Controversies and Unresolved Issues in the Design of Randomized Controlled Trials Testing Clinical/Behavioral Public Health Interventions

Part I: Control Group Design

UCSF CAPS Methods Core Seminar

October 23, 2018

Steve Gregorich
Broad Overview

Topics on designing/conducting behavioral/clinical RCTs in public health

. Part I:  Control group design
. Part II:  Adjustment for multiple testing?
. Part III: Goals and design of Pilot RCTs

Guiding principles

. Inform policy: Improve health, well-being, QoL, life expectancy
. Evidence-based medicine requires medicine-based evidence †
. Ethical considerations

Overview of Part I: Control group design

Focus on Efficacy, Effectiveness, and Implementation RCTs
  . Not on Comparative Effectiveness RCTs
  . RCTs and threats to internal validity
  . Usage of health behavior theories in research practice
  . Conceptual decomposition of generic effect types
  . Testing a theory vs. testing a theory-informed intervention
  . Intervention Testing: Efficacy-Effectiveness
  . NIH Stage model
  . Control groups in RCTs of behavioral/clinical interventions
  . Impact of control group design on anticipated effects
  . Ethical considerations
  . Proposal writing and manuscript strategies
Gold Standard: The Randomized Controlled Trial

$$\begin{align*}
\text{Rnd} & \quad \begin{cases} 
\text{Intv: } O_{t1} & \text{Tx} & O_{t2} \\
\text{Ctrl: } O_{t1} & O_{t2}
\end{cases}
\end{align*}$$

- Rnd: Equivalent groups at $t_1$.

- If 'closed-system' maintained, then sound basis for causal inference about Tx effects.

Offers protection from threats to internal validity listed below:

<table>
<thead>
<tr>
<th>Selection</th>
<th>History</th>
<th>Maturation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing</td>
<td>Instrumentation</td>
<td>Regression</td>
</tr>
</tbody>
</table>

Ambiguous temporal sequencing of measurement
Usage of health behavior theories in research practice

Four broad usage categories †

. **Mention**: A theoretical framework was mentioned, but research components & measures don't seem to derive from the theory

. **Application**: Theoretical framework mentioned and seems to have informed research components and measures

. **Testing**: Theoretical framework mentioned and theoretical constructs were tested, or two or more theories were compared

. **Theory building**: Research intended to develop a new or revised theory

My focus is on the **Application** and **Testing** categories

Conceptual decomposition of generic effect types †

<table>
<thead>
<tr>
<th>Specific effects:</th>
<th>Change in a specific outcome that is attributable to mechanisms postulated within the targeted theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specific effects:</td>
<td>Change in a specific outcome that is attributable to mechanisms not postulated within the targeted theory. E.g., contextual factors during intervention delivery; placebo effect</td>
</tr>
<tr>
<td>Common effects:</td>
<td>Non-specific effects that are shared across alternative interventions. I.e., a subset of all Non-specific effects</td>
</tr>
<tr>
<td>Total effects:</td>
<td>The combination of Specific + Non-specific effects</td>
</tr>
</tbody>
</table>

Estimating Specific & Non-specific effects in most designs requires untestable assumptions

A 'unified theory' view regards Non-specific effects as theory shortcomings

Common effects are typically & inaccurately labeled 'Common factors'. 'Common effects' is a more accurate and preferred label

Testing a theory vs. testing a theory-informed intervention

Testing a Theory: RCT **must** be designed to estimate Specific effects
Scientific evidence for or against a theory rests upon Specific effects
E.g., Comparing experimental drug versus placebo
Theory Testing: Largely, the realm of basic science

Testing an Intervention: RCT **should** be designed to estimate Total effects
Efficacy, Effectiveness & Implementation RCTs focus on Total effects
E.g., Giving an Rx w/ confidence improves patient outcome †
Theory application: Largely, the realm of applied science

Public health investigators should more deliberately consider the Theory Testing vs Intervention Testing distinction

Intervention Testing focuses on Efficacy & Effectiveness

Does the intervention do more good than harm… †
  …when delivered under **optimal** circumstances? **Efficacy**
  …when delivered under **real-world** circumstances? **Effectiveness**

"…more good than harm…" Compared to what?
To inform policy, the comparator needs to be clinically relevant
I.e., Evidence-based medicine, requires medicine-based evidence ‡

**Efficacy** trials maximize internal validity; external validity suffers
**Effectiveness** trials maximize external validity; internal validity suffers

<table>
<thead>
<tr>
<th>Intervention setting</th>
<th>Intervention delivery</th>
<th>Intervention maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient selection</td>
<td>Provider skill level</td>
<td>Study staffing</td>
</tr>
<tr>
<td>Costs</td>
<td>Outcomes</td>
<td>Study duration</td>
</tr>
</tbody>
</table>

Intervention Testing: Efficacy-Effectiveness Continuum

Efficacy and Effectiveness exist along a continuum

When designing an RCT of a behavioral/clinical intervention, 'push' the design as far toward effectiveness as reasonably possible

We have limited human capital, time, and money

Maximizing benefits and reducing risks to participants. Minimize participant burden: Carefully choose research to conduct

Investigators should consider this very carefully
NIH Stage Model for Behavioral/Clinical Interventions †

Stage 0
  . Basic Science

Stage 1A
  . De novo creation of a new intervention, or
  . Modification, adaptation, or refinement of an existing intervention

Stage 1B
  . Acceptability, feasibility, & pilot testing of the new intervention

Stage 2 (may be skipped to move directly to community-based research)
  . Efficacy trial in research setting; research interventionists/providers

Stage 3
  . Efficacy trial in community setting; community interventionists/providers

Stage 4
  . Effectiveness trial in community setting. Maximizing external validity

Stage 5
  . Dissemination and implementation research

Control groups in RCTs of behavioral/clinical interventions

Control group designs that I focus on today

**Attention Placebo Control** (APC; sham therapy)
- APC must appear to credibly impact the targeted health outcome
- APC and INT are matched on all non-specific factors
- APC is *inert*: includes no specific factors
- Holy grail of psychotherapy research. Attainable?

**Time & Attention Control** (TAC; *AKA* Attention Control)
- TAC & INT have differing content, e.g., Diet/Exercise v Sex Risk Beh
- TAC & INT matched on frequency, manner & duration of contact;
  matched on participant attention required

**Usual Care** (UC; *AKA* Unrestricted Standard of Care)
- Patients receive UC 'in the wild'
- Natural variation in UC content is documented
- May result in an Additive Design: i.e., UC versus INT+UC
- Provides base-rate for safety monitoring
Control groups in RCTs of behavioral/clinical interventions

Other control group designs that I will not discuss

Standard of Care (SoC; AKA Restricted/Protocolized Standard of Care)
  . Identify 'the' standard of care, manualize it as part of study protocol
  . Fidelity of SoC is monitored
  . Care must be taken to avoid 'practice misalignment'
  . Higher internal validity & lower external validity than UC

Waitlist Control (WLC)
  . Participants assigned INT immediately or after a prescribed interval
  . Those assigned to WLC effectively are UC during the waitlist period

Dismantling Control (DIS)
  . Multiple experimental groups with differing combinations of intervention components.
Review: Conceptual decomposition of effect types

**Specific effects:** Change in a specific outcome that is attributable to mechanisms postulated within the targeted theory

**Non-specific effects:** Change in a specific outcome that is attributable to mechanisms not postulated within the targeted theory. E.g., contextual factors during intervention delivery; placebo effect

**Common effects:** Non-specific effects that are shared across alternative interventions. I.e., a subset of all Non-specific effects

**Total effects:** The combination of Specific + Non-specific effects

*Next*

Expected effects w/in RCTs of a sex risk reduction intervention versus…

- Attention Placebo Control (APC): Sham risk reduction intervention
- Time & Attention Control (TAC): Diet & Exercise TAC
- Usual Care (UC): Whatever participants access
### Attention Placebo Control: Decomposition of effect types

**Ideal RCT Outcome**  (Assuming APC is credible; untestable assumptions)

<table>
<thead>
<tr>
<th>Rnd Group</th>
<th>Specific Effects</th>
<th>Non-Specific Effects</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT (↓ sex risk)</td>
<td>$S_{INT}$</td>
<td>$C$</td>
<td>$S_{INT} + C$</td>
</tr>
<tr>
<td>APC (sham)</td>
<td>.</td>
<td>$C$</td>
<td>$C$</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>$S_{INT}$</td>
<td>.</td>
<td>$S_{INT}$</td>
</tr>
</tbody>
</table>

$S_{INT} = \text{Specific effects of experimental intervention (INT) on sex risk}$

$C = \text{Common non-specific effects of INT and APC on sex risk}$

**IFF**  APC is truly inert and participants view it as a plausible intervention

. APC will have no specific effects on sex risk: $S_{APC} = 0$

. Non-specific effects of INT and APC might be common

### Many argue that APC is unattainable in behavioral/clinical research

Time & Attention Control: Decomposition of effect types

**Ideal RCT Outcome** (*implausible & untestable assumptions*)

<table>
<thead>
<tr>
<th>Rnd Group</th>
<th>Specific Effects</th>
<th>Non-Specific Effects</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT (↓sex risk)</td>
<td>$S_{\text{INT}}$</td>
<td>$C^*$</td>
<td>$S_{\text{INT}} + C^*$</td>
</tr>
<tr>
<td>TAC (lifestyle)</td>
<td>.*</td>
<td>$C^*$</td>
<td>$C^*$</td>
</tr>
<tr>
<td>Δ</td>
<td>$S_{\text{INT}}$</td>
<td>.*</td>
<td>$S_{\text{INT}}^*$</td>
</tr>
</tbody>
</table>

**Plausible RCT Outcome** (*implausible & untestable assumptions*)

<table>
<thead>
<tr>
<th>Rnd Group</th>
<th>Specific Effects (S)</th>
<th>Non-Specific Effects (N)</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT (↓sex risk)</td>
<td>$S_{\text{INT}}$</td>
<td>$N_{\text{INT}}$</td>
<td>$S_{\text{INT}} + N_{\text{INT}}$</td>
</tr>
<tr>
<td>TAC (lifestyle)</td>
<td>.</td>
<td>$N_{\text{TAC}}$</td>
<td>$N_{\text{TAC}}$</td>
</tr>
<tr>
<td>Δ</td>
<td>$S_{\text{INT}}$</td>
<td>$N_{\text{INT}} - N_{\text{TAC}}$</td>
<td>$S_{\text{INT}} + N_{\text{INT}} - N_{\text{TAC}}$</td>
</tr>
</tbody>
</table>

- No expectation T&A yield **only** common effects on sex risk across INT & TAC.
- Within INT, T&A expected to have **positive** Non-specific effects on sex risk.
- Within TAC, T&A may have **no or negative** Non-specific effects on sex risk b/c focus on lifestyle factors may distract from a focus on sex risk †.
- RCT Result: Total effect estimate for an uninformative research question

Time & Attention Control: Decomposition of effect types

Even if you assume that $N_{TAC}=0$…

**RCT Outcome when $N_{TAC}=0$ (* implausible and untestable assumptions)**

<table>
<thead>
<tr>
<th>Rnd Group</th>
<th>Specific Effects (S)</th>
<th>Non-Specific Effects (N)</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INT</strong> (↓sex risk)</td>
<td>$S_{INT}$</td>
<td>$N_{INT}$</td>
<td>$S_{INT} + N_{INT}$</td>
</tr>
<tr>
<td><strong>TAC</strong> (lifestyle)</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>$S_{INT}$</td>
<td>$N_{INT}$ *</td>
<td>$S_{INT} + N_{INT}$ *</td>
</tr>
</tbody>
</table>

…what you end-up with is an estimate of the Total effect of the experimental intervention

$S_{INT}+N_{INT}$ is a useful quantity to know, but…

. That estimate may be unlikely to obtain in a TAC-controlled RCT

. The resulting group difference would include INT T&A effects

  I.e., The exact circumstance that the investigator tried to avoid

. Whether you obtain this estimate rests upon untestable assumptions
Usual Care Control: Decomposition of effect types

**RCT Outcome**

<table>
<thead>
<tr>
<th>Rnd Group</th>
<th>Specific Effects (S)</th>
<th>Non-Specific Effects (N)</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INT</strong> (↓sex risk)</td>
<td>(S_{\text{INT}})</td>
<td>(N_{\text{INT}})</td>
<td>(S_{\text{INT}} + N_{\text{INT}})</td>
</tr>
<tr>
<td><strong>UC</strong> (as observed)</td>
<td>(S_{\text{UC}})</td>
<td>(N_{\text{UC}})</td>
<td>(S_{\text{UC}} + N_{\text{UC}})</td>
</tr>
<tr>
<td>Δ</td>
<td>(S_{\text{INT}} - S_{\text{UC}})</td>
<td>(N_{\text{INT}} - N_{\text{UC}})</td>
<td>(S_{\text{INT}} + N_{\text{INT}} - S_{\text{UC}} - N_{\text{UC}})</td>
</tr>
</tbody>
</table>

Result: Total effect of INT vs. UC

- Because **UC** is relevant, this Total effect is relevant to policy decisions
- No required assumptions about Specific, Non-specific, Common effects
Sometimes TAC clearly is contraindicated

Example
. An implementation science intervention to prompt physician-patient discussion on a particular topic (e.g., breast cancer screening).

INT: Pt completes Qx about breast cancer risk in waiting room.
. Printed risk summary generated and available during MD visit
. Goal: Promote corresponding MD-Pt discussion during clinic visit

Impact of control group design choices
UC: Doctor and patient complete clinic visit as usual
. Likely result: Estimation of INT effect versus UC

TAC: Pt completes Qx about exercise habits in waiting room
. Printed summary generated and available during MD visit
. Likely result of TAC:
. 'Stacking the deck' against discussing breast cancer risk
. Overstating effect of INT relative to usual care
Ethical Considerations

Minimize human subjects activity required to inform a policy decision

Consider whether your study is Theory Testing or Intervention Testing. Which more likely will maximize benefits & minimize harms to society?

*If Theory Testing*
- Can you design a control group that plausibly allows estimation of Specific effects?

*If Intervention Testing*
- Consider where on the Efficacy-Effectiveness continuum the study optimally should lie
  - Move as far toward Effectiveness as you can
  - Choose a control group design that will best inform policy decisions

**TAC**: Is there equipoise?
Ethical considerations: Usual Care control conditions

UC provides a basis for safety monitoring/excess risk assessment

A UC control group “will enhance clinical value and increase the ability [of the trial] to stop early if needed to protect subjects” † (p. 852)

Although UC may not be up to current guidelines/bests practices (SoC), that does not make UC an unethical control group choice. Numerous evidence-based interventions are not implemented.

“[If] trials lack a control group representative of standard practices, they will not be able to redefine the standard of care.” † (pp. 852-853)

UC allows estimation of clinically relevant Total effects

"[R]esearchers need to think carefully about why it is important to know the extent of an intervention’s nonplacebo effects…If researchers are interested applying their results to clinical practice, a usual care…control group may be more appropriate." ‡ (p.160)


Proposal Strategies

Many reviewers will insist that RCTs incorporate a TAC

Such reviewers don't get much resistance from other reviewers. Therefore, the proposal needs to make a strong case for UC

If you propose a two-group RCT design with UC, then it will help if…
  . You can plausibly claim yours is an Effectiveness/Pragmatic trial
  . Argue that your primary goal is to inform health care policy. Understanding the Total effect of INT vs UC is needed. Not an understanding of the Specific Effects of INT
  . You plausibly argue that TAC would 'stack the deck' and over-estimate the INT effect
  . You plausibly argue that UC is needed for safety monitoring

Not guaranteed. This is very 'churchy' territory
Unfortunately, proposing a TAC design is popular with reviewers!
Manuscript strategies

Occasionally, journal reviewers complain that TAC should have been chosen instead of UC.

It is easy enough to write a response negating that critique.

You should be able to convince all concerned, but only have to convince the editor!
## Summary: Theory Testing versus Intervention Testing

<table>
<thead>
<tr>
<th>Theory Testing</th>
<th>Intervention Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Realm:</strong> Basic science</td>
<td><strong>Realm:</strong> Applied science</td>
</tr>
<tr>
<td><strong>Goal:</strong> Inform theory</td>
<td><strong>Goal:</strong> Inform policy</td>
</tr>
<tr>
<td><strong>Question:</strong> Is the theory supported?</td>
<td><strong>Question:</strong> What works best?</td>
</tr>
<tr>
<td><strong>Estimates of interest:</strong> Specific effects</td>
<td><strong>Estimates of Interest:</strong> Total effects</td>
</tr>
<tr>
<td><strong>Comparator:</strong> Control group identical to Intervention group except for theoretically postulated mechanisms</td>
<td><strong>Comparator:</strong> Control group represents current practice. Often UC for efficacy-effectiveness &amp; implementation trials</td>
</tr>
<tr>
<td><strong>Mechanisms of action:</strong> Tested via Specific effects and/or mediation analysis</td>
<td><strong>Mechanisms of action:</strong> Tested via mediation analysis</td>
</tr>
</tbody>
</table>
Evidence-based medicine requires medicine-based evidence