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

† Knottnerus JA, Dinant GJ (1997). Medicine based evidence, a prerequisite for evidence based medicine. British Medical Journal (315) 309–10.

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

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

- 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

Selection	History	Maturation
Testing	Instrumentation	Regression
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

[†] Painter JE, Borba CP, Hynes M, Mays D, Glanz K. (2008). The use of theory in health behavior research from 2000 to 2005: A systematic review. *Annals of Behavioral Medicine*, 35, 358–62.

Conceptual decomposition of generic 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

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

† Bootzin & Bailey (2005). Understanding placebo, nocebo, and iatrogenic treatment effects. Journal of Clinical Psychology, 61, 861-870.

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

† Thomas KB (1987). General practice consultations: Is there any point in being positive? *British Medical Journal* (294) 1200-1202.

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

Intervention setting	Intervention delivery	Intervention maintenance
Patient selection	Provider skill level	Study staffing
Costs	Outcomes	Study duration

[†] Flay BR (1986) Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.

[‡] Knottnerus JA, Dinant GJ (1997). Medicine based evidence, a prerequisite for evidence based medicine. British Medical Journal (315) 309–10.

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

† Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M (2014). Reenvisioning Clinical Science: Unifying the Discipline to Improve the Public Health. *Clinical Psychological Science*, 2, 22-34.

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)

Rnd Group	Specific Effects	Non-Specific Effects	Total Effects
INT (↓sex risk)	S_{INT}	C	S_{INT} + C
APC (sham)	.	C	C
Δ	S_{INT}	.	S_{INT}

S_{INT} = Specific effects of experimental intervention (INT) on sex risk
C = 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

Borkovec TD, Onken LS. Recommendations for research concerning the use of placebos in clinical trials to test behavioral interventions. In: Guess HA, Kleinman A, Kusek JW, Engel LW, eds. The Science of Placebo: Toward an Interdisciplinary Research Agenda. London: BMJ Books; 2002:306-310.

Penzien DB, Andrasik F, Freidenberg BM, Houle TT, Lake AE, Lipchik GL, Holroyd, KA, Lipton RB, McCrory DC, Nash JM, Nicholson RA, Powers SW, Rains JC, Wittrock DA (2005). Guidelines for Trials of Behavioral Treatments for Recurrent Headache, First Edition: American Headache Society Behavioral Clinical Trials Workgroup. Headache, 45[Suppl 2], S110-S132.

Wampold BE, Frost ND & Yulish NE (2016). Placebo Effects in Psychotherapy: A Flawed Concept and a Contorted History. *Psychology of Consciousness: Theory, Research and Practice*, 3, 108-120

Kirsch I (2005). Placebo psychotherapy: synonym or oxymoron? J Clin Psychol, 61, 791–803.

Time & Attention Control: Decomposition of effect types

Ideal RCT Outcome (* implausible & untestable assumptions)

Rnd Group	Specific Effects	Non-Specific Effects	Total Effects
INT (↓sex risk)	S_{INT}	C*	S_{INT} + C*
TAC (lifestyle)	.*	C*	C*
Δ	S_{INT}	.*	S_{INT}*

Plausible RCT Outcome (* implausible & untestable assumptions)

Rnd Group	Specific Effects (S)	Non-Specific Effects (N)	Total Effects
INT (↓sex risk)	S_{INT}	N_{INT}	S_{INT} + N_{INT}
TAC (lifestyle)	.	N_{TAC}	N_{TAC}
Δ	S_{INT}	N_{INT} - N_{TAC}	S_{INT} + N_{INT} - N_{TAC}

. 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

† Pagoto S, McDermott MM, Reed G, Greenland P, Mazor KM, Ockene JK, Whited M, Schneider K, Appelhans B, Leung K, Merriam P, Ockene I (2013). Can attention control conditions have detrimental effects in behavioral medicine randomized trials? Psychosomatic Medicine, 75, 137-143.

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)

Rnd Group	Specific Effects (S)	Non-Specific Effects (N)	Total Effects
INT (↓sex risk)	S_{INT}	N_{INT}	S_{INT} + N_{INT}
TAC (lifestyle)	.	.*	.*
Δ	S_{INT}	N_{INT} *	S_{INT} + N_{INT}*

…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

Rnd Group	Specific Effects (S)	Non-Specific Effects (N)	Total Effects
INT (↓sex risk)	S_{INT}	N_{INT}	$S_{INT} + N_{INT}$
UC (as observed)	S_{UC}	N_{UC}	$S_{UC} + N_{UC}$
Δ	$S_{INT} - S_{UC}$	$N_{INT} - N_{UC}$	$S_{INT} + N_{INT} - S_{UC} - N_{UC}$

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)

† Silverman HJ, Miller FG (2004) Control group selection in critical care randomized controlled trials evaluating interventional strategies: An ethical assessment. Crit Care Med 32:852–857

‡ Vickers AJ & de Craen, AJM (2000). Why use placebos in clinical trials? A narrative review of the methodological literature. Journal of Clinical Epidemiology, 53, 157-161.

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

Theory Testing	Intervention Testing
Realm: Basic science	Realm: Applied science
Goal: Inform theory	Goal: Inform policy
Question: Is the theory supported?	Question: What works best?
Estimates of interest: Specific effects	Estimates of Interest: Total effects
Comparator: Control group identical to Intervention group except for theoretically postulated mechanisms	Comparator: Control group represents current practice. Often UC for efficacy-effectiveness & implementation trials
Mechanisms of action: Tested via Specific effects and/or mediation analysis	Mechanisms of action: Tested via mediation analysis

Evidence-based medicine requires medicine-based evidence

Knottnerus JA, Dinant GJ (1997). Medicine based evidence, a prerequisite for evidence based medicine. *British Medical Journal* (315) 309–10.

END