Design, Measurement, & Evaluation

Enola Proctor & David Chambers

August 2, 2011
Outline

- Research Question to Research Strategy
- Identifying the Research Question(s)
- Considering Multiple Design Options
- Selection of Measures
- Conducting the Evaluation
Objectives

- Gain Familiarity with Design Options
- Review/Discuss Candidate Measures
- Understand Fit between Study Questions and Design
- Learn about Evaluation in implementation research
- Discuss Common Challenges/Complexities
Key Design Questions

- What is my primary question?
- Where am I looking to answer it?
- How could it BEST be answered?
- How could it FEASIBLY be answered?
- What data is currently available?
- What data do I need to gather?
- What do I have control over?
A Range of D and I Questions…

- Applied—e.g. how does intervention X get implemented in setting Y?
- Basic D&I Research—e.g. how do specific stakeholders interpret information?
- Measurement questions—e.g. how do I validly measure implementation outcomes (or processes/strategies or contexts)
- Design questions—e.g. how to account for variation in organizational characteristics?
Res Question Considerations

- **Mechanism**
  - R03, R21, R34, R01, K01…

- **Level of Innovation**
  - For D and I research; for applied field

- **Impact**
  - As Above

- **Approach**
  - Rigorous, Relevant, Robust
What is the impact of a natural experiment to implement an Evidence-Based Intervention within state primary care?

What is my primary question?

- Does the EBI get implemented and What Happens as a Result?

Where am I looking to answer it?

- Multiple Primary Care Clinics

How could it OPTIMALLY be answered?

- Comparison Group

How could it FEASIBLY be answered?

- Phased Roll-out permitting comparators among sites
Illustrative Example #1

- What is the impact of a natural experiment to implement an Evidence-Based Intervention within state primary care?
  - What data is currently available?
    - Claims, Chart abstraction
  - What data do I need to gather?
    - Way in which Intervention was delivered; Client outcomes;
  - What do I have control over?
    - State willing to phase-in roll-out
What was the impact of a natural experiment to implement an Evidence-Based Intervention within state primary care?

- Who’s doing the implementing? SOMEONE ELSE
- What is the timeframe? TWO YEARS
- What is my primary outcome? Secondary outcome(s)? EFFECTIVE DELIVERY of EBI; Patient Outcomes, System Outcomes
- Can I control the design? MAYBE
What is the comparative effectiveness of two strategies to disseminate Evidence-Based Guidelines for diet and exercise to teachers in schools?

What is my primary question?
- Is one strategy superior to the other?

Where am I looking to answer it?
- Schools

How could it OPTIMALLY be answered?
- Matched samples of teachers

How could it FEASIBLY be answered?
- Same
What is the comparative effectiveness of two strategies to disseminate Evidence-Based Guidelines for diet and exercise to teachers in schools?

- What data is currently available?
  - ??

- What data do I need to gather?
  - Teacher behavior, student outcomes

- What do I have control over?
  - Dissemination strategy, timeframe, data collection
What is the comparative effectiveness of two strategies to disseminate Evidence-Based Guidelines for diet and exercise to teachers in schools?

- Who’s doing the implementing? I AM
- What is the timeframe? ONE YEAR
- What is my primary outcome? Secondary outcome(s)? Use/Adherence to Guidelines; Student Outcomes
- Can I control the design? YES
What is the effectiveness of an organizational implementation strategy, ARC, on agency climate & culture, adoption of evidence-based treatments, and service effectiveness?

Who’s doing the implementing? ARC team, from study

What is the timeframe? 3 years

What is my primary outcome?
- Agency climate & culture
- Agency adoption of EBP’s

Secondary outcome(s)? Service effectiveness
- Child outcomes

Can I control the design? YES
Illustrative Example #4

(Insert your favorite)

- What is my primary question?
- Where am I looking to answer it?
- How could it BEST be answered?
- How could it FEASIBLY be answered?
- What data is currently available?
- What data do I need to gather?
- What do I have control over?
Key ingredients in IR

- Study question
- Implementation process
- Outcomes, at multiple levels
- Contextual factors

Each has implications for design & measurement
Types of research questions:

- discovering barriers to the delivery of evidence-based treatment

- mapping pathways to the uptake and sustainability of those treatments in the real world, and shaping implementation strategies

- and testing the comparative effectiveness of implementation strategies.
Types of research to develop and test implementation strategies

**Exploratory**
- Identify Barriers/Facilitators
- Derive implications for strategies

**Developmental**
- Develop implementation strategies that fit service setting & stakeholder preference

**Experimental**
- Test comparative effectiveness of implementation strategies

---

© 2009 Washington University in St. Louis. Any use of these written and graphic materials requires permission from the Brown School at Washington University in St. Louis. This product was developed and supported from the Center for Mental Health Services Research; This advanced center for interventions and services research is funded by NIMH – P30 MH068579.
Actual

- Retrospective
  - Reliance on secondary data and recall
  - Find Comparator
- Observational
  - Primary and secondary data
  - Find Comparator
- Prospective
  - Primary and secondary data
  - Find/Design Comparator
- Simulation: a new way of doing science
Implementation:

Multiple levels of action in implementation

Because implementation is a process, multiple stages, phases, points in time

At what level do you capture a phenomenon?

At what level is measurement taken, and at what level is variation analyzed?

What levels of aggregation are appropriate?

When do you measure and when do you aggregate?

Multiple types of aggregation: common data and mixed methods
Randomization in prospective studies

Yes/ no

At what level does randomization occur?
provider
practice
collaborative/region
state
Questions, Discussion, group think
How to conceptualize and measure success of implementation processes and their impact on care?

Distinct outcomes needed

Outcomes are distinct from clinical outcomes
  - Could have an effective intervention, poorly implemented
  - Could have an ineffective treatment, successfully implemented
Implementation Outcomes

State of field*

- Widely varying constructs used, including clinical outcomes
- Lack of detail regarding constructs
- Unit of analysis errors
- Poor measurement quality

*Grimshaw et al., 2006
Implementation Outcomes

Functions:
1. Indicators of implementation success
2. Proximal indicators of implementation processes
3. Key intermediate outcomes in relation to service system & clinical outcomes (treatment will not be effective if not implemented well)
Conceptual Model: three types of outcomes

What?
QIs
ESTs

How?
Implementation Strategies

Implementation Outcomes
Feasibility
Fidelity
Penetration
Acceptability
Sustainability
Uptake
Costs

Service Outcomes*
Efficiency
Safety
Effectiveness
Equity
Patient-centeredness
Timeliness

Health Outcomes
Satisfaction
Function
Health status/symptoms

Implementation Research Methods

Proctor et al 2009 Admin. & Pol. in Mental Health Services
When treatment efforts are unsuccessful, is failure due to:

- Treatment didn’t work (intervention failure)?
- Treatment was not implemented well (implementation failure)?
- Could have an effective intervention, poorly implemented
- Could have an ineffective treatment, successfully implemented
Types of outcomes evaluated in IR

Implementation Outcomes
- Acceptability
- Adoption
- Appropriateness
- Costs
- Feasibility
- Fidelity
- Penetration
- Sustainability

Service Outcomes*
- Efficiency
- Safety
- Effectiveness
- Equity
- Patient-centeredness
- Timeliness

Client Outcomes
- Satisfaction
- Function
- Symptomatology

*IOM Standards of Care
Implementation outcomes: multiple stakeholders & multiple perspectives

- service consumers
- families
- providers
- administrators
- funders
- legislators
Our scan of implementation outcomes & their measurement

<table>
<thead>
<tr>
<th>Outcomes</th>
<th># Measurement Approaches or Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptability</td>
<td>40</td>
</tr>
<tr>
<td>Adoption</td>
<td>27</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>10</td>
</tr>
<tr>
<td>Feasibility</td>
<td>10</td>
</tr>
<tr>
<td>Fidelity</td>
<td>14</td>
</tr>
<tr>
<td>Penetration</td>
<td>4</td>
</tr>
<tr>
<td>Sustainability</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
</tr>
</tbody>
</table>
Service system outcomes

- Proposed by IOM

- Efficiency
- Effectiveness
- Equity
- Safety
- Timeliness
- Patient-centeredness
Dimensions to consider in conceptualization & measurement

- Referent/ stakeholder/ level of analysis
- Theoretical basis
- Varying terms
- Salient point or stage in implementation
- Types of existing measures
- Note cross walk to some RE-AIM constructs
Studies in SUD measuring implementation outcomes

*Glass et al. 2010
Measurement: Acceptability

Typically brief (10-20 item) Likert scales summarized and dichotomized (Karlsson and Bendtsen (2005)).

Administered via interview & questionnaires, trending toward on-line administration.
Acceptability: Evidence Based Practice Attitude Scale (EBPAS)*

15-items, 5-point Likert scale (Aarons, 2004)

EBPAS administered by professional research staff

One higher order factor/total scale and four lower order factors/subscales:

- Appeal (intuitive appeal of EBPs),
- Requirements (likelihood of adopting EBPs when required),
- Openness (to new practices),
- Divergence (between research-based/academically developed interventions and current practice).

Properties: subscales ranging from .91 to .67; total scale coefficient of .74 (Aarons et al., 2010).
Measurement: Fidelity

Typically multiple item, Likert measures
Summed up scale yields continuous measure of fidelity, often dichotomized
Assessment via:
Self-report (e.g., of components delivered)
Face-to-face or telephone interviews
Observation by research teams
Psychometric validation of many, most scales
Fidelity Measurement: Dartmouth Assertive Community Tx Scale*

Excellent psychometric properties.

Licensed measure in the public domain and is included in SAMHSA’s ACT toolkit.
[http://store.samhsa.gov/shin/content//SMA08-4345/EvaluatingYourProgram-ACT.pdf](http://store.samhsa.gov/shin/content//SMA08-4345/EvaluatingYourProgram-ACT.pdf)

*Teague and colleagues’ (1998) for scale
*McHugo (2007) for its use by SAMHSA in a national EBP project.
Rarely directly measured
Often inferred or judged by researchers
  Program may be deemed feasible if highly rated on other implementation outcomes (acceptability)
Often inferred retrospectively on basis of burden
  Program, screener, or treatment may require too much time
Measurement: Adoption

Dichotomous measure:

- is intervention being used? (Henggeler et al, 2008)

Continuous measures of adoption:

- adding number of program components adopted (Li, Simon, Bodenheimer, Gillies, Casalino, & Shortell, 2004).
- considering adoption intent.

Consistent with transtheoretical model of human behavior change (stages of change).

Stage-based measures have not been psychometrically validated.
“An individual-level measure of participation”

- Measured as a proportion
  - # of persons receiving the EBI / # of persons in the population who would benefit from the EBI

- Individual characteristics important:
  - Is numerator representative of the population who needs it?
Example:

- “Reach” measured prospectively in a workplace obesity prevention program *Step Ahead* Estabrook et al (In press)
- # of employees in evaluation cohort reporting intervention components exposure (e.g. brochures, signage) / # of employees in evaluation cohort
- Evaluation cohort selected to be representative of worksites participating in the intervention
Measurement: Penetration

- Reflects “depth” of implementation in target sites
- Measured as a proportion
  - # sites within a network adopting EBI/
    # network sites exposed to EBI
  - # of providers delivering the EBI / # of
    providers trained
  - # of providers’ cases receiving the
    EBI/ # eligible clients served by
    provider
Adoption:
McGovern et al readiness to adopt

1 – We are not interested and do not think this practice would be effective in our program.
2 – We have considered this practice but see many pros and cons.
3 – We are leaning in the direction of adopting this practice in our program.
4 – We have just begun to implement this practice in our work.
5 – We have been using this practice and efforts are in place to maintain it.

Consistent terminology needed
Clear referent of “what” is being evaluated
   one EBT, the implementation approach, several new Tx’s at once
Specify level of analysis
Test and report measurement properties
Assess salience of outcomes to stakeholder groups
Model interrelationships among outcomes
   Among implementation outcomes
   Between IO’s, service outcomes, client outcomes
Implementation Outcomes: Moving the field forward
Match measurement to study questions & target of implementation strategies

- Providers
- Patients/ health consumers/ families
- Administrators
- Organization or setting
- Communities
Aspects of setting, context, environment are measured to account for additional variance
- Variation in needs among patients/clients/consumers?
- How large is the “implementation gap?” (can determine targeted amount of change)
- Policy, environmental factors
What factors need to be controlled?

What variance needs to be maximized?

What are moderator variables?
Mixed methods

- Qualitative data
- Quantitative data
- Informing one with the other
  - Simultaneous, at risk of silo’ed
  - Sequential
  - Integrative

*The whole picture is greater than the sum of the parts*

- Examples from the group?
Mixed research methods:

- Record abstraction
- Stakeholder preference assessment
- Qualitative research methods
- Multi-level analysis
- Data management unit
Mixed methods of data collection & analysis

- Document collection & review
- Key informant interviews to identify modifiable dimensions (e.g., of implementation strategy)
- Conjoint analysis (CA) to quantify stakeholder preferences
  - Measurement approach from psychology, used in marketing and product development; participants rate dimensions
- Group model building with stakeholders
DIAMOND study

- Natural experiment
- Key Question—does facilitated implementation and financing lead to uptake of collaborative care for Depression
- Phased-in rollout of CC across multiple PC clinics
- Not randomized
- Assess individual depression outcomes, use of CC at provider and practice level, quality of care provided
MTFC Implementation at County Level

- Randomized Controlled Trial
- Key Question—does Community Development Team model lead to better implementation and sustainability of MTFC for children with disruptive behavior disorders
- 54 counties randomized to CDT or Implementation as Usual
- Dynamic Wait-list Controlled Design
- Assess individual kid outcomes, use of MTFC with Stages of Implementation Measure
Implementing EBPs with ARC in child MH

- Randomized trial
  - Randomization at agency level
  - Comparison group will receive intense ARC workshop at study end

- Key Question—does ARC, as an implementation strategy, lead to better organizational climate & culture for EB practices, better uptake of EB practices, and better service effectiveness (agency & child outcomes)

- 15 agencies
  - 7 experimental arm, 8 observation (oversubscribed)

Assess agency climate & culture, adoption of
Points for Discussion (all Group)

- Fitting Question to Design—Examples?
- Levels of Analysis?
- Qualitative and Quantitative
  - Mixed Methods—Separate or Integrated?
- From Theoretical to Empirical
  - Answering a practical question
  - Advancing theory
- Trade-offs
- Salvage Strategies—when something goes wrong…
Match Design, Measures and Evaluation Process to Question

Each design has clear strengths and weaknesses

Valid measures exist, but not for all constructs
  - Opportunity to contribute to measure development

Funded research studies have tremendous variation in design
  - No one-size fits all

Goal is to maximize rigor, relevance and feasibility of research design.
**Key References**


Questions, Discussion, group think
THANKS!
dchamber@mail.nih.gov
ekp@wustl.edu
Encouragement Designs

- All exposed to same intervention
- Encouragements built in
  - reminders, lotteries, or incentives
  - incorporation of preferences; additional supportive training; technical assistance or coaching;
- Encouragements for individuals and families, physicians, other providers, organizations, or agencies.
- These designs test whether participation in or adherence to an assigned intervention can be modified using motivators.
- These designs can also be used to test the causal effect of an intervention with imperfect
Preference Trial

- Prefer A
  - Receive A
    - Assess Outcomes
  - Indifferent Randomised
    - Receive A
    - Assess Outcomes
  - Prefer B
    - Receive B
    - Assess Outcomes

www-users.york.ac.uk/~djt6/N1course/preference.ppt