Design, Measurement, & Evaluation

David Chambers & Enola Proctor
June 4, 2013
• 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.
Outline

- Research Question to Research Strategy
- Selection of Measures
- Conducting the Evaluation (Practical Exercise)
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 are currently available?
- What data do I need to gather?
- What do I have control over?
What is the impact of a natural experiment to implement an Evidence-Based Intervention within state primary care clinics?

- 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 are 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 are 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
• Mechanism
  – R03, R21, R34, R01, K01…
• Level of Innovation
  – For D and I research; for applied field
• Impact
  – As Above
• Approach
  – Rigorous, Relevant, Robust
Basic D&I Research—e.g. how do specific stakeholders interpret information about implementation?

Applied—e.g. how does (EB) intervention X best get implemented in setting Y?

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?
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
Recent Study Examples

• 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
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
Key concepts

- Acceptability
- Adoption
- Appropriateness
- Feasibility
- Fidelity
- Implementation cost
- Penetration
- Sustainability
Why distinct implementation outcomes?

When services are unsuccessful, is failure due to:

– Services didn’t work (service or treatment failure)?
– Services or treatments were not implemented well (implementation failure)?

– Could have an effective treatment, poorly implemented
– Could have an ineffective treatment, successfully implemented
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

*IOM Standards of Care

Implementation Research Methods

Proctor et al 2009 Admin. & Pol. in Mental Health Services
Implementation outcomes

Multiple stakeholders & multiple perspectives

- service consumers
- families
- providers
- administrators
- funders
- legislators
Studies in SUD measuring implementation

- Acceptability
- Adoption
- Appropriateness
- Costs
- Feasibility
- Fidelity
- Penetration
- Sustainability
- Other

Percent of studies

0% 10% 20% 30% 40% 50%
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).
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
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

* Teague and colleagues’ (1998) for scale
* McHugo (2007) for its use by SAMHSA in a national EBP project.
• Reflects “depth” of implementation in target sites; akin to population “reach”
  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
Consistent terminology needed

Clear referent of the “what” being implemented:
  One EBT?
  An implementation approach/ strategy?
  Several new Tx’s at once?

Specify level/ unit of analysis

Assess salience of outcomes to stakeholder groups
Test and report measurement properties
GEM
SIRC (in Supplemental Slides)
Rarely defined & poorly described *
   Referenced by many terms
      “tower of Babel”
   May be “least developed” aspect of implementation science **
   Are complex social interventions***

*Michie, Fixsen, Grimshaw, Eccles, 2009)
** Laura Damschroeder, personal communication
*** Mittman, IRI, June 25, 2012
Strategies definitions:

68 implementation strategies and definitions, grouped by six key processes

– Planning
– Educating
– Financing
– Restructuring
– Managing quality
– Attending to policy context

*Powell, McMillen, Proctor et al., Medical Care Research and Review, 2012
Measurement challenge:

**Discrete**

involve one process or action, such as “meetings,” “reminders”

**Multifaceted**

uses two or more discrete strategies, such as “training + technical assistance”

**Blended**

several discrete strategies are interwoven & packaged as protocolized or branded strategies, such as “ARC,” IHI Framework for Spread”

*Powell, McMillen, Proctor et al., 2012

**Grimshaw et al., 2001, Grol & Grimshaw, 2003

*
Consolidated Framework for Implementation Research (CFIR):*

- Intervention characteristics
- Outer setting
- Inner setting
- Characteristics of individuals involved
- Process of implementation

*Damschroder et al., 2009
Strategies have many potential actors

- Payers
- Administrators
- Providers/ clinicians
- Clients/ patients/ consumers
- Community stakeholders
Measurement Challenge: Operational definitions

Implementation strategies carry same demands as interventions
  • Operational definitions
  • Protocols & manuals
    • Undeveloped with few exceptions (ARC)*
  • Fidelity
    • Few tools and fidelity measurement procedures

*Glisson & Schoenwald, *Mental Health Services Research*, 2005
• What contextual factors might affect implementation?

• Guidance from conceptual models:
  – Intervention characteristics
  – Inner setting
  – Outer setting
  – Individual
• Networks and communication
• Organizational Structure
• Organizational culture and climate
• Implementation climate
  – Tension for change, relative priority, learning climate
• Readiness for implementation
Examples: Outer Setting

- Patient, consumer needs
- Peer pressure
- Financing Structures
- External incentives, policies
Take Home Messages

• 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


THANKS!

dchamber@mail.nih.gov
ekp@wustl.edu
Supplemental Slides
Dissemination and Implementation Measures and Methods Initiative

https://www.gem-beta.org/ (GEM Homepage)
http://cancercontrol.cancer.gov/IS/resources.html (IS Team Website)
# What, Why, and Who

## What is the D&I Measures and Methods Initiative?

**Purpose:** Bring together an international community of researchers and practitioners to create a growing and evolving resource for standardized, vetted D&I measures that can lead to comparable datasets and facilitate collaboration and comparison across disciplines and regions.

The D&I Measures and Methods Initiative and resource enables researchers and practitioners to:
- Identify and define constructs relevant to D&I research and practice;
- Learn about, comment on, and rate existing measures for D&I;
- Share new D&I measures;
- Identify missing D&I measures;
- Learn about strategies/methods relevant to D&I

## Why should I get involved?

If you are interested in advancing the D&I field, this Initiative is an excellent way to contribute to the field and engage with colleagues. The D&I Measures and Methods Initiative gives you access to D&I constructs, measures, and methods developed by other colleagues and also provides you with a platform to share your own D&I measures.

## Who Should Participate?

Researchers and practitioners involved or interested in D&I research
A Comprehensive Review of Dissemination and Implementation Science Instruments

Cara C. Lewis, Cameo Borntrager, Ruben Martinez, Phil Fizur, & Kate Comtois
Three Primary Goals

- Step 1: Instrument Repository
- Step 2: Comprehensive Review
- Step 3: Evidence Based Assessment
- Step 4: Consensus Battery
Comprehensive Review

-Organized by
  -Consolidated Framework for Implementation Research (39 constructs)
    -Damschroder, Aron, Keith, Kirsh, Alexander, & Lowery, 2009
  -Outcomes for Implementation Research (17 constructs)
    -Proctor, Silmere, Raghavan, Hobmand, Aarons, Bunger, et al., 2011
-Construct reviews resulted in over 400 measures
Survey of Organizational Functioning

The TCU Survey of Organizational Functioning is a tool used to assess organizational functioning via scales that explore job attitudes, workplace practices, motivational factors, program resources, staff attributes, and organizational climate.

TCU Survey of Organizational Functioning (SOF)
TCU Survey of Organizational Functioning (SOF) Scoring Guide

Articles related to this measure


Evidence-Based Assessment Criteria

EBA

0 = None  1 = Minimal  2 = Adequate  3 = Good  4 = Excellent

Psychometric Properties

- Norms
- Reliability Information: Inter-rater, Test-retest, Internal
- Validity Information: Criterion, Content, Construct
- Validity Information: Predictive, Discriminant, Convergent, Concurrent
- Usability
Volunteer implementation experts

- Core workgroup members representing Universities of Washington, Montana, Indiana, North Carolina

- Nearly 50 members

- IRTF members have agreed to rate the measures using the EBA criteria
- Rating has begun and will be completed by the end of this calendar year
- From the ratings, a consensus battery will emerge
Measurement challenges:

data quality may vary

• Observational studies:
  – may offer rich detail but limit data availability
    • Retrospective studies
      – Capture agency records, access treatment purveyor records
    • Prospective
      – Interview key participants
      – Direct observation

• Experimental studies
  – Opportunity to vary strategy content, target, actors, timing, & dose
<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Challenges</th>
</tr>
</thead>
</table>
| Survey approaches ("what strategy did you use?" “Did you use XXX?”) | Limited reliability of self-report  
Respondent limited perspective |
| Medical, clinic records, archival data | Inconsistent terminology for strategies  
Lack of detail in recording process  
Lack of “procedure codes” for strategies |
| Ethnographic observation            | Access,  
Potential incomplete picture                                                                |
| Purveyor, EST disseminator data     | Access                                                                                         |
Assessing strategy effectiveness

Test strategy effectiveness in relation to implementation outcomes*

- Acceptability to stakeholders
- Adoption:
  - Was the strategy actually used?
- Timeliness
  - Stages of Implementation Completed (SIC)*
- Penetration of EST within organization
- Sustainability of EST over time

* Proctor et al., *Admin Pol MH Serv*, 2010, GEM, NCI

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
• 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