Go Big or Go Home?
Designs and methods for studying scale up and spread

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

J.D. Smith, PhD
Associate Director, Center for Prevention Implementation Methodology for Drug Abuse and HIV
Associate Professor, Feinberg School of Medicine, Northwestern University
Greenhalgh and Papoutsi, *BMJ*, May 2019:

“…there is no simple or universally replicable way of implementing change at scale in a complex system. A technology or pathway that works smoothly in setting A will operate awkwardly (or not at all) in setting B.”
OK, so we **go home** then?
No, wait...

• This is exactly why we still have a lot of work to do!
• This is why we need to design strong scale up STUDIES to see what works and what doesn’t!
• So what designs should we consider? *(go big, then?)*
  – Well, *that depends*…
    • How much money you got?
    • How much time you got?
    • What do your system/community partners need/want?
    • How variable are the “units” within your system/community?
Looking at UP not OUT

• Talking about Scale-**UP** today
  – Same thing [intervention], same/similar settings, same population
  – Focus on getting the thing *spread*
    • Measure *how much* and *how well* the thing gets done
  – We expect effectiveness of the thing to be the same/similar
    • So, might not even measure that part

• Will NOT cover Scale-**OUT**
  – Adapting the thing for new people/places/targets
How do we get to Scale Up?

• Present briefly on 2 frameworks:
  – US VA QUERI: 4 Phase Pipeline Framework
    • US Department of Veterans Affairs (national healthcare system for Veterans)
    • Quality Enhancement Research Initiative
  – Institute for Healthcare Improvement (IHI) Framework for Going to Full-Scale

• Both include *testing* scale up and then *doing* scale up
  – Assumption: already have best practice/intervention to be scaled
US VA QUERI: 4 Phase Pipeline Framework

- **Sequence** of implementation projects from initial feasibility assessment to national roll-out:
  - **Phase 1**: Pilot project to develop/refine an implementation program and assess basic feasibility (1 clinic or facility)
  - **Phase 2**: Small trials to further refine and evaluate an implementation program (4-6 facilities in 1-2 regions)
  - **Phase 3**: “Regional roll-out” projects (10-20 facilities in 3-5 regions)
    - Preparing for “hand-off” of implementation to operational partners
    - Integrate partners in implementation program
  - **Phase 4**: “National roll out” (all facilities/locations)
    - Operational partners support implementation fully
Institute for Healthcare Improvement (IHI)
Framework for Going to Full-Scale

• **Phases of Scale-Up:**

- **Best Practice Exists**
- **Set-Up**
  - Early demonstration phase
  - Highly iterative
  - Figuring out “all parts needed” for scale, but testing in small # units
- **Develop Scalable Unit**
- **Test Scale-Up**
  - More units
  - Range of units
  - Still learning
  - *Compare strategies?*
- **Go to Full Scale**
  - Lots of units
  - Adding units
  - Less learning
  - Using strategies
Institute for Healthcare Improvement (IHI) Framework for Going to Full-Scale

• **Phases of Scale-Up:**


**Diagram:**

- Best Practice Exists
- Set-Up
- Develop Scalable Unit
- Test Scale-Up
- Go to Full Scale

**Note:** Designs covered today mostly for here...
An Overview of Research and Evaluation Designs for Dissemination and Implementation


The Annual Review of Public Health is online at publhealth.annualreviews.org

https://doi.org/10.1146/annurev-publhealth-031816-044215
Design Categories

• **Within-Site Designs**
  – We use these when we want or need to expose all places to the same strategy (or package)
  – No comparison places (compare to prior levels of performance)

• **Between-Site Designs**
  – We use these when we want to expose places to different strategies
  – Causal inference improved when places are randomized to different exposures (not always possible or called for)

• **Within- and Between-Site Designs**
  – Use of crossover; begin with one strategy/condition and then move to another
  – Stepped wedge designs
  – “Roll out” implementation designs
For “test scale up”, consider **comparing** strategies:

- **Remember:** Assuming you’ve already done some smaller “scalable unit” iterative stuff
- Let’s start with some **Between-Site Designs**
  - “New Strategy vs. Implementation as Usual” randomized trial
  - “Head to Head” randomized implementation trial of 2 new strategies
OK, so WAIT A SEC…

• You don’t certainly don’t have to compare strategies during a “test scale up” phase
• Common to iterate “one” strategy (package) using within-site designs all the way
• But, if you have the desire, funders, and supportive partners, consider designs that allow comparing strategies across places
So, Head to Head combat!
Example of *Head to Head*: Go NAPSACC (NIH R01)

• Intervention = diet and exercise program in child care settings
• Regional “Technical Assistance providers” (TAs)
  – Trained to initiate and support uptake at centers
  – Existing role within the state system
• Compare Basic vs. Enhanced implementation strategies across 96 childcare centers (targeted mostly at TAs)
  – Basic = online tools to train and support TAs (and programs)
  – Enhanced = basic + facilitation (problem solving, audits, etc...)
TA Providers randomized to received Basic or Enhanced strategy

Then they work in *48* Centers per arm

BOTH TA Providers and Centers measured on implementation outcomes
Example 2-3: REDUCE Trial, eCRT SStudy

- Juszczyk et al., 2016; Gulliford et al., 2014
- Reduce antibiotic prescribing in primary care
  - Decision-support training, reminder, patient sheet, feedback in EMR (all remote delivery) (about 50 clinics)
  - Vs. Usual care (about 50 clinics)
- All remote delivery of “simple” implementation strategies could speed the time to scale up
- Could also lend it self to larger trial sizes
- 100 units seems a common target for cRCTs in D&I
What if you can’t get a large # of units?

- Consider matched-pair randomized designs
  - Improves power with lower number of units
    - Match for balance between pairs
  - Common in VA during “regional roll-out” phase
  - Useful with 12+ units *(H. Brown, personal communication)*

- Consider “Roll Out of Repeated Pairs” design:

Wyman et al., 2015
Brown et al., 2009
Pros/Cons of randomized implementation trials for testing scale up

• Compared to non-randomized designs, better case for causality of your strategy

• If you specified your “scalable unit” well, positive trial could be good precursor to “going to scale” evaluation
  – The idea with both GoNAPSACC and REDUCE examples...

• Randomization in D&I research can be a difficult “sell”

• Strategies are often treated as fixed, so usually not allowing/encouraging/addressing adaptation
  – There are some options for this, including...
Let’s Get SMART
More Between-Site: SMART

- Sequential Multiple Assignment Randomized Trial
- Built on the notion of **stepped-care** (adaptive approach)
  - Begin with low intensity strategy, go up based on response
Real SMART: Kilbourne et al., 2018

Multi-arm trial alert!
Pros/Cons of SMART for testing scale up

• Built in evaluation of adapting strategy approach
• Learn *why* different types of places need different strategies
  – Aided by assessment of context
• Learn these things more quickly
  – Same study informs about what different contexts might need
  – Address “failing” sites now, not waiting until next study…
• Randomization problems like talked about before…
• How to set threshold of success/failure?
• Indeed *Go Big*…
  – Often big trials, multiple randomizations, big team…
• Not too popular yet in D&I but growing
Wedges anyone?

- Type of crossover design: start Strategy A then switch to Strategy B
- Timing of start randomized
Randomized **Stepped Wedge** Implementation Trial
Comparing Two Strategies (n=20 units)

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th></th>
<th></th>
<th>Year 2</th>
<th></th>
<th></th>
<th>Year 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
</tr>
<tr>
<td>COHORT 1 (n = 4)</td>
<td>c</td>
<td>c</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>COHORT 2 (n = 4)</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>COHORT 3 (n = 4)</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>i</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>COHORT 4 (n = 4)</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>COHORT 5 (n = 4)</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>i</td>
</tr>
</tbody>
</table>
Randomized *Roll Out* Implementation Trial Design
*(Incomplete wedges; n= 56 units: 7 clusters, 8 units each)*

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td>Q1</td>
</tr>
<tr>
<td>C</td>
<td>c</td>
<td>C</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Cluster 4</td>
<td>Cluster 5</td>
</tr>
</tbody>
</table>

- Reduces respondent burden
- Increases representativeness of the baseline/pre-implementation
- Can stage enrollment and keep going if needed for power or demand within a system
### Randomized Roll Out Optimization Trial Design

Implementing Guidelines for Pediatric Hypertension Diagnosis

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
<td>Q1</td>
</tr>
<tr>
<td>Cluster 1 (6 units)</td>
<td>C1 Strategy Development</td>
<td>C1 (Initial) Implementation Strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 2 (6 units)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 3 (6 units)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 4 (6 units)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q2</td>
<td>Q3</td>
<td>Q4</td>
</tr>
<tr>
<td>Cluster 2 (6 units)</td>
<td>C2 Optimized Implementation Strategy #1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 3 (6 units)</td>
<td>C3 Optimized Implementation Strategy #2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 4 (6 units)</td>
<td>C4 Optimized Implementation Strategy #3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data for strategy optimization

Implementation Data

Pre-Implementation Data

---

### Analysis:

Time to achievement of criterion between Initial (C1) and Later (C2–C4) implementation strategy (primary outcome: rate of guideline-concordant diagnosis of hypertension)

Smith, Brown, et al., 2019

***NS***

Add more clusters and keep optimizing

Go to Full Scale!
Pros/Cons of Wedge/Rollout Designs

• Intuitive appeal, each cluster ultimately gets the hypothesized winning strategy and acts as its own control
• Clinical/Community partners can prefer this over cRCT
• Phased delivery of strategies more practical, sometimes needed
• Optimized version allows adaptation
• Pretty darn popular these days; lots of examples
• Can be slower than cRCTs to findings
• Temporal trends potentially more problematic…
  – But you can control for these statistically (to some extent)
Go to Full Scale!
OK, so Go to Full Scale

- Looking at *Within-Site Designs* here
  - One strategy (package) used “everywhere”

- Commonly seen when decision has been made to “do the thing” in a system *(county, state, country…)*
  - Guideline (to be) adopted
  - Policy mandate

- Common designs used here:
  - Post-only; Pre-Post
  - Interrupted Times Series *(pre, pre, pre… / post, post, post…)*
  - Network analysis of the spread/scale up process *(observational or experimental designs)* *(Valente et al. 2015)*
Prevention of Mother-to-Child [HIV] Transmission Program (PMTCT; South Africa)

- Barron, 2015
- Intervention = testing, initiating anti-retrovirals, follow-up
- Implementation Strategy = *Learning Collaborative* plus *package* of implementation strategies
- Followed IHI scale up phases
  - Scalable unit: health district (hospital and 20+ feeder primary care clinics)
  - 3 districts, to 8, *to all 52 districts*
- Iterating collaborative activities, package of strats, and use of a “campaign”
VA Roll Out example: ReachVet

- Nationwide roll-out of *suicide prevention intervention* (N = 144 healthcare systems)
- Initial plan was phased pre-post by region
  - CHANGED to “all at once” due to ethics concerns, political will ($$), use of national data systems and training systems
- Implementation strategies
  - Policy memo, Dashboard, local coordinator, web-based training, education materials, technical assistance
- ADDED *research-funded* Stepped Wedge of +facilitation
  - For under-performing sites (4 lowest performers in 7 regions)
<table>
<thead>
<tr>
<th>VISN A</th>
<th>FY17</th>
<th>Year 1 (FY18)</th>
<th>Year 2 (FY19)</th>
<th>Year 3 (FY20)</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>Aug. 29th</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>Aug. 30th</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>Aug 31st</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>Sept. 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site X</td>
<td>Sept. 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VISN G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What’s that design called...?

• So, the roll-out evaluation and SW add-on changed a number of times during the planning AND post-funding of the research grant...

• “Sara cries under her desk until she understands this is *highly* partnered research” design
Wrap Up for Scale Up

• Mostly dealt with designs for comparing strategies across units— for “testing scale up” phase
• Included some design options which allow us to test adapted strategies
• Didn’t cover mediator/moderator stuff
  – But DO IT! Test mechanisms, preconditions, conditional effects...
• Didn’t cover qualitative or MM process evaluations
  – But DO THEM! Helps understand mechanisms of action, complexity, why worked here but not there
  – See MRC guidance (Moore et al., 2015)
• Didn’t cover observational research approaches
Drivers of Selecting Designs
(Non-exhaustive....)

• Partners
  – What do they need? What will they agree to?
  – “…design is only as good as what your partners allow.”

• Practicality
  – Can you randomize? Want to randomize?
  – How much time, money, expertise, staff... you got?
  – How many units do you have access to?

• Power
  – How many units do you have access to?
  – How much power do you NEED?
    • Who is asking? Who cares about this?
    • Power analysis program for randomized roll-out implementation trial designs is forthcoming from Brown, Smith, et al. (STAY TUNED!)
Had enough? Me, too...
Many thanks to:

- C. Hendricks Brown, PhD
- John Landsverk, PhD
- Brian Mittman, PhD
- Niajua Duan, PhD
- Greg Aarons, PhD
- David Chambers, DPhil
- Ken Wells, MD
- Songthip Ounpraseuth, PhD
- Keith Williams, PhD
- Ken Wells, MD
Questions? Comments?
Questions? Comments? Heckling?
Scale OUT

• Adapt intervention for new population, context, or target...

• Aarons et al (2017)
  – Don’t need whole new effectiveness trial
  – “Borrow strength” from evidence of impact in prior trial(s)
  – Levels of evidence
    • Test implementation and intervention fidelity
    • Test underlying mechanisms (mediators) present in prior work

• and/or Consider effectiveness-implementation hybrid design