Hybrid Studies Should Not Sacrifice Rigorous Methods

Collaboratory Grand Rounds

October 6, 2023

David Murray, Ph.D.,
Associate Director for Prevention,
Director, Office of Disease Prevention
National Institutes of Health

NIH National Institutes of Health
Office of Disease Prevention

Today's Presentation

- The Complaint About Hybrid Studies
- The DECIPHeR Initiative
- Technical Assistance Workgroup
- Design and Analytic Issues
- Lessons Learned

The Complaint About Hybrid Studies

- Implementation research relies heavily on hybrid designs as suggested by Curran et al., 2012 and revised in Curran et al., 2022.
- “Hybrid Design” was an unfortunate choice of words, suggesting that implementation research had different methods than other research and might not be held to the same standard as other research.
- Instead, we should use the same rigorous methods for implementation research that we use for other research, changing only the focus.
- As with any other area, we will not advance the science with weak methods.

- Curran GM, et al. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012;50(3):217-26. PMID: 22310560.
- Curran GM, et al. Reflections on 10 years of effectiveness-implementation hybrid studies. Front Health Serv. 2022;2:1053496. PMID: 36925811.

The NHLBI DECIPHeR Initiative

- Key indicators of cardiovascular and pulmonary health outcomes vary markedly by race, ethnicity, sex or gender, geographic location, and SES.
- The Center for Translation Research and Implementation Science (CTRIS) at NHLBI launched the DECIPHeR Initiative in September 2019.
  - Disparities Elimination through Coordinated Interventions to Prevent and Control Heart and Lung Disease Risk (DECIPHeR)
- Two RFAs sought UG3 Clinical Centers and one U24 Coordinating Center.
- Awards were made in September 2020.
- Seven Clinical Centers transitioned to UH3 awards in September 2023.

- https://grants.nih.gov/grants/guide/rfa-files/RFA-HL-20-003.html
- https://grants.nih.gov/grants/guide/rfa-files/RFA-HL-20-004.html

Expectations for the Clinical Centers

- Test late-stage implementation research strategies for optimally and sustainably delivering proven-effective multilevel interventions to reduce or eliminate cardiovascular and/or pulmonary health disparities;
  - Employ validated theoretical or conceptual implementation research frameworks;
  - Include implementation research study designs;
    - e.g., experimental, quasi-experimental, observational, modeling, cluster randomization, stepped-wedge, Type III hybrid effectiveness, etc.
  - Include implementation measures as primary research outcomes;
    - e.g., acceptability, adoption, appropriateness, affordability, cost, feasibility, fidelity, reach, etc.
  - Inform understanding of key mediators and mechanisms of action of the implementation strategy.

Technical Assistance Workgroup

- Building on the model established by the NIH Pragmatic Trials Collaboratory, NHLBI created a Technical Assistance Workgroup.
  - David M. Murray and Jon Moyer, methodologists from the NIH Office of Disease Prevention.
  - James Troendle, Deputy Director of the Office of Biostatistics Research at NHLBI.
  - Patrick J. Heagerty and Feng-Chang Lin, biostatisticians from the Coordinating Center at UNC Chapel Hill, with Dr. Heagerty joining from the U of Washington.
  - June Stevens, Kim Truesdale, and Leslie Lytle, two epidemiologists and an interventionist from the Coordinating Center.
- The TA Workgroup met with each Clinical Center team 6-7 times during the 2nd half of the UG3 phase to review their UH3 plans and provide advice.
- The goal was to help the Clinical Centers have the strongest possible application for the UH3 phase.

Technical Assistance Workgroup

- The TA Workgroup began with the specific aims for each project.
- Once the aims were clear, the workgroup turned to the study design.
- That often led to revisiting the aims.
- Once the aims and design were aligned, the workgroup turned to the statistical analysis plan.
- That often led to revisiting the design.
- Once the design and analytic plan were aligned, the workgroup turned to the power analysis.
- That often led to revisiting the analytic plan and sometimes the design.
- In the end, the aims, design, analytic plan, and power analysis were aligned.
- The TA Workgroup reviewed each Clinical Center's protocol before it went to the DSMB and after it was revised based on feedback from the DSMB, just before review by NHLBI for transition to the UH3 phase.

Design and Analytic Issues

- Emphasis on implementation outcomes
- Research designs for Type I, II, and III Hybrid Studies
- Intervention vs implementation strategies
- The need to address clustering
- Cross-classification and multiple membership
- Time-varying intervention effects
- Data based parameter estimates
- Blinding
- Adaptations of intervention and implementation strategies

Emphasis on Implementation Outcomes

- Trials in implementation science study both health outcomes and implementation outcomes.
- Implementation outcomes often include measures of acceptability, adoption, appropriateness, affordability, cost, feasibility, fidelity, reach, etc.
- All DECIPHeR Clinical Centers initially included implementation measures, but usually as secondary outcomes.
- NHLBI expected the Clinical Centers to use an implementation measure as their primary outcome.
- The TA Workgroup worked with the Clinical Centers to select an implementation measure as the primary outcome.
- Four chose to measure reach, two adoption, and one fidelity as the primary outcome.

Example

- Tulane began with a study comparing a community-health-worker-led multifaceted intervention to reduce cardiovascular risk to an enhanced usual-care arm with a health measure as the primary outcome.
- Tulane gradually moved to delivering the same multifaceted intervention to reduce cardiovascular risk in two arms that differed in their implementation strategy.
- The revised estimand focuses on the effect of the implementation strategy, not the effect of the intervention.
- The primary outcome is intervention fidelity, an implementation measure.

Example

- UCLA began with a stepped-wedge group-randomized trial with 3 sequences, five periods, a clinic-based intervention to manage blood pressure, and a health measure as the primary outcome.
- UCLA gradually moved to a parallel group-randomized trial,
  - In Year 1, one arm will receive usual care and two arms will receive the same intervention with different implementation strategies.
  - In Year 2, the usual care clinics will employ both implementation strategies whlle the other two arms will cross to the other implementation strategy.
- The estimands in the first year are focused on the effect of the implementation strategy and the effect of the intervention. The estimand in the second year is focused on the effect of the sequence of implementation strategies.
- The primary outcome is adoption, an implementation measure.

Research designs for Type I, II, and III Hybrid Studies

- Curran et al. (2012) introduced the hybrid effectiveness-implementation designs.
  - Hybrid Type I tests a clinical intervention while gathering information on implementation.
  - Hybrid Type II simultaneously tests a clinical intervention and an implementation intervention or strategy.
  - Hybrid Type III tests an implementation intervention or strategy while gathering information on effectiveness.

- Curran GM, et al. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012;50(3):217-26. PMID: 22310560.

Research designs for Type I, II, and III Hybrid Studies

- Curran et al (2022) updated their original description of hybrid designs, labeling them as hybrid studies without offering designs for each type.
- The TA Workgroup, in a discussion led by June Stevens, suggested design prototypes for each of the three types (cf., Stevens et al, under review).
  - Type I requires at minimum a two-arm trial with a comparator for the intervention.
    - Intervention vs No Intervention
  - Type II requires at minimum a three-arm trial with a comparator for the intervention and a comparator for the implementation strategy.
    - No Intervention vs Intervention vs Intervention with Enhanced Implementation Strategy
  - Type III requires at minimum a two-arm trial with a comparator for the implementation strategy.
    - Intervention vs Intervention with Enhanced Implementation Strategy

- Curran GM, et al. Reflections on 10 years of effectiveness-implementation hybrid studies. Front Health Serv. 2022;2:1053496. PMID: 36925811.
- Stevens J, et al. Design of a Dual Randomized Trial in a Type 2 Hybrid Effectiveness-Implementation Study. Implementation Science. under review.

Research designs for Type I, II, and III Hybrid Studies

- All DECIPHeR Clinical Centers proposed hybrid studies.
  - Five of the seven Clinical Centers described their study as a Hybrid Type II study.
  - One described their study as a Hybrid Type III study.
  - The seventh did not label their Hybrid Type.
- Close examination revealed that most were Hybrid Type I studies.
  - These studies had a comparator for the intervention but not for the implementation strategy.
- With the focus of NHLBI on Type III studies, the TA Workgroup worked with each Clinical Center to develop designs that provided the appropriate comparators.
  - Six Clinical Centers shifted to Type III studies with a comparator for the implementation strategy and one shifted to a Type II study with comparators for both the intervention and the implementation strategy.

Example

- Colorado wanted to test their intervention and their implementation strategy.
- The TA Workgroup worked with Colorado to develop a design that provided a comparator for both.
- In Year 1, there will be 3 arms in a parallel group-randomized trial design.
  - Usual care
  - Intervention via standard implementation
  - Intervention via enhanced implementation
- Usual care provides a comparator for the intervention delivered via standard implementation strategy.
- Intervention via standard implementation provides a comparator for the enhanced implementation strategy.
- This is a Hybrid Type II study in Year 1.

Example

- In Year 2 at Colorado, the usual care clusters will be randomized to one of the two implementation strategies.
- This will create a 2-arm parallel group-randomized trial design.
  - Intervention via standard implementation
  - Intervention via enhanced implementation
- Intervention via standard implementation provides a comparator for the enhanced implementation strategy.
- This is a Hybrid Type III study beginning in Year 2.
- The analysis will reflect the duration of the intervention, with a lower dose in the clusters that began in usual care.

Intervention vs Implementation Strategies

- The usual trial evaluates a single intervention strategy delivered with a single implementation strategy as a package and it is not possible to distinguish the effects of the two strategies.
- In contrast, implementation trials compare intervention strategies and/or implementation strategies.
- Distinguishing these strategies is essential to evaluate the effect of the intervention strategies and/or the effect of the implementation strategies (Pinnock et al., 2017a, 2017b).

- Pinnock H, et al. Standards for Reporting Implementation Studies (StaRI) Statement. BMJ. 2017a;356:i6795. PMID: 28264797.
- Pinnock H, et al. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. BMJ Open. 2017b;7(4):e013318. PMID: 28373250.

Intervention vs Implementation Strategies

- Experts in intervention and implementation strategies joined the TA Workgroup.
- The discussion focused on distinguishing intervention and implementation strategies.
- The discussion also focused on distinguishing the strategies used in different implementation approaches.
- The goal was to have the same intervention delivered using at least two different implementation strategies at each Clinical Center.

Example

- Tulane began using a single intervention delivered using a single implementation strategy.
- This confounded the intervention and implementation strategies and is typical of trials that are not focused on implementation.
- The TA Workgroup helped Tulane distinguish intervention vs implementation strategies.
- The TA Workgroup helped Tulane identify two distinct implementation strategies (community-health-worker led vs group led) delivering the same intervention (multifaceted cardiovascular risk reduction).

Example

- Northwestern proposed the Kaiser Bundle as their intervention.
- It includes components that focus on change at the system level, e.g., clinic-wide adoption of a comprehensive hypertension registry.
- Distinguishing between these system-level intervention activities and system-level activities designed to encourage implementation was challenging.
- The TA Workgroup helped Northwestern develop a plan to compare two different implementation strategies (practice facilitation vs no practice facilitation) delivering the same intervention (the Kaiser Bundle).

The Need to Address Clustering

- Most of the Clinical Center applications proposed a parallel group- or cluster randomized trial.
  - Positive intraclass correlation is expected among observations taken on members of the same group or cluster, violating the assumption of independence of errors that underlies the usual analytic methods.
  - Alternative methods are required for analysis and sample size to avoid an underpowered study or an inflated Type I error rate (e.g., Murray, 1998).
  - The TA Workgroup helped Clinical Centers address clustering due to assignment of identifiable groups to study arms.
  - The https://researchmethodsresources.nih.gov/ website provides a sample size calculator for parallel group- or cluster-randomized trials.

- Murray DM. Design and Analysis of Group-Randomized Trials. New York, NY: Oxford University Press; 1998.

The Need to Address Clustering

- One Clinical Center proposed an individually randomized group-treatment (IRGT) trial.
  - Participants are randomized to interventions but interact with other participants or with a shared intervention agent post-randomization.
  - Some level of positive intraclass correlation will develop among participants who interact with each other or with the same intervention agent.
  - Alternative methods are required for analysis and sample size to avoid an underpowered study or an inflated Type I error rate (cf., Pals et al, 2008).
  - The TA Workgroup helped that Clinical Center address clustering due to the method of intervention delivery.
  - The https://researchmethodsresources.nih.gov/ website provides sample size calculator for individually randomized group-treatment (IRGT) trials.

- Pals SL, et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. Am J Public Health. 2008;98(8):1418-24. PMID: 18556603. Errata: Am J Public Health. 2008;98(12):2120.

Example

- Northwestern's application proposed a quasi-experimental design with one area of Chicago to get the intervention and another to serve as a control.
  - This one-group-per-condition design has no valid analysis (Varnell et al., 2001).
- The TA Workgroup helped Northwestern understand the weaknesses of their proposed design and to consider alternatives.
- Northwestern moved to a parallel group- or cluster-randomized trial comparing two implementation strategies (practice facilitation vs no practice facilitation) using the same intervention strategy (the Kaiser Bundle).
- Northwestern used the https://researchmethodsresources.nih.gov/ website to perform their sample size calculations.

- Varnell SP, et al. An evaluation of analysis options for the one-group-per-condition design. Can any of the alternatives overcome the problems inherent in this design? Eval Rev. 2001;25(4):440-53. PMID: 11480307.

Cross-Classification and Multiple Membership

- Cross-classification occurs when the same intervention agent interacts with participants in more than one unit of randomization.
- Multiple membership occurs when participants receive some components of their intervention from one intervention agent or in one setting and other components from a different agent or in a different setting, all within the same unit of randomization.
- Both create an additional source of variation due to the intervention agent or setting that must be accounted for in the analysis and sample size methods to avoid an underpowered study or an inflated Type I error rate (Barker et al., 2020; Cafri et al., 2015, Teerenstra et al., 2023).

- Barker KM, et al. Cross-classified multilevel in health research: A systematic review of published empirical studies and recommendations for best practices. SSM - population health. 2020;12:100661. PMID: 32964097.
- Cafri G, et al. An introduction and integration of cross-classified, multiple membership, and dynamic group random-effects models. Psychol Methods. 2015;20(4):407-21. PMID: 26237504.
- Teerenstra, S et al. Sample size for partially nested designs and other nested or crossed designs with a continuous outcome when adjusted for baseline. Stat Med. 2023;42(19):3568-92. PMID: 37348855.

Example

- Hopkins and Michigan proposed an individually randomized group-treatment (IRGT) trial.
- Individuals were to be randomized to study arms but would interact with a shared intervention agent during the course of the trial.
  - That meant the trial was an IRGT rather than an RCT.
- Some participants would interact with a second shared intervention agent during the course of the trial.
  - That would create a multiple-membership situation for those participants.
- The TA Workgroup helped the Clinical Center develop appropriate sample size and analytic plans to reflect both the IRGT design and the multiple-membership structure.

Example

- Colorado will randomize school nurses to study arms and each nurse will work with one or two schools.
- Colorado also will employ navigators who will work with several nurses in the same study arm.
- As such, participants will be cross-classified by both nurses and navigators.
- The TA Workgroup helped Colorado recognize this issue and address it in their analytic plan by including an additional random effect for navigators and by using Kenward-Rogers (2009) degrees of freedom.

- Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. Computational Statistics and Data Analysis. 2009;53:2583-95.

Example

- After randomizing churches to study arms, Tulane planned to have a limited number of community health workers manage a large number of churches.
- Participants would have been cross-classified by community health worker and by church.
- The TA Workgroup helped Tulane understand the challenges of cross-classification and Tulane moved to having a separate community health worker for each church.
- That avoided the cross-classification issue entirely.
- Alternatively, Tulane could have modeled the CHW as an additional source of variation, but that would have reduced power.

Time-varying intervention effects

- An intervention effect may change over time for a variety of reasons.
- These patterns are readily addressed for parallel group- or cluster-randomized trials so long as there are enough measurement occasions.
  - The sample size and analytic methods can account for the expected pattern.
  - This is also true for individually randomized group-treatment (IRGT) trials.
- We learned recently that this can be particularly problematic in a SWGRT.
  - The standard analytic methods assume a rapid and persistent intervention effect.
  - If the effect varies over time, the standard analytic methods can give wildly misleading results (Kenny et al., 2022).
  - Kenny et al. (2022), Maleyeff et al. (2022), Hughes et al. (2023) offer solutions.

- Kenny A, et al. Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect. Stat Med. 2022;41(22):4311-39. PMID: 35774016.
- Maleyeff L, et al. Assessing exposure-time treatment effect heterogeneity in stepped-wedge cluster randomized trials. Biometrics. 2023;79(3):2551-64. PMID: 36416302.
- Hughes JP, et al. Sample Size Calculations for Stepped Wedge Designs with Treatment Effects that May Change with the Duration of Time under Intervention. Prevention. 2023. PMID: 37728810.

Example

- NYU proposed stepped-wedge design to evaluate their enhanced implementation strategy to improve adoption of a proven-effective blood pressure management program.
  - At the time of their application, Kenny et al. had not been published.
  - The TA Workgroup worked with NYU to understand the problem and the alternative methods proposed in Kenny et al. and Maleyeff et al.
  - There was no sample size calculator available that implemented the Kenny et al. solution so NYU had to develop one on their own.
  - In the end, the analytic plan and power methods were aligned with the stepped wedge design and the expected pattern of the intervention effect over time.
  - The NYU team co-authored a paper with the Kenny et al. team that describes the solution they developed (Hughes et al.).
  - The https://researchmethodsresources.nih.gov/ website provides a sample size calculator that implements the Kenny et al. and Hughes et al. solutions for stepped-wedge group-randomized trials.

Data-Based Parameter Estimates

- Data-based estimates are recommended for the parameters needed for sample size calculations.
- Such estimates are increasingly available for health outcomes but rarely available for implementation outcomes.
- The TA Workgroup helped the Clinical Centers consider how to use existing data to develop the necessary parameter estimates.

Example

- Northwestern did not have any estimates available for the parameters needed for their sample size calculation, in part because of the substantial change in their study design.
- The TA Workgroup advised that the best estimates would come from the same clinics that were going to participate in their Hybrid Type III study, from the same population of participants, and from variables defined and measured as they would be in the Type III study.
- Northwestern was able to work with their clinics to develop those estimates before they had to submit their protocol to the DSMB for review.
- The TA Workgroup had helped them run through the methods they would use for their sample size calculations using non-data-based estimates so they would be ready to go when the data-based estimates were available.

Example

- NYU did not have any estimates available for the parameters needed for their sample size calculation, in part because of the substantial change in their analytic methods.
- The TA Workgroup advised that the best estimates would come from the same clinics that were going to participate in their Hybrid Type III study, from the same population of participants, and from variables defined and measured as they would be in the Type III study.
- NYU was able to access data from a previous stepped wedge trial focused on blood pressure management to develop their estimates.
- NYU was able to use those estimates for sample size calculations using the revised analytic methods planned for their stepped-wedge trial.

Blinding

- Blinding of study arm assignment is routine in most clinical trials.
- Blinding is more challenging in trials involving complex interventions or those conducted in community settings.
- Implementation studies do not have established practices for blinding.
- NHLBI developed guidance for the DECIPHeR Clinical Centers on blinding:
  - No one outside the unblinded study statistician and the DSMB will see trial results by study arm until all data have been collected and cleaned and the RCC announces that the blind can be broken, except as noted below.
  - Data on primary and secondary outcomes that typically serve as process variables (e.g., reach, adoption, fidelity to the intervention) can be shared only with implementation intervention staff, only for the single implementation arm in which they work, and only for the purposes of encouraging adherence to the approved protocol.
  - Feedback to stakeholders should be based on data different from the primary and secondary outcomes.

Adaptation

- Adaptation of intervention strategies is uncommon in most clinical trials.
- Adaptive intervention trials allow adaptations of the intervention protocol based on a process that is prespecified in the protocol and describes what changes can be made and what will trigger those changes.
- Implementation studies do not have established practices for adaptation.
- NHLBI developed guidance for the DECIPHeR Clinical Centers on adaptation:
  - Any changes in the protocol, including changes to the implementation intervention, must be approved in advance by NHLBI and the DSMB.
  - If the trial anticipates that changes may be required for the implementation intervention, the trial is strongly encouraged to follow standard practice for an adaptive intervention trial by prespecifying, in the protocol, the process by which such changes will be made.

Lessons Learned

- Implementation research has its own practices with regard to outcomes, intervention and implementation strategies, study design, analysis, blinding, adaptation, and other issues.
- Consensus is lacking in many areas, even within the implementation research community.
- Methodologists outside the implementation community often do not understand the features common to implementation research.
- Bringing those two communities together in a careful review of the proposed studies led to appreciable modifications of each study.
- Involvement of methodologists familiar with clustered designs and their analytic and power issues was particularly important.
- The result was a much stronger set of proposals for the UH3 phase of DECIPHeR.