

Breakout 2.3: Measurement and Evaluation Issues in Hybrid Designs

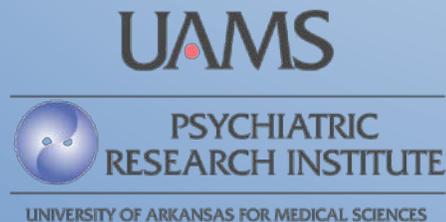
Geoffrey M. Curran, PhD

Director, Center for Implementation Research

Professor, Departments of Pharmacy Practice and Psychiatry

University of Arkansas for Medical Sciences

Research Health Scientist, Central Arkansas Veterans Healthcare System



Structure of the Session

- Go through hybrid types focusing on measurement issues
- Provide examples
- Give a sense of some of the “new thinking” we have on hybrid designs as we have been reviewing 80+ papers reporting hybrid designs
- Open the discussion
 - Structured activity?
 - Problem-solve cases from the audience?
 - Open Q&A?

Evaluation Issues for Type 1

- The original definition of a type 1 emphasized secondary aims/questions and exploratory data collection and analysis *preparatory* to future implementation activity
- Type 1 examples range from conventional effectiveness study with limited exploratory B/F measurement, to intensive parallel process evaluation explaining effectiveness findings and elucidating implementation factors
- Some get into development of implementation strategies in response to process eval, diagnostic data

Example of the former... (previously described)

- **Curran et al., 2012, *Implementation Science***
- **Qualitative process evaluation alongside trial**
 1. How did CALM operate in your clinic?
 2. What worked and what didn't work?
 3. How did CALM affect workload, burden, and space?
 4. How was CALM received by you and others in your site and how did that change over time?
 5. Were there "champions" or "opinion leaders" for CALM?
 6. Did the communication between the ACS, the external psychiatrist, and local PCPs work?
 7. What outcomes are/were you seeing?
 8. What changes should be made to CALM?
 9. What are the prospects for CALM being sustained and why?

Example of the latter...

- **Zoellner et al., 2014, *Contemp Clin Trials***
 - Patient-level RCT of intervention to reduce consumption of sugar-sweetened beverages
 - RE-AIM framework guided evaluation
 - Process evaluation: Reach, implementation
 - Impact evaluation: Effectiveness, maintenance
 - Interviews assessed perceptions of intervention components (small group sessions, personal action plans, drink diaries/exercise logs, teach back call, IVR calls, resources provided)
 - Adoption not measured: research staff delivered intervention

Measurement Issues in Type 1

- **Effectiveness stuff “as usual” ...**
 - Consider PRECIS 2 tool to assist with trial specification along the continuum from “very explanatory” to “very pragmatic”
- **Process evaluation best guided by implementation framework**
 - I know we sound like broken records with this one, but...
 - You’ll use the framework from here all the way through development and evaluation of the implementation strategies that will result
- **Process evaluation usually going to be mixed method**
 - Do you have that expertise on your team?
 - Use framework to pattern interview/observation guides
- **The process evaluation is usually not a “budget-buster”**
 - Fits well into a traditional R01 budget along with a trial

Evaluation Issues for Type 2

- The original definition of a type 2 described possibilities of dual focused, dual randomized, factorial designs & randomized effectiveness trials nested in *pilots* of implementation strategy
 - Majority of current published studies are the latter
- Either way, there are 2 interventions/strategies being evaluated and 2 sets of outcome data– “intervention” outcomes and “Implementation strategy” outcomes
- CLARITY when it comes to BOTH becomes paramount

Evaluation Issues for Type 2 (cont.)

- RE-AIM is a very common evaluation framework for hybrid type 2 (and 3) studies
 - Reach: Who got intervention and who didn't?
 - Effectiveness: Did people who got it get better?
 - Adoption: To what extent was the intervention adopted?
 - Implementation: How "good" was the intervention delivered? (fidelity)
 - Maintenance: How long did these changes last?
- Have a good argument about which measure "goes" where...
- Use a table to quickly depict your measures by RE-AIM

Example of Type 2 with *pilot* Impl Strat (noted earlier...)

- **Cully et al., 2012, *Implementation Science***
 - Clinical trial of brief cognitive behavioral therapy in treating depression and anxiety
 - Patient randomization only; Pilot study of implementation strategy (online training, audit and feedback, facilitation)
 - Intent-to-treat analysis of clinical outcomes
 - Feasibility, acceptability, and “preliminary effectiveness” data collected on implementation strategy
 - Measured knowledge acquisition, fidelity to model
 - Qualitative data on implementability, time spent, etc.
 - Measured sustainability of provision of brief CBT after trial
 - Preparatory to implementation trial of strategy

Specification issues for Type 2

- Important to have an explicit implementation strategy as well as clinical/prevention intervention
 - For a Type 2, if the effectiveness portion of the study is successful, we expect the next step to be testing the implementation strategy in a comparative way
 - So the impl strategy needs to be clearly specified and measured
- Consider using terms consistent with literature
 - Implementation outcomes (Proctor et al., 2011)
 - Implementation strategies (Powell et al., 2015)

Reporting Considerations important for Type 2

- Think of a parallel process to be followed in the description of:
 - Aims (effectiveness and implementation)
 - Description of intervention and strategy components
 - Evidence and theoretical background of interventions and strategy components
 - Outcomes measures
 - Data analysis process

Measurement Issues in Type 2

- **Effectiveness measurement often the same as “regular” effectiveness trial**
 - Perhaps smaller sample than if whole study focused on this, however
- **As type 2 is usually an initial test of the implementation strategy, mixed method approach common with implementation outcomes measurement**
 - Try to have the adoption/fidelity measures be as “quantitative” as possible
 - Feasibility, satisfaction, recommendations for changes, etc., usually qualitative
- **Budget and time considerations usually impact sample sizes and measurement choices across both aspects of the study**
- **Type 2 studies often involve formative evaluation, which adds complications in terms of needing outcomes and analyses quickly at many intervals during the study upon which to base adaptations/changes**

Guidance on *Formative Evaluation*

- A number of studies, especially type 2 designs, include formative evaluation to improve/adapt the implementation strategies (Stetler et al., 2006)
- When considering formative evaluation, it is important to have multiple time points of evaluation to capture the effect of changes
- It's helpful to have more formalized iterations (e.g., making changes at specific time points)
- Audit and feedback data often part of the FE
 - Everyone looking at the implementation outcomes together...

Evaluation Issues for Type 3

- These are mostly implementation trials *plus* an evaluation of clinical/prevention outcomes
 - Many compare a “standard” or “low intensity” implementation strategy to an “enhanced” or “higher intensity” strategy
 - Standard strategy often consists of training, limited technical support, limited consultation
 - Enhanced strategy often adds extended facilitation, PDSA cycles with adaptation, leadership engagement, mentoring, etc.
 - Some randomize at clinic/hospital level, some randomize at provider level, some don’t randomize but site-match and select, some randomized the timing of the start of the strategies

Example 1 of Type 3

- **Lewis et al., 2015, *Impl Science***
 - Dynamic cluster randomized trial of “standard” vs. “tailored” strategies supporting implementation of measurement-based care for depression
 - 12 sites, 150 providers, 500 patients (targets)
 - *Tailored implementation* refers to responsive application of implementation strategies and content matched to determinants of practice (i.e., barriers) identified via a needs assessment and formative evaluation
 - Primary outcome variables are around implementation– measurement-based care fidelity, plus extensive data on contextual mediators; qualitative data used to drive tailored strategy and provide elaboration/interpretation of main findings
 - Secondary outcomes measures are around effectiveness– e.g., depressive symptoms

Example 2 of Type 3

- Kilbourne et al., 2014, *Impl Science*
 - Large implementation trial comparing a “standard” implementation strategy (Replicating Effective Programs) with “enhanced” strategy (REP plus *external facilitation*) to promote the uptake of Re-Engage, an outreach program to Veterans with SMI who are lost to care (150+ VA sites)
 - Context was a system-level mandated “roll-out”
 - Adaptive trial– sites “failing” under initial attempt at the standard strategy randomized to enhanced
 - Main outcome variables are around implementation– extent of employing Re-Engage behaviors (locating and contacting behaviors), extensive contextual covariates; qualitative data on barriers/facilitators to implementation
 - Secondary outcomes measures from medical records are around percentages of Veterans who re-engaged in services, service utilization (*proxies* of clinical outcomes)

Measurement Challenges in Type 3

- **Implementation outcomes/instruments**
 - “Implementation outcomes instrumentation is underdeveloped with respect to both the sheer number of available instruments and the psychometric quality of existing instruments.” (Lewis et al., 2015)
 - Domains: acceptability, adoption, appropriateness (fit), feasibility, fidelity, cost, penetration, sustainability (Proctor et al, 2011)
 - Still, many measures are new and often study-specific
 - RE-AIM is another useful tool to drive selection of outcome domains
 - Reach, Effectiveness, Adoption, Implementation, Maintenance
 - Studies using RE-AIM and other models appear to have more clearly-defined outcomes measures
- **Effectiveness**
 - Readily-available (i.e., administrative) effectiveness measures are scarce for many conditions (e.g., mental health)
 - Primary data collection for effectiveness measures can severely limit overall study power

Power Challenges in Type 3

- Trade-offs between # of sites, providers, patients
- Many applying mixed-methods and case-comparative qualitative methods to improve understanding of implementation outcomes when number of sites is small
 - Does team have appropriate expertise?
- Overall power (# of implementation sites) can be increased if secondary data sources are available for effectiveness outcomes
- Some using multi-level modeling, which can help maximize power

Type 3 clarifications, revisions, extensions

- At the 2014 D&I Meeting session on hybrids a person asked the following question– “When wouldn’t we want to collect clinical outcomes during an implementation trial?”
 - We based these types originally on the assumption that there comes a time when we don’t “need to” anymore. Is that correct?
 - “Probably not” – even as interventions move down the implementation pipeline there is a need to monitor outcomes to assure stakeholders that results are sustained for their patients – the question is how much time, effort, dollars should be invested to monitor clinical outcomes?
 - We expect clinical outcomes to vary by level/fidelity of implementation and continuing adaptation of the clinical intervention itself... how much do we want/need to know?
 - Dynamic Sustainability Framework from Chambers et al., 2013 (interventions and implementation cannot be “frozen”)

Lets Open This Up!

- **Structured activity?**
 - How to turn your effectiveness study into a hybrid 1 or 2?
 - How to turn your implementation trial into a hybrid type 3?
- **Problem-solve cases from the audience together?**
 - Or break-out groups on several cases?
- **Open Q&A?**