repeated measurement of the dependent variable
   - If preintervention measurements cannot be taken, retrospective data may be used.
2. Baseline phase (A)
   - Intervention not offered to subject
   - Acts in place of a "control group"
   - Repeated measurements of the DV are taken until a pattern emerges
3. Treatment phase(s) (B)
   - Intervention is implemented
   - Repeated measurements of the DV are taken
   - Should be as long as the baseline phase

Phases and Phase Changes

- Series of observations made under same conditions
- Baseline (A) - absence of treatment
- Treatment (B) – during treatment
- (C and D) = other treatments
- Modifications = B1, B2….
- BC – phase involving combination of treatments B & C
- Min. 3 observations in Phase

Evaluating Results

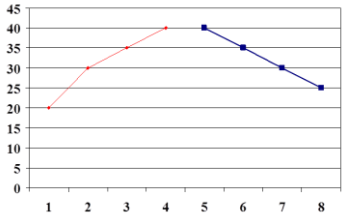
Graphic display
- Facilitates monitoring and evaluating the impact of the intervention
- No control over extraneous variables
- Assessing practical (clinical) significance is of primary importance
  - Set criteria for success with individual or community
  - Use clinical cut-off scores
  - Weigh costs and benefits of producing the change



Sleep Treatment Introduced

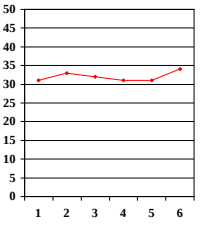
Trends

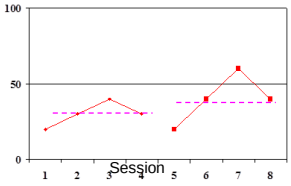
- Direction in the pattern of the data points
- Consistent increase or decrease in magnitude of behavior across phase



Levels

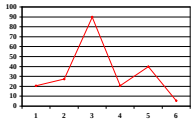
- Level =
  - magnitude of participant's responses
  - magnitude of the target variable; typically used when the observations fall along relatively stable lines
- Must be clear pattern WITHIN a phase
- Then show that patterns change from one phase to the next

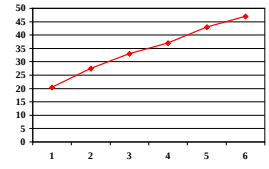


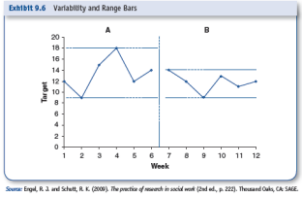


Examination of Variability

- Variability = how different or divergent the scores are within a baseline or intervention phase
- Stability = straight line with only minor deviations
- Unstable – large differences/high variability from one observation to the next

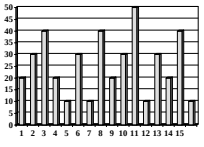


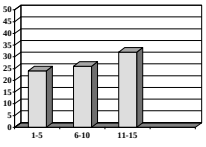




Dealing with Unstable Data

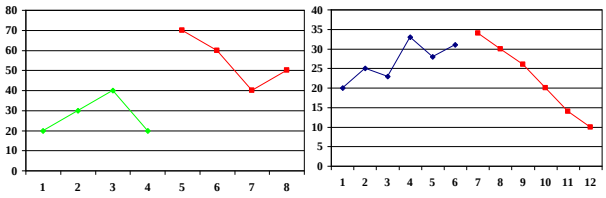
- Keep observing and hope data will stabilize
- Average a set of observations
- Look for pattern within inconsistency
  - Morning sessions differ from afternoon sessions
  - Id and control extraneous variables





Immediate change in level

Latency of change



Changing Phases

- Phase change = manipulation of IV
- Implementing, withdrawing or changing a treatment
- Look for change in pattern of behavior
- Do NOT introduce treatment if baseline phase shows trend toward improvement
- DO introduce treatment early if behaviors are reaching dangerous levels in baseline
- STOP treatment early if negative effects apparent

Basic Design (A-B)

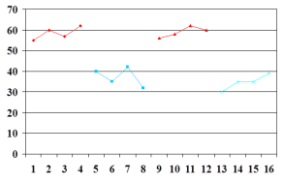
- Baseline phase (A) with repeated measurements and an intervention phase (B) continuing the same measures
- Fluctuations are difficult to interpret
- Cannot rule out other extraneous events, so causality cannot be established

Withdrawal Designs

- Intervention is concluded or is temporarily stopped during the study
- A-B-A Design
  - Behavior is measured (Baseline period; A)
  - Independent variable is introduced (B)
  - Behavior is measured (A)
  - Includes post-treatment follow-up
  - Follow-up period should include multiple measures

Withdrawal Designs (cont.)

- A-B-A-B Reversal Design
  - Adds second intervention phase that is identical to the first
  - Replication of treatment phase reduces the possibility that an event or history explains the change
  - Pattern in each baseline phase must be different from pattern in each treatment phase
  - Changes are same for each phase-change point in exp.
  - Return to baseline



ABAB Reversal Design cont.

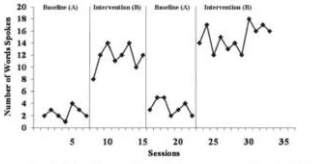


Figure 1. ABAB design: Hypothetical representation of child communication outcomes.

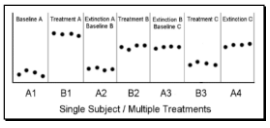
- Controls for influence of extraneous variable
- Can't evaluate treatments expected to have long-lasting effects
  - Carryover effects
- Ethical issues

Criteria for Cause-Effect

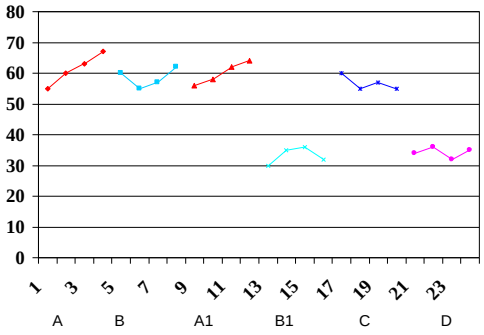
- Clear change in behavior when treatment introduced
- At least one replication of the change
- More difficult to determine with more complex designs

Multiple-Treatment (Multiple I) Designs

- Nature of the intervention changes over time
  - Each change represents a new phase
- Yields a more convincing picture of the effect of the treatment program
- Can change:
  - Intensity of the intervention
  - Number of treatments
  - Nature of the intervention

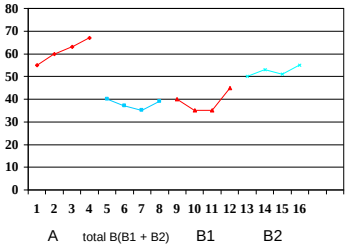


Complex Phase – Change Designs



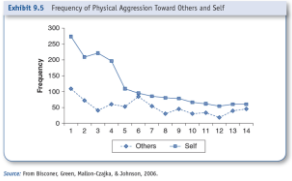
Dismantling or Component Analysis Design

- Breaking treatments down into component parts
- Each phase adds or eliminates one component of the treatment



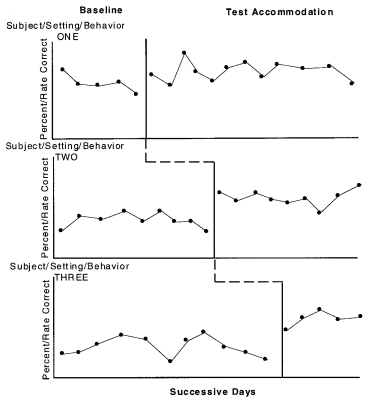
Multiple Baseline Designs

- Two or more behaviors studied simultaneously
  - Obtain baseline on all behaviors
  - Introduce an independent variable that is predicted to affect only one behavior
- Allows the researcher to show that the independent variable is causing the target behavior to change and is not affecting the other behaviors



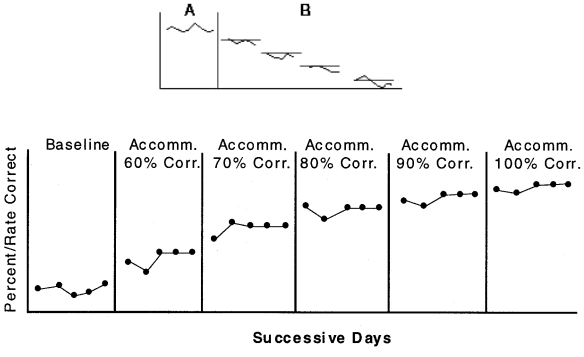
Multiple Baseline Designs

- Eliminates need for return to baseline
- Well suited for evaluating treatments with long-lasting effects
- Controls for history effects
- Only one phase change from baseline to treatment
- Replicates phase change in different situation (multiple baseline across contexts), for second participant (multiple baseline across subjects), or second behavior (multiple baseline across behaviors)
  - Treatment introduced earlier for one participant or one behavior/context



The Changing Criterion Design

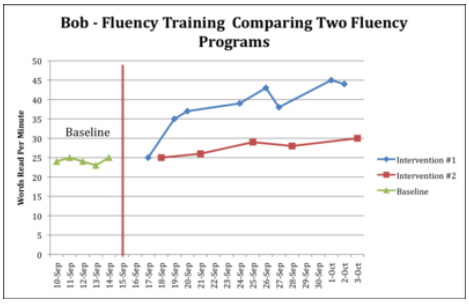
- Treatment involves series of target levels or criteria that can be set by the researcher
- Participant's behavior should change in accordance with changing criterion
  - E.g., smoke 3 packs a day, then 2 packs a day, then 1 pack a day – criterion changes each week
- To differentiate between following trend and stepwise tracking of criterion:
  - Vary length of criterion phases randomly
  - Incorporate backward steps - if criterion is steadily decreasing add one or more phases where it increases



The Alternating Treatments Design

- Also called Discrete Trials or Simultaneous Treatments design
- Allows a test of the relative effectiveness of several treatments in one experiment
- Equal times are created, one for each treatment
  - Each treatment is used during its time period
  - Order of treatment is counterbalanced
- Control condition can be added
  - Helps rule out history and maturation effects
- Participant's behavior must show immediate response to treatment
- Data is grouped by treatment conditions rather than grouped into blocks of time
- Rapidly alternating succession independent of level of responding

The Alternating Treatments Design



Advantages

- Establish cause and effect with only single participant
- Can integrate experimental research into applied clinical practice
- Flexibility
  - No need to standardize treatments

Disadvantages

- External validity
- Internal validity
  - Awareness of continuous observations
  - Reactivity or sensitization
  - Absence of statistical controls
- Small effects not seen in graphs
- Neglect of interactions among variables
- Ethical issues
  - Example: Do you withdraw an effective treatment from a particularly troubled client in a reversal design?

Problems of Interpretation

- Widely discrepant scores in the baseline
- Delayed changes in the intervention phase
- Improvement in the target problem scores during the baseline phase
- Act of graphing can create visual distortions
- Requirements of the statistical test may be difficult or impossible to meet in a small-N design

Generalizability

- Difficult to demonstrate in small-N designs
- Requires replication:
  - Direct replication = same study with different clients
  - Systematic replication = same interventions in different settings
  - Clinical replication = combining different interventions into a clinical package to treat multiple problems

Case Study Research

- **Case study** – a detailed study of a single individual, group, or event
- May use information from numerous sources: observation, interviews, questionnaires, news reports, and archival records
- All information is compiled into a narrative description
- Psychobiography – applying concepts and theories from psychology in an effort to understand famous people
- Illustrative anecdotes

Case Studies

- In depth record of an individual's experience
- No manipulation
- Idiographic approach = intensive study of individuals
- Often used in clinical research
- Demonstrate exception to a rule
- Rare phenomena
  - E.g., woman found alive after being buried under rubble for 60 days in Pakistan earthquake (Naqsha Bibi)
  - H.M.
  - Sybil

Case Studies: Advantages

- Limited focus allows detailed examination of subject
  - More vivid and personal
- Use several different techniques to gather data
- Best way to gather detailed information about subject
- Can suggest directions for future research

Case Studies: Disadvantages

- Time-consuming
- Subject to biases in observing and recording data
  - Selective bias –report most successful or dramatic case
  - All observations may be conducted by a single researcher
  - No way of determining reliability and validity of these observations
- Lack breadth
- Lack both internal and external validity
  - Failure to control extraneous variables
- Cannot demonstrate cause-and-effect relationships
- Limited generalizability
- Exaggerated sense of credibility

Statistical Analysis

- Inferential statistics for single-case experiments are being developed
  - E.g., Bayesian Hypothesis-testing for Single subject designs, permutation (randomization) test, interrupted time-series analysis (ITSA), multi-level modelling
  - Used to compare level, variability, and trend of baseline data to treatment data
    - Examines whether change occurred by chance
    - Is a more sensitive test than visual analysis
  - These techniques are relatively new
  - Evaluation of their effectiveness is ongoing
  - Requires more data points than most single-case researchers collect

POWER

- The probability of getting a significant result when you SHOULD get one
- Correct decision to reject false null hypothesis (accept H_E)
- (1- β = power)

Error

	H_0 True	H_E True
Reject H_0	Type I Error α (no effect)	Correct Decision 1-β (Power) (effect)
AcceptH_0	Correct Decision 1- α (no effect)	Type II Error β (effect)

Power

- When power is < .50 – chance of successful outcome is up to chance
- Cohen – aim for power of .80 (80% chance of success)
  - Type II error rate will be no worse than 20% (one quarter as bad as Type I errors (.05/.20 = .25) or 4:1 ratio; meaning we're more concerned about Type I than Type II

Increasing Power

- Use more powerful statistical tests
- Fewer df in numerator for F tests
- Parametric tests

Increasing Power

- A function of :
  - Sensitivity of study
    - Reliability of measures
    - Control over extraneous variables
    - Accuracy of observations
    - Larger sample sizes
  - Type I error rate
    - Reducing Type I errors reduces power
    - Use less stringent α
  - Effect size
    - Larger difference between null and alternative hypothesis
      - "top and tail" – select participants at extreme ends
      - Increase strength of manipulation
    - Increase association between variables

Effect Size

- Standardized mean difference
- *Cohen's d* = $(M_E - M_C)/SD$
- *d* = 1.0 means the groups differed by a full SD
- Negative *d* can mean treatment was detrimental

Effect Size

- Percentage of variance accounted for = r^2
- $r^2 = \frac{d^2}{d^2 + 1/pq}$
- p and q are proportion of total sample in each group
- Also can switch back to $d = 2r/\sqrt{(1-r^2)}$

Effect Size Conventions

	r^2	d
Small	.01	.20
Medium	.10	.50
Large	.25	.80

Sample Size

- N needed for power of .80 with two-tailed tests assuming $\alpha = .05$
- One-sample Tests
  - $7.85/ d^2$
- Two-sample Tests
  - $7.85/ r^2$

Uses of Power Analyses

- Post hoc power based on observed effect size not very useful
- Should frame power analysis around N needed
- Generally don't know the actual effect size a priori
- Obtain estimates of effect size from prior research or conventions
- Power is always an approximation at best

One-Sample t-test

- Personality of musicians
- Costa & McCrae (1992)
  - Scores falling above or below .5 SD from population mean on each trait considered outside average range
  - .5 SD = *d* of .50
  - $7.85/ .50^2$
  - 32
  - Thus need N of 32 for power of .80

Two-Sample *t*-test

- Compare personalities of singers and instrumentalists
- See pg. 166 in Leong & Austin

Other Applications

- Solve for smallest effect you can reasonably expect to find given a particular sample size
- Solve for α to find significance level you should aim for to obtain desired power level

- https://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx