Institute for Healthcare Improvement

IHI//BMJ Forum, Copenhagen
April 2023

C1: **Designing for Success, Evaluating for Learning and Credibility**

Don Goldmann, MD

Chief Scientific Officer Emeritus & Senior Fellow, IHI

Professor of Pediatrics, Harvard Medical School

Professor of Epidemiology, Harvard TH Chan School of Public Health

@DAGoldmann

dgoldmann@ihi.org

Pierre Barker, MD, MB ChB

Chief Scientific Officer, IHI

Clinical Professor, Maternal and Child Health, Gillings School of Global Public Health, UNC Chapel Hill USA

pbarker @IHI.org

Disclosure

Pierre Barker and Don Goldmann *have no relevant financial relationship(s) to disclose with eligible companies whose primary business is producing, marketing, selling, re-selling, or distributing healthcare products used by or on patients.*

Session Objectives

After attending this session, attendees will be able to:

1. Appreciate current gaps in the rigor and credibility of many major QI programs
2. Understand and apply methods for mitigating bias and strengthening causal inference
3. Understand the importance of collaboration among designers, implementers, and evaluators
4. Understand and apply rigorous approaches to designing QI programs that have a higher likelihood of yielding credible results
5. Understand, explain, and test evolving IHI frameworks for putting these rigorous methods into practice and improving the effectiveness and credibility of our QI work
6. Understand, explain, and test commonly used evaluation frameworks and measures

Roadmap

- The current (problematic) state of QI design, implementation, and evaluation
- Epidemiological principles for mitigating bias and strengthening causal inference
- Causal (program) theory and how to display it (driver diagrams, logic models)
- Commonly used evaluation and measurement frameworks
- Evolving IHI frameworks for building rigorous designs and evaluating the results in order to improve learning and credibility of our QI work
- Study design options and their strengths and limitations for demonstrating causality
- Synthesis and conclusions

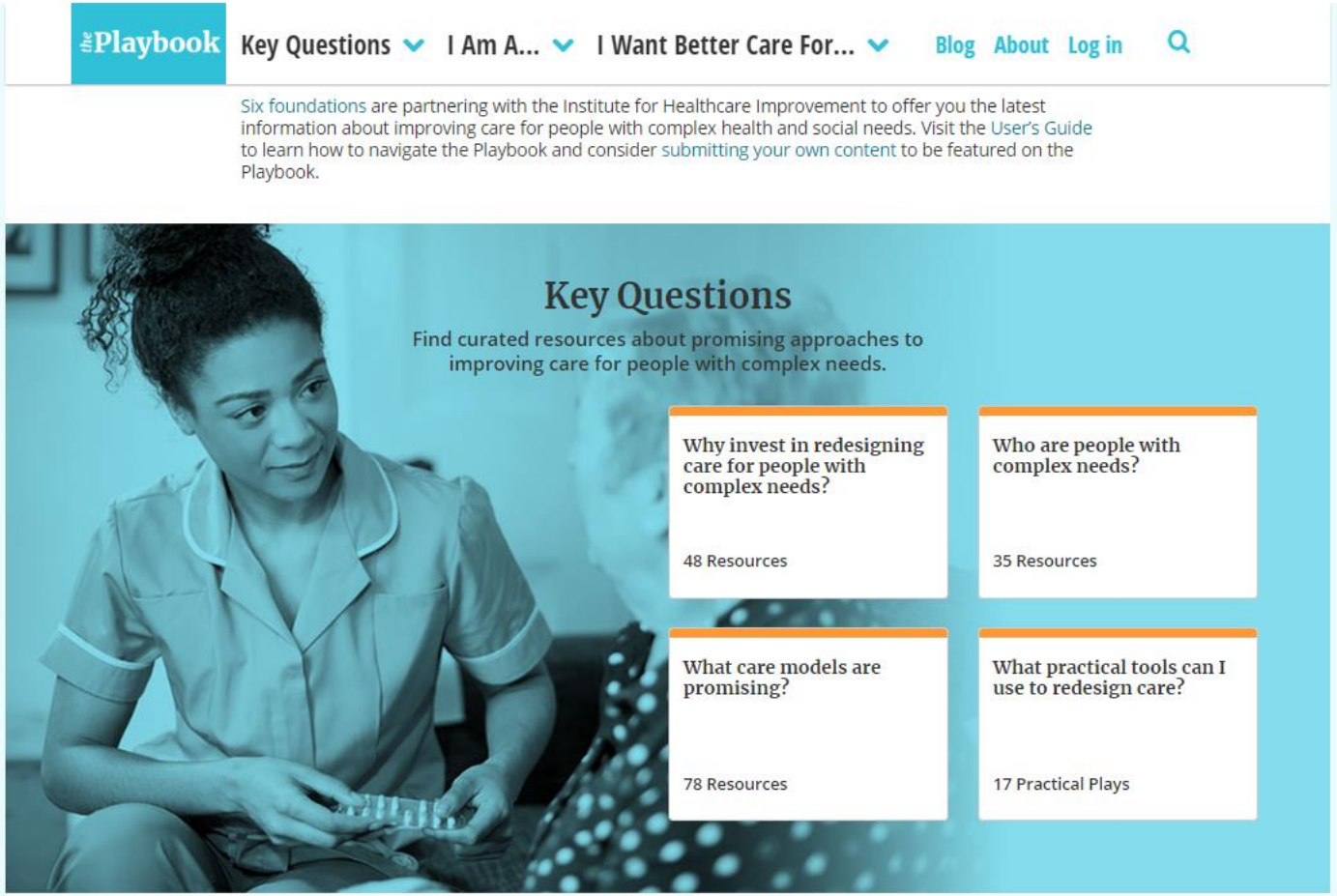
The Current Credibility Gap: Opportunities for Improving Design and Evaluation of Improvement Programs

- Why do so many promising 'bright spots" and innovative programs (including those that appear to be successful in "collaboratives") disappoint when scaled up?
  - Example: ELC and EPOCH emergency laparotomy programs
  - Other examples (for self-study, citations below):
    - Camden Coalition "Hotspotting"
    - Peter Pronovost's Keystone CLABSI collaborative and "Matching Michigan"
    - Atul Gawande's surgical safety checklist as implemented in Ontario
- Can we evaluate the context in which we will test and implement a new practice in advance rather than just "learning as we go?"
  - Example (for self-study): ZamCAT chlorhexidine umbilical care in Africa

Finkelstein, et al. Health care hotspotting – a randomized, controlled trial. doi: 10.1056/NEJMsa1906848
Pronovost, et al. An intervention to reduce catheter-related bloodstream infections in the ICU. doi: 10.1056/NEJMoa061115
Urbach, et al. Introduction of surgical safety checklists in Ontario, Canada. doi: 10.1056/NEJMsa1308261
Dixon-Woods, et al. Explaining Matching Michigan: an ethnographic study of a patient safety program. doi: 10.1186/1748-5908-8-70

Caution – Most Innovative Models to Improve Care Rest on Shaky Evidence: “The Better Care Playbook”

- > 100 care models,
- Only a few with **strong evidence**
- Only a few **replicated successfully**
- None prioritized **equity**



Based on Cochrane Collaboration EPOC (Effective Practice and Organisation of Care), GRADE, and other sources

Perhaps We Need…

- More humility about our promising findings
- More attention to basic principles of epidemiology
  - Association v. causation: causal inference and "counterfactuals"
- Clear display of our causal theory (aka "program" theory) with driver diagrams, logic models, and other methods
- Stronger designs, including randomized designs
- Better evaluation of why a program did or did not work, including process evaluation, established and evolving evaluation frameworks, and qualitative and ethnographic methods

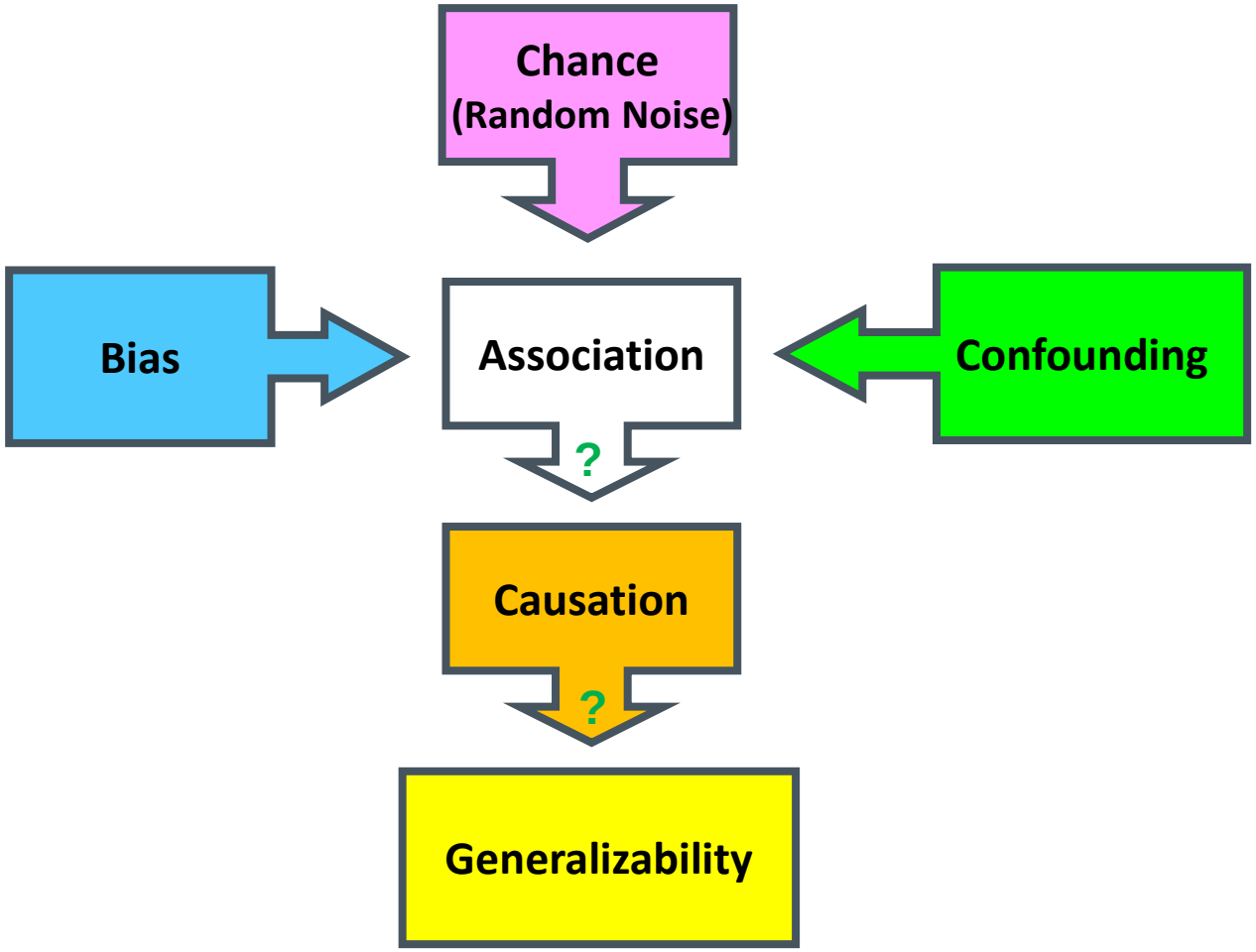
Epidemiology Informs Sound Design and Evaluation

Association and Causation in Improvement Science

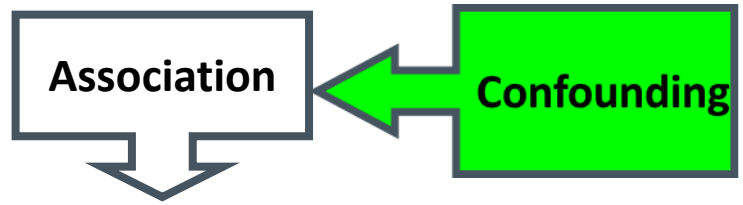
Be careful attributing improved outcomes to your interventions without formally considering how the improvements you are claiming may not be causal and could be misleading or even wrong

Whenever we infer cause from association, we can be wrong – in *three ways*

When we act based on associations, we can be wrong in three ways…

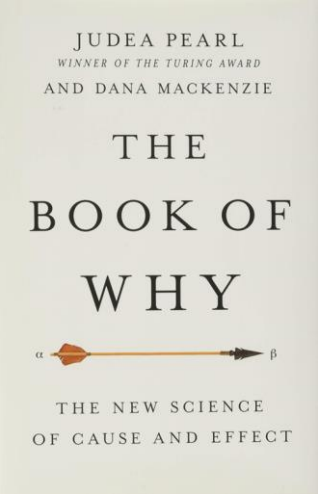


Properties of Confounders



- A confounder must be a cause (or be a risk factor for) the outcome
- The confounder must be related to the exposure (intervention)
- The confounder must **not** be on the causal pathway between the exposure and outcome (not an "*effect modifier*" or "*mediator*")
- Case mix/severity is a classic confounder

"Directed Acyclic Graphs" (DAGs) show causal inference, including potential confounding variables.
Highly recommended: The Book of Why, Judea Pearl



Smoking and Lung Cancer: Association versus Causation: Hill vs. Fisher

- Whether smoking caused lung cancer used to be controversial
- R.A. Fisher (architect of the randomized controlled trial), argued that a "confounding factor" – a lung cancer gene – explained the apparent effect of smoking
- A RCT was impossible, but a 1950 observational study by Richard Doll and Austin Bradford Hill showed a clear association between smoking and cancer
- Surgeon General convened a committee to develop criteria for assessing causation in observational studies
  - These criteria evolved to Hill's 9 Criteria, widely used in epidemiology

Hill, Austin Bradford 1965: Proceedings Royal Soc of Med **58** (5): 295–300

Updated by Fadek, E et al: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4589117/

Reflections of Hill's 9 Criteria in Improvement Science

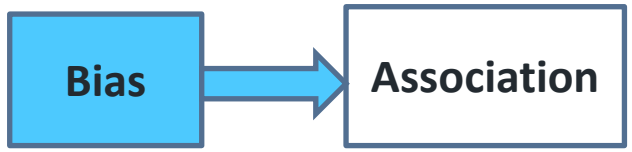
- **Plausibility** (credible causal theory expressed in a driver diagram)
- **Strength** (magnitude of change and "special cause")
- **Consistency** (changes produce results when implemented in different settings and contexts)
- **Temporality** (annotated run charts and statistical process control charts)
- **Biological gradient** (relation of dose of an implementation activity delivered, dose received, and magnitude of observed effect)
- **Experiment** (PDSAs along hypothesized causal pathway)
- **Analogy** (If Michigan Keystone CLABSI collaborative worked, maybe this approach will work for SSIs in Michigan, or for CLABSIs in England – spoiler alert: it did not!)
- **Specificity** (comparison groups in some QI studies to rule out secular trends *unrelated* to the intervention ("counterfactual")

Practice using these by applying them to Pronovost's famous Michigan CLABSI Collaborative: Pronovost P, et al., An intervention to decrease catheter-related bloodstream infections In the ICU. N Engl J Med 2006;355:2725-32

So What's a "Counterfactual" – And Why is it Important in QI Design and Evaluation?

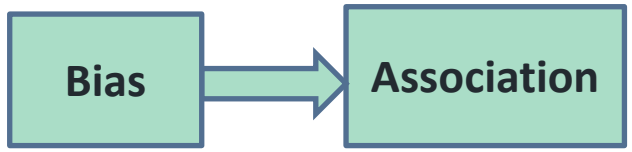
- A comparison between what actually happened and what would have happened in the absence of the intervention
  - "What if" statements:
    - What if I had not intervened?
    - What if I had not taken aspirin for my headache?
    - What if I had not smoked?
- My personal problem with "bright spotting."
  - Were there "bright spots" where you did not intervene, not just where you did.
  - Were there comparable numbers of "dark spots" in intervention and non-intervention sites?
  - What biases may have distorted bright spot results?

Bias



Systematic error, or deviation from the “truth,” introduced during design, subject selection, project implementation, data collection, or analysis

Do You Have a Bias Checklist for your Project?



- Systematic error, or deviation from the "truth," introduced during design, subject selection, project implementation, data collection, or analysis
  - Selection bias (includes volunteer and "enthusiast" bias in QI)
  - Performance bias ("trying harder" when blinding is not possible)
  - Detection/ascertainment bias (e.g., interview bias, looking for "positive"data)
  - Reporting/publication bias (just the good news, please, preferably with a small p-value)
  - Attrition bias (people drop out) (not fully mitigated by "intention to treat" analysis)
  - Protopathic bias (disease already underway) (mitigated by inserting a lag time before the outcome)
  - Indication bias (existing risk factor or condition influences both a decision to treat and the outcome of interest) (mitigated by propensity scoring)
  - Misclassification bias
  - Lead time bias (early screening picks up a condition before it would be manifest clinically)(distort incidence estimates and outcomes)
  - Immortal time bias (subjects could not have died during study timeframe)

Most Published Studies of QI Collaboratives (QICs) are Low Quality and Biased

Conclusions of a systematic review of published collaboratives:

Overall, the QICs included in this review reported significant improvements in targeted clinical processes and patient outcomes. These reports are encouraging, but most be interpreted cautiously since fewer than a third met established quality and reporting criteria (SQUIRE 2.0), and publication bias is likely.*

Are quality improvement collaboratives (QICs) effective? A systematic review. Wells S, et al. BMJ Qual Saf. doi: 10.1136/bmjqs-2017-006926.

* http://squire-statement.org/index.cfm?fuseaction=Page.ViewPage&PageID=471
Cochrane Handbook for Systematic Reviews of Interventions:
http://handbook.cochrane.org/

An Example of Failed Efforts to Scale-Up “Bright Spots” or Promising Results from Small or Potentially Biased Studies

Example: Reduce Mortality in Emergency Laparotomy Collaborative (ELC)

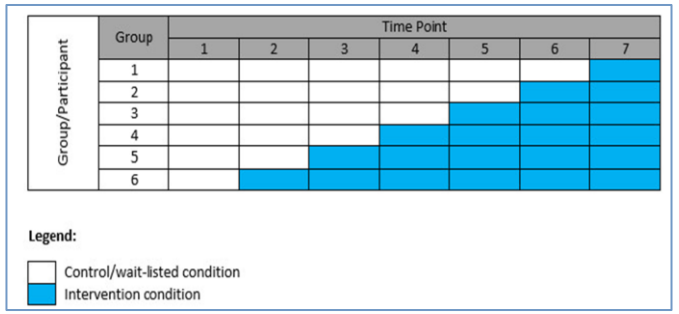
- 26-trust Breakthrough Series Collaborative drew on promising results from a 4-trust pilot study
- 6-component "bundle"
  - Early warning score or lactate measurement
  - Early identification of sepsis and prompt antibiotics
  - Prompt transfer to operating theatre
  - Consultant surgeon and anesthesiologist in operating theatre
  - Goal-directed fluid therapy
  - Post-operative care in ICU
- Primary outcomes: in-hospital mortality and risk-adjusted length of stay.

ELC Collaborative Results

- Unadjusted mortality decreased from 9.8% to 8.3% (15.3% reduction) pre-post (risk adjusted mortality from 5.3% to 4.5%) (15.1% reduction)
- Length of stay declined from 20.1 days to 18.9 days
- “Significant” improvement in 5/6 measures, but timely antibiotic administration for sepsis declined
- “A collaborative approach using a quality improvement methodology and care bundle appeared to be effective in reducing mortality….suggesting hospitals should adopt such an approach to see better outcomes…”

NHS QI Programme to Improve Survival After Emergency Laparotomy (EPOCH)

- Stepped-wedge, cluster randomized controlled trial (cRCT)
  - 15 geographical clusters of trusts, 16 “steps” of 5 weeks each (longest exposure to intervention 80 weeks)
- 36-component intervention based on Delphi consensus process (*not* a “bundle”)
- 10 components selected for emphasis (some not strongly evidence-based)
  - Pre-op documentation of risk
  - Time to operating theatre
  - Time to operating theater by level of urgency
  - Goal directed fluid therapy
  - Serum lactate measured at end of surgery
  - Admission to intensive care after surgery
  - 4 measures based on consultant present/participating
  - (no surgical safety checklist at that time)

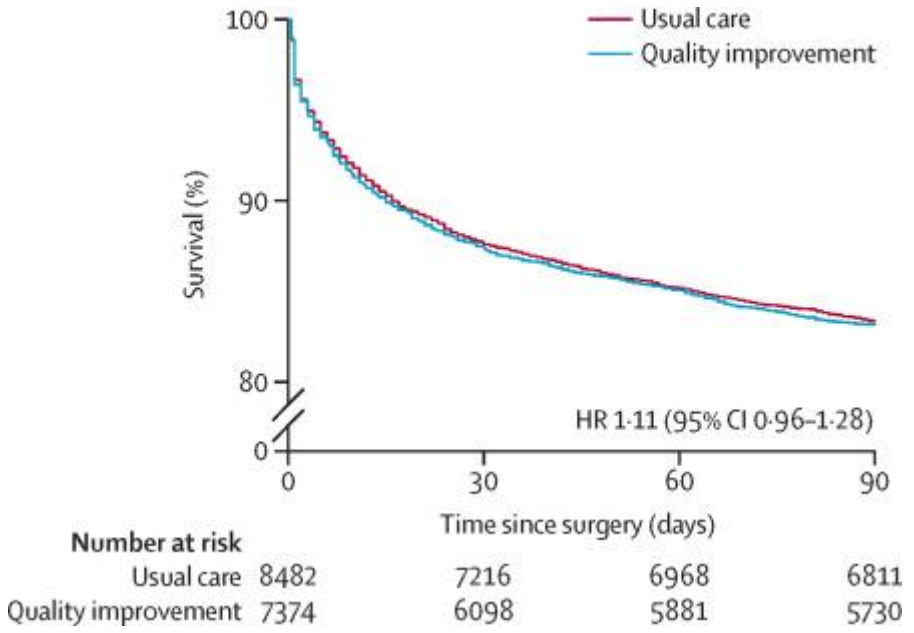


Peden CJ, et al. Lancet 2019;393:213-21

Important Note: EPOCH was designed before the ELC Collaborative was finished

EPOCH Results

- 16% 90-day mortality reduction in both intervention and control groups in intention-to-treat analysis
  - No difference in 180-day mortality
- Length of stay not meaningfully or statistically different
- Readmission rates within 180 days similar
- Some improvement (generally modest) in 7/10 key measures

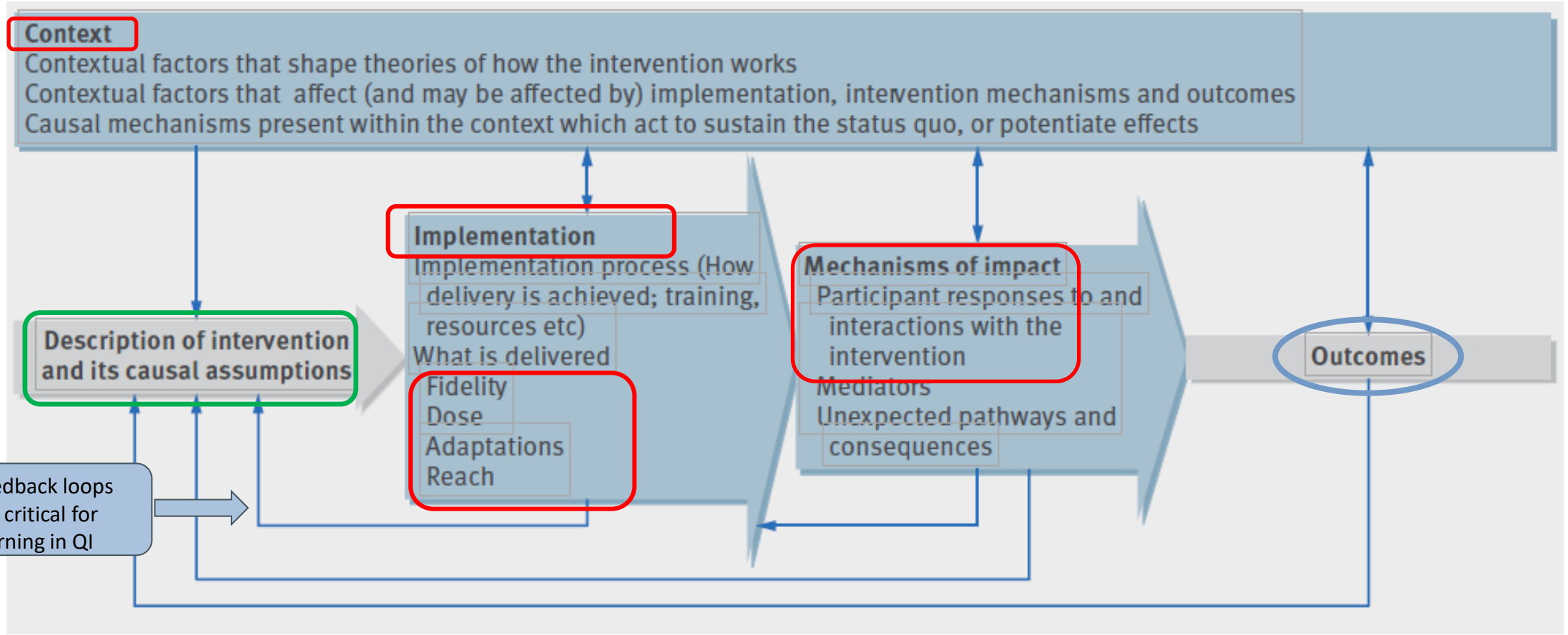


Process Evaluation of EPOCH

- Kudos for the research team on doing a terrific process evaluation
- What IS a “process evaluation?”
  - Determines whether program activities have been implemented as intended and have (or have not) resulted in the predicted outputs and outcomes
  - Can be used to assess the quality of implemented activities, clarify what worked or did not work, and determine why and how the results were as hoped and predicted (or not)
    - Emphasizes qualitative inquiry about contextual barriers and enablers for success that were encountered during the project
  - Main features are adapted from implementation science principles and are compatible with evaluation frameworks, such as RE-AIM (more about this later)

Stephens TJ, et al. Implementation Science 2018;13:142-158

Key Elements and Interactions of a Process Evaluation



Note the need for a clear description of the intervention and implementation activities, as well as the context and a clear causal (program/change) theory, and consideration of context

Process evaluation of complex interventions: Medical Research Council guidance
Moore G, et al. BMJ. 2015 Mar 19;350:h1258. doi: 10.1136/bmj.h1258

Process Evaluation of EPOCH

- Many challenges reported, including time (especially for collecting and entering data) and resources.
- Large variation in fidelity to recommended QI methods
  - 4/74 teams used PDSAs “often”
- Large variation in which of the 36 components teams chose implement
- Timeframe for implementation as short as 5 weeks
  - But no evidence that outcomes were better in hospitals with up to 80 weeks to implement changes
- In fairness, the PI has pointed out that there was substantial improvement in some of the processes of care

Critical Question: If QI Training is Critical, How Much is Enough?

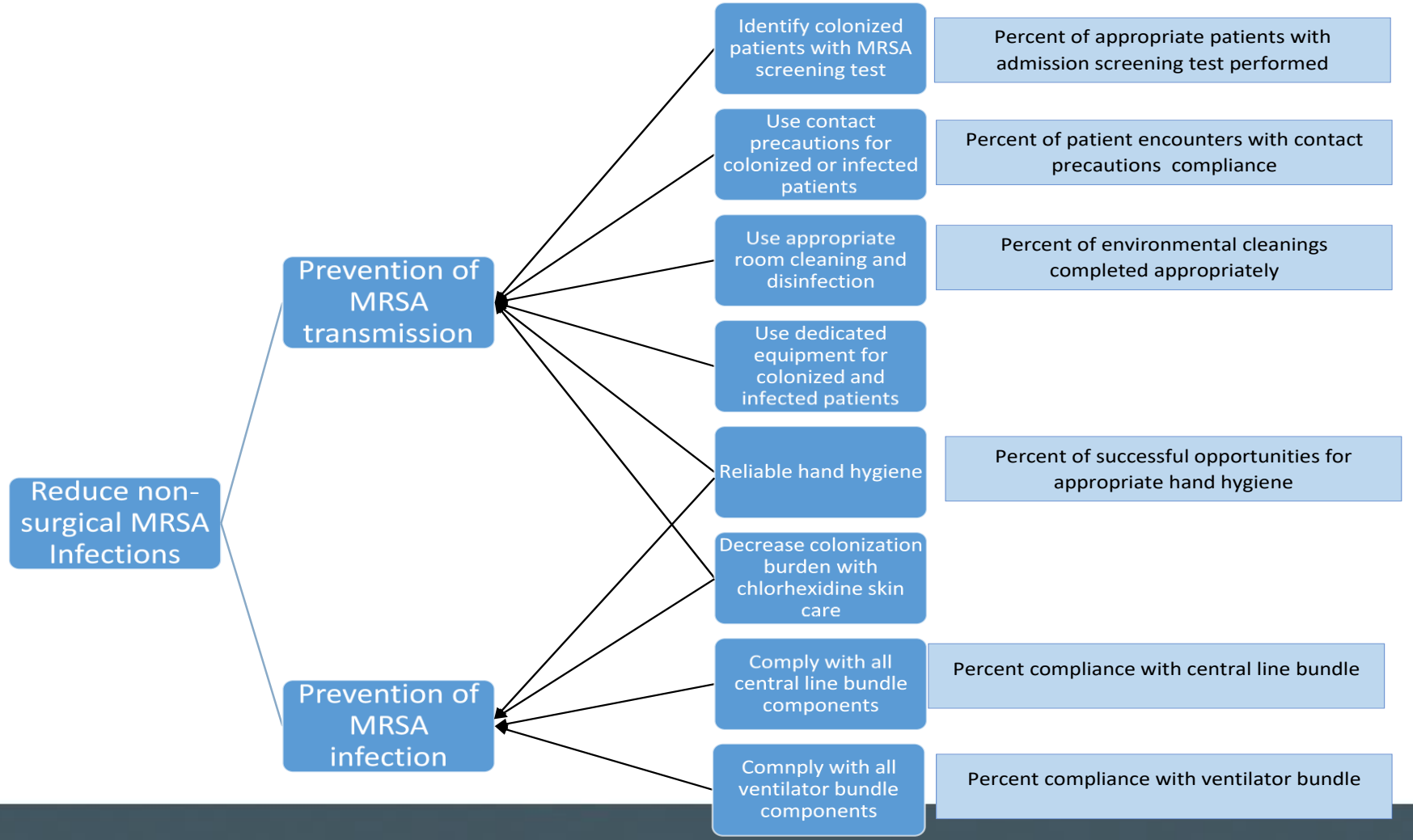
- EPOCH QI "dose" was described as "light touch" –
  - QI support *much* less intense than in the ELC breakthrough series collaborative

Causal (program) theory and how to display it (driver diagrams, logic models)

Reminder: Rigor in Design and Evaluation Requires a Strong Causal Theory

- To reiterate, it is essential to have a strong "causal theory" (aka "program theory")
- Logic models, driver diagrams, other frameworks are useful for articulating and displaying the theory. They:
  - Clarify the theory and inform strategy for achieving outcomes
  - Ensure that designers, implementers, and evaluators are on the same page
  - Provide a framework for measurement
  - Inform evaluation
  - Usually are required for competitive grants and contracts
  - Allow other organizations or researchers to compare their project/study design to what others have used

Example: MRSA Control Driver Diagram

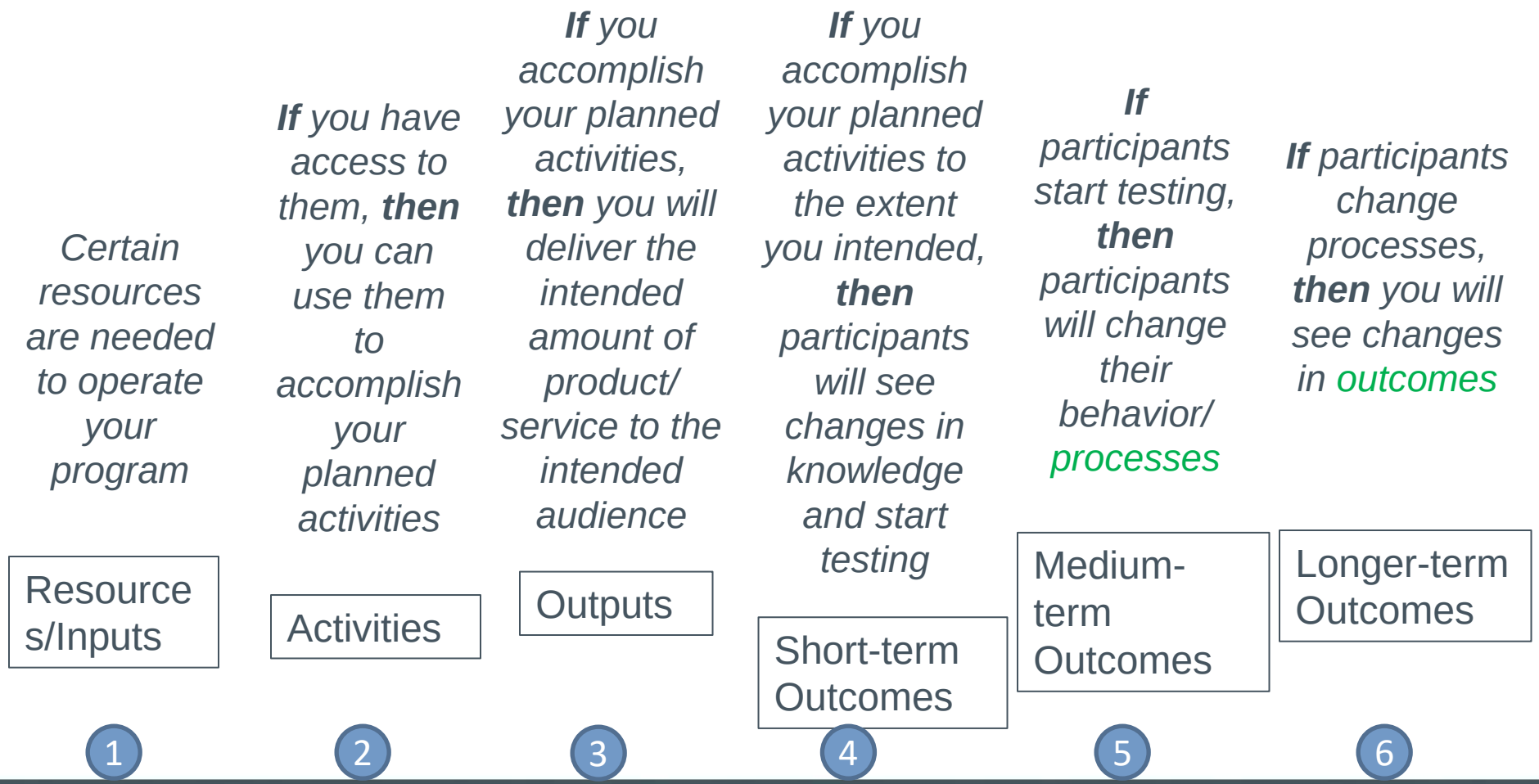


Logic Models

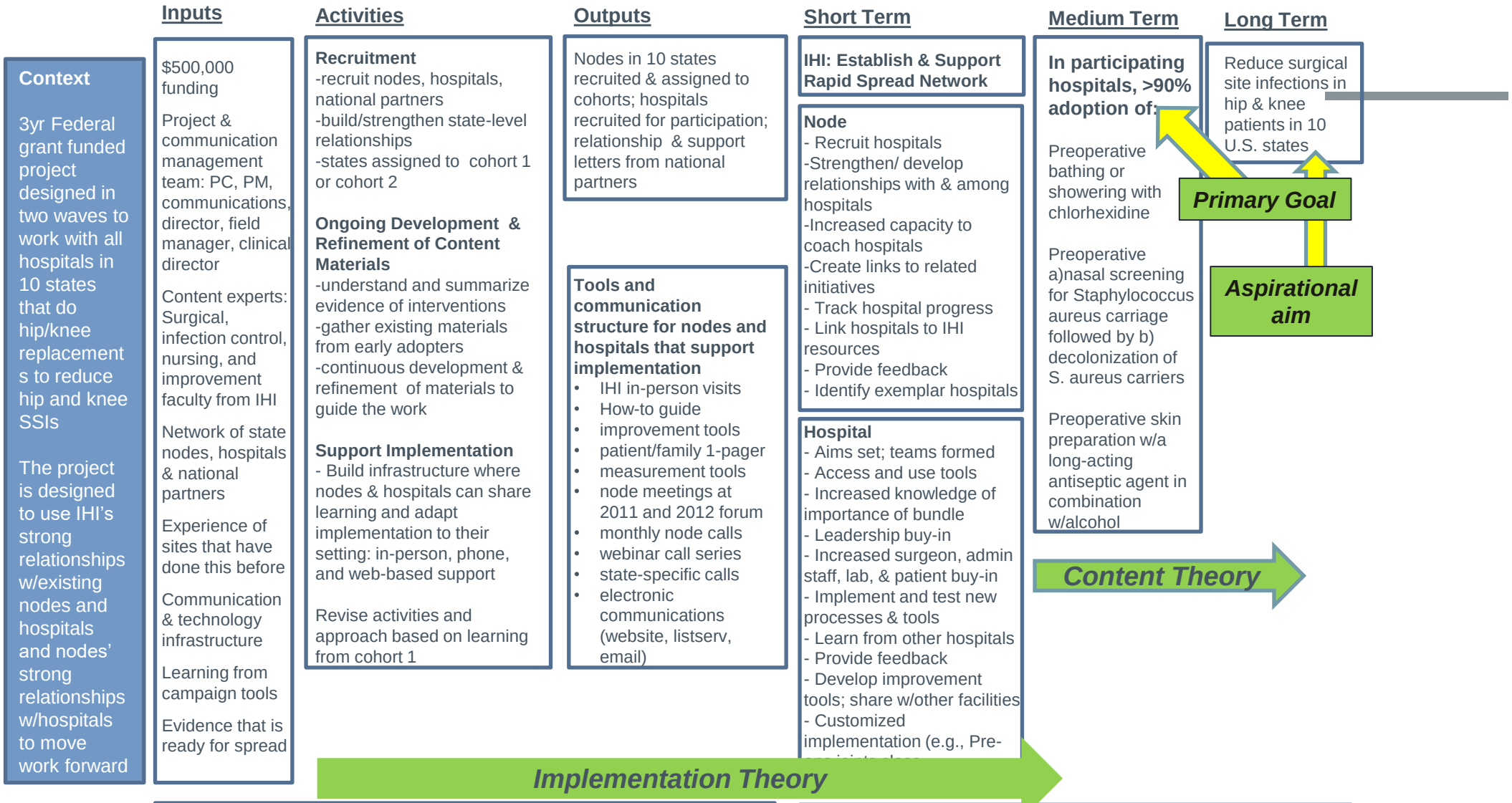
- Show a clear delineation of what will determine a given outcome, including
  - Inputs, resources, assumptions, and context
  - Activities and their predicted outputs
  - Key processes that need to be implemented to produce desired outcomes (so-called "content theory")
  - Embody both *implementation* theory and *content theory* (often expressed in a driver diagram)
- Enable evaluators to see exactly what was done and how effectively it was done, including fidelity, the intended "dose," and the received "dose"
- Support replication elsewhere

Note: Pierre will discuss logic models and driver diagrams again when he talks about designing QI and evaluating projects later

The Basic Logic Model: "If-Then"



Use of IHI's Rapid Spread Network to Reduce Hip & Knee Surgical Site Infections in Ten States in the U.S. (Sept 2010 – Oct 2012)



A Brief Guide to Evaluation Frameworks

The “How” and “Why” of Implementation: Making Sense of Diverse Frameworks and Terms

Some or all of the frameworks we will present are relatively unfamiliar to many quality improvers (they were to me for many years)

The terminology can be jargony, idiosyncratic, and frankly, confusing

I will try to reconcile the methods and terminology of these frameworks and provide guidance for choosing key elements for evaluating the “how” and “why”

Regardless of the evaluation framework, evaluation and design must be aligned – Collaboration is essential

Reminder about Process Evaluation

- Determines whether program activities have been implemented as intended and have (or have not) resulted in the predicted outputs and outcomes
- Can be used to assess the quality of implementation activities, clarify what worked or did not work, and determine why and how the results were as we predicted/hoped they would be (or not)
  - Emphasizes qualitative inquiry about contextual barriers and enablers for success that were encountered during the project

Process evaluation is the foundation for most evaluation and related measurement frameworks

Frameworks, Frameworks, Frameworks

Five (Selected) Frameworks for Evaluating the "How" and "Why" of Implementation and Outcomes

Self-study

1. Kirkpatrick Framework
2. Realist Evaluation
3. RE-AIM
4. Implementation Science Outcomes
5. Consolidated Framework for Implementation Research (CFIR)
   - A framework for assessing **context** in terms of existing or potential **barriers** and **facilitators** to successful implementation

A General Framework that is Foundational for all 5 Evaluation Frameworks, as well as process Evaluation

Everett Rogers' Adoption of Innovations Theory

- Characteristics of innovations that influence decisions **whether or not to adopt** an innovation

Self-study

Kirkpatrick Model

- Developed in 1950s based on organizational and industrial psychology
- Widely used to evaluate training programs

1	Experience/Reaction	What was the participants' experience? Did the participants have an excellent experience working on the improvement project?
2	Learning	What did participants learn? Did they learn improvement methods and begin testing?
3	Process/Behavior	Did participants modify their behavior? Did they work differently and see change in their process measures?
4	Outcomes	Did the organization improve its performance and see improvements in its outcome measures?

For education and training, take a look at Millers' Pyramid, which emphasizes competency, not just learning

Kirkpatrick Partners. The Kirkpatrick Model. Kirkpatrick Partners, LLC. https://www.kirkpatrickpartners.com/the-kirkpatrick-model/

Miller's Pyramid: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246123/

RE-AIM – Widely Used in Public Health: 5 Domains

- Reach (or penetration)
  - WHO is intended to benefit and who actually participates or is exposed to the initiative?
- Effectiveness
  - WHAT is the most important outcome you are trying to achieve and what are possible negative outcomes?
- Adoption
  - WHERE was the program adopted and WHO adopted it?
- Implementation
  - HOW was the intervention implemented? Was it implemented consistently and reliably (WHY or WHY NOT); What adaptions were made and WHY?
- Maintenance
  - WHEN was the program operational, and how long are the results sustained?

Gaglio B, et al. The RE-AIM Framework: A Systematic Review of Use Over Time
Am J Public Health. 2013;103(6):e38-e46. doi:10.2105/AJPH.2013.301299.

Proctor's Implementation Outcomes

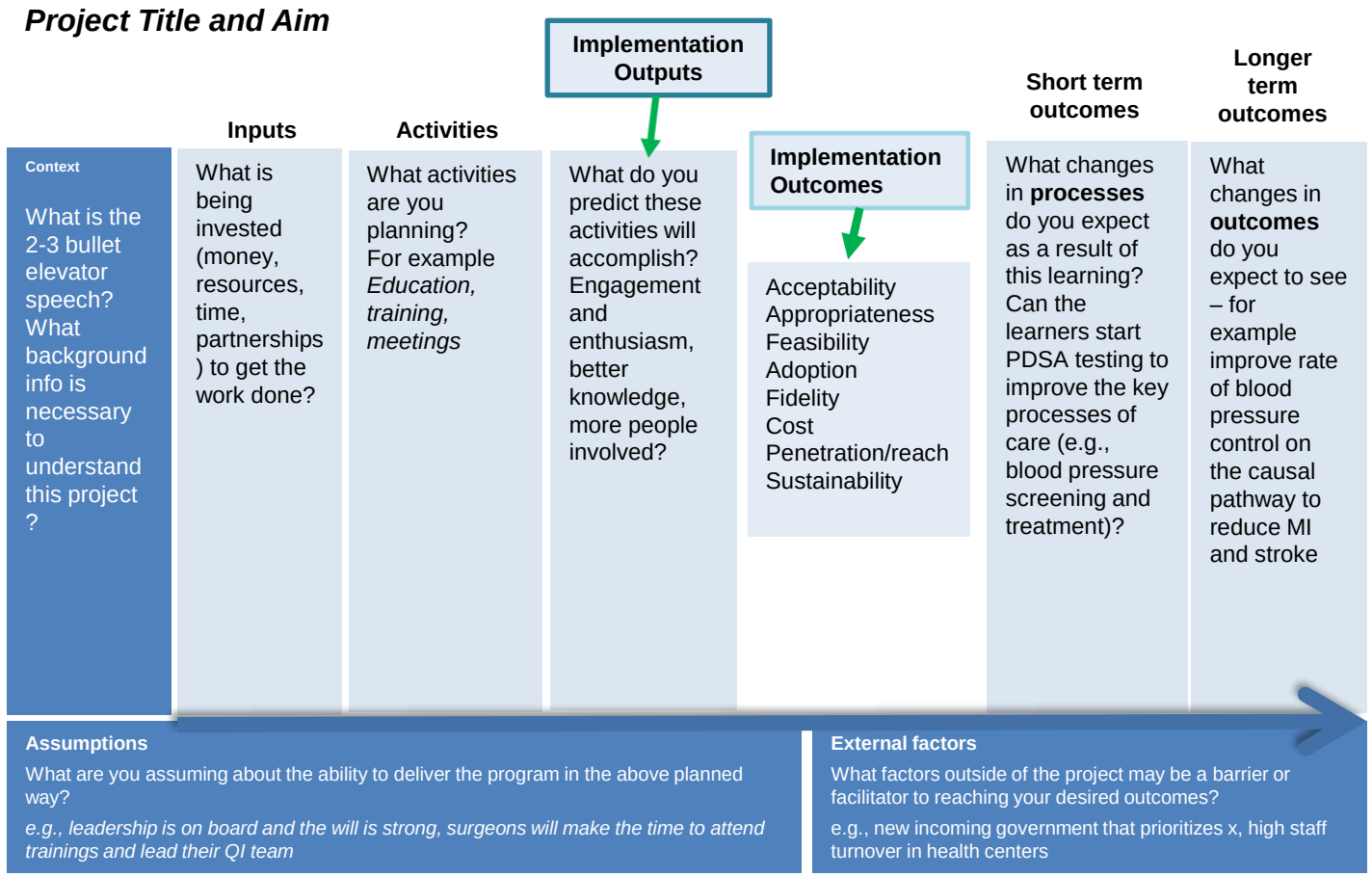
- *Acceptability:* Perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory (Rogers' relative advantage and complexity
- *Appropriateness:* Perceived fit, relevance, or compatibility of an innovation or evidence-based practice for a given practice setting, provider, or consumer (Rogers' compatibility)
- *Feasibility:* Extent to which a new treatment, or innovation, can be successfully used or carried out within a given agency or setting (Rogers' trialability)
- *Adoption:* Intention, initial decision, or action to try or employ an innovation or evidence-based practice (RE-AM adoption)

Proctor E, et al. Adm Policy Ment Health. 2011;38(2):65-76. doi:10.1007/s10488-010-0319-7
https://implementationoutcomerepository.org/implementation-outcomes

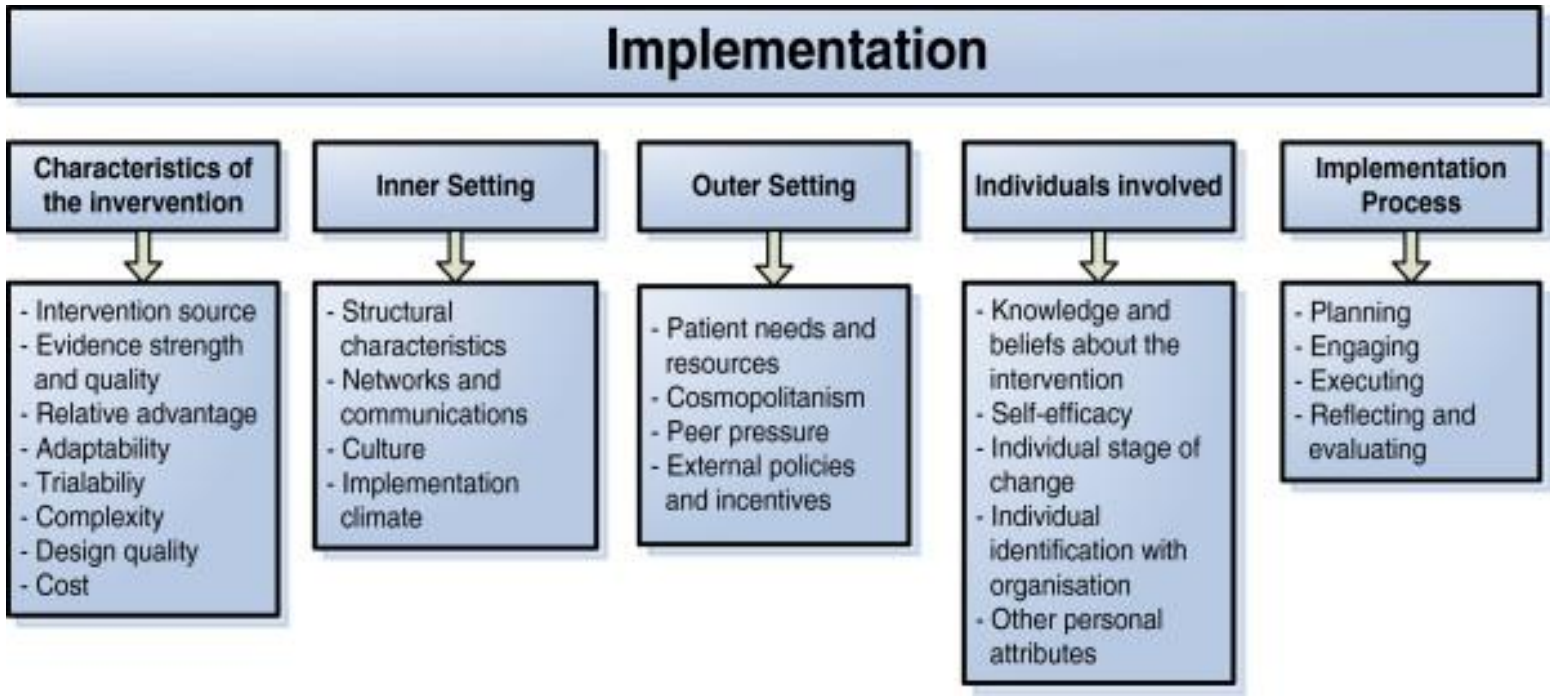
Proctor's Implementation Outcomes

- *Fidelity:* Degree to which an intervention was implemented as prescribed/intended in the original protocol/design (part of "implementation" in RE-AIM) [reliability]
- *Implementation cost*
- *Penetration:* Degree to which an intervention is implemented by the target audience (RE-AIM reach)
- *Sustainability:* Extent to which a newly implemented treatment is maintained or institutionalized within a service setting's ongoing, stable operations (RE-AIM maintenance)

Implementation Science Logic Model Template



Consolidated Framework for Implementation Research (CFIR): Domains and Sub-Domains



Should be considered in both design and evaluation

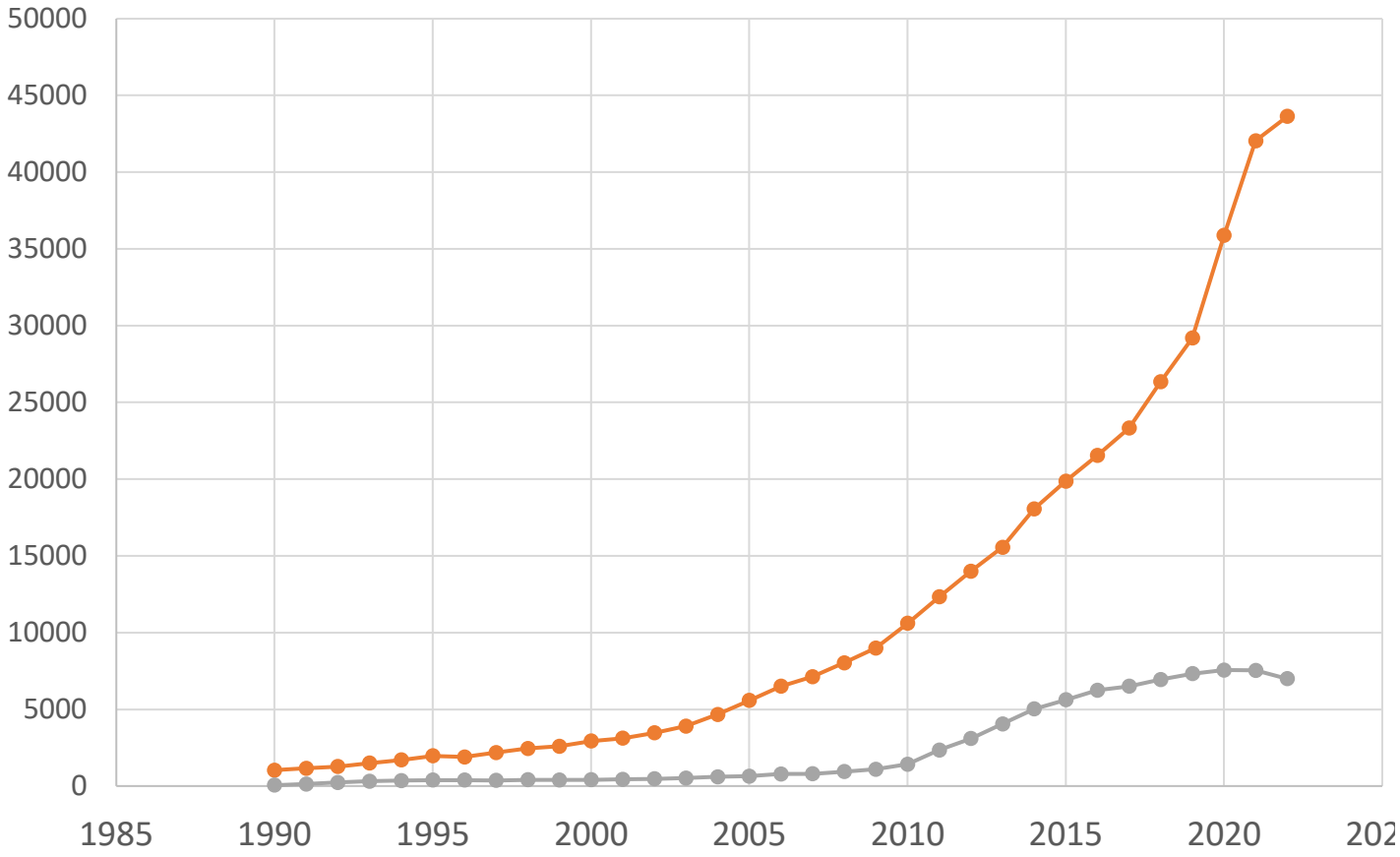
Damschroder LJ, et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50. doi:10.1186/1748-5908-4-50

“Right-Sizing” Evaluation

- Strong program designs and rigorous evaluation tend to be resource-intensive and take too long
  - Timeliness of evaluation is more and more challenging given the pace of innovation, especially in IT and tech
  - Great ideas “can’t wait”
- Rapid learning is easier if the designers and evaluators collaborate and evaluators are embedded in program implementation
- Formal evaluation is not necessary nor practical for small projects along well-travelled roads
  - For example, implementation of the central line associated blood stream infection bundle

“QI/Improvement” vs “Implementation” Publications since 1990

Number of QI Publications per Year



Implementation Science Studies – study of methods to adopt implementation practice

QI/Quality Improvement Studies – study of a theories and ideas to improve implementation practice

Key Question: How can we Improve Deployment of ~~QI~~ Science?

Does quality improvement improve quality?

Authors: Mary Dixon-Woods[A] and Graham P Martin[B]

ABSTRACT

Although quality improvement (QI) is frequently advocated as a way of addressing the problems with healthcare, evidence of its effectiveness has remained very mixed. The reasons for this are

US studies suggest that nurses deal v
8.4 work system failures per 8-hour
continually interrupted.[5,6] The need
re-learn, associated with the variabil

4 Issues Identified

1. **Does it work?** Variable fidelity in the application of QI methods.

2. **Can we Scale promising QI work** – pilot projects not designed for scale. Failed efforts to scale/integrate into larger system

3. **Can we effectively learn from our work?** Lack of rigorous evaluation (studies subject to bias)

4. **Can we share what we are learning?**: Lack of sharing of successes and failures.

3 Ideas for Improving our Use Of QI Science

1. Establish your Project Theory: Core Components of Design
2. Use a Framework for Improvement Research and Evaluation (FIRE)
3. The “Holy Grail”: Use a Study Design that supports the causal pathway

3 Ideas for Improving our Use Of QI Science

1. Establish your Project Theory: Core Components of Design
2. Use a Framework for Improvement Research and Evaluation (FIRE)
3. The “Holy Grail”: Use a Study Design that supports the causal pathway

Establishing the Project/Program Theory: 5 Core Components of Design:

1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Execution Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*

Langley, J. et al. The Improvement Guide. Jossey-Bass Publishers, 2009.

Core Design: 1. Goals and Aims

Learning Questions +

1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Execution Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*

Aim Statement:
Numeric, time-bound Aim, that calls out equity

"How much, by when, for whom"?

e.g. Decrease post operative mortality and morbidity by 25% across 250 African hospitals over 5 years

Core Design: 2. Content Theory

1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What theoretical drivers and change ideas will drive us to our Aim?*

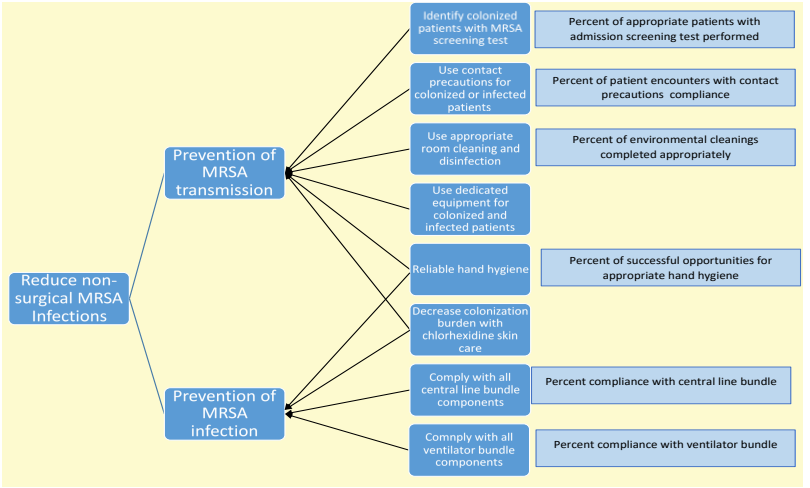
3) **Execution Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*

Role of Content Theory in establishing the Causal pathway

*If we create the capabilities and environment that supports change (**Primary Environmental Drivers**)* ⇦ *Then providers will reliably delivery evidence-based care (**Primary Clinical Drivers**)* ⇦ *Then the clinical **processes and outcomes will improve***



The Driver Diagram

Core Design: 3. Implementation Theory

Learning Questions +

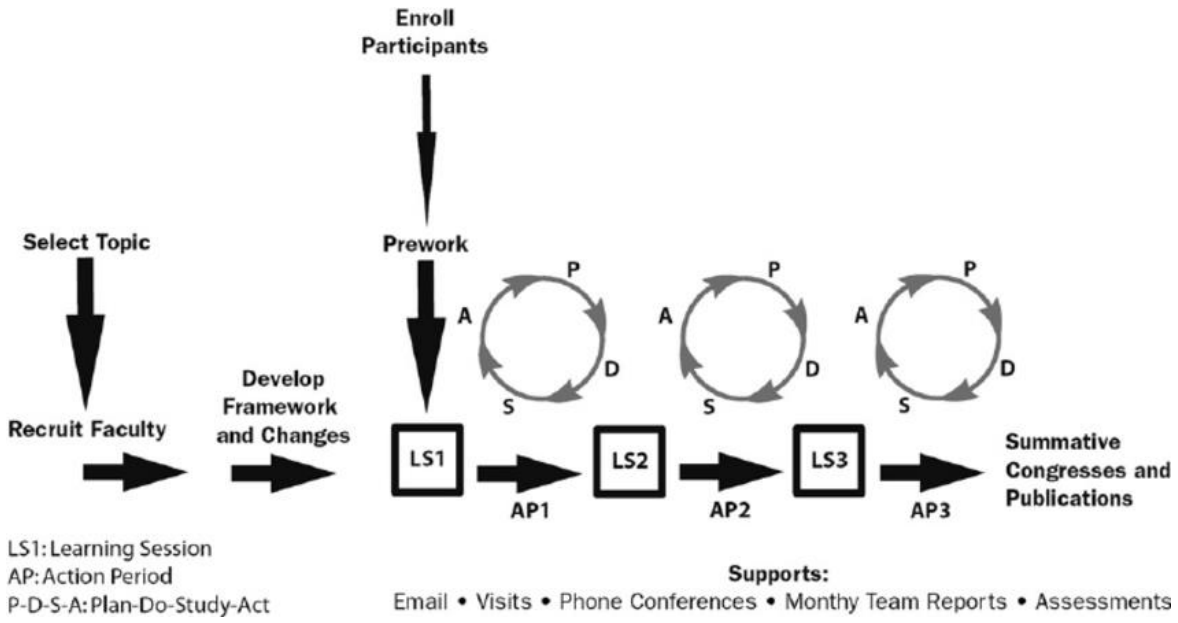
1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Implementation Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*



Core Design: 3. Implementation Theory

Learning Questions +

1) **Goals:** *What are we trying to Accomplish*

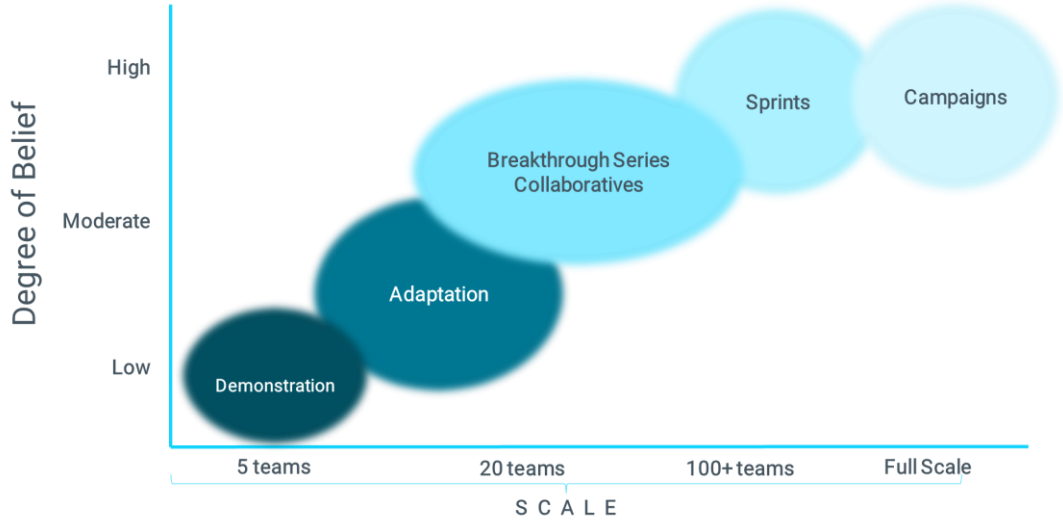
2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Implementation Theory:** *What Activities will test and implement successful ideas?*

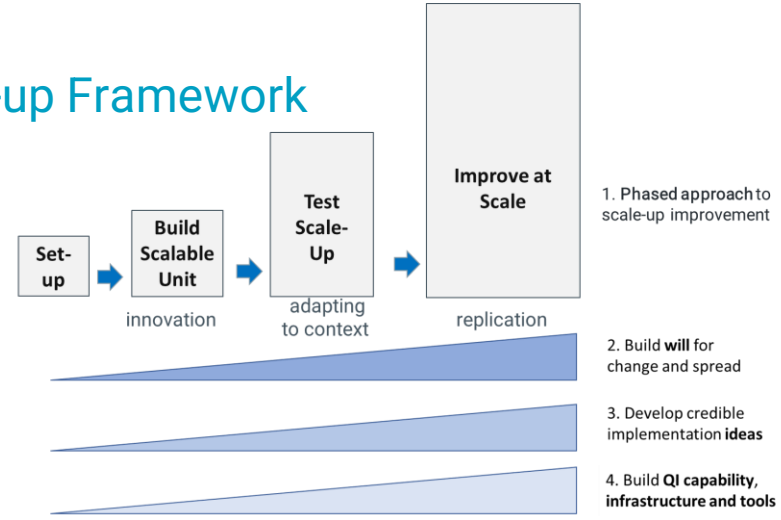
4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*

IHI's Implementation Designs



IHI Scale-up Framework



Core Design: 4. Measures and Learning

Learning Questions +

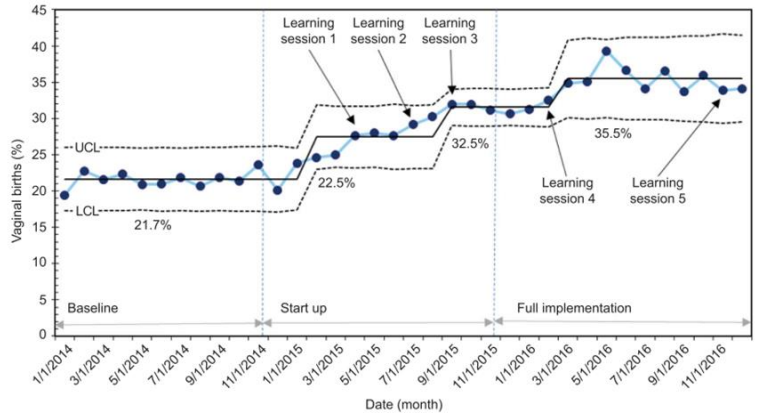
1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Execution Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) **Sharing & Communication:** *How and to whom will we communicate our results and learning?*



- **Data collection plan** for effectiveness and learning (quantitative and qualitative data)
- **Data plan linked to Study Design**
  - Quasi-experimental, mixed methods
  - Randomization designs

Core Design: Disseminating Learning

Learning Questions +

1) **Goals:** *What are we trying to Accomplish*

2) **Content Theory:** *What drivers, change ideas will drive us to our Aim?*

3) **Execution Theory:** *What Activities will test and implement successful ideas?*

4) **Results & Learning:** *What measures and learning will guide us to success*

5) Sharing & Communication: *How and to whom will we communicate our results and learning?*

- Crucial role of data in building will (for spread and scale)
- Power of stories *and* data
- Importance of disseminating learnings (successes *and* failures)

Langley, J. et al. The Improvement Guide. Jossey-Bass Publishers, 2009.

Logic Model: Assembling Core Components of Design into a Project Theory

Core Design Components

1) **Goals**

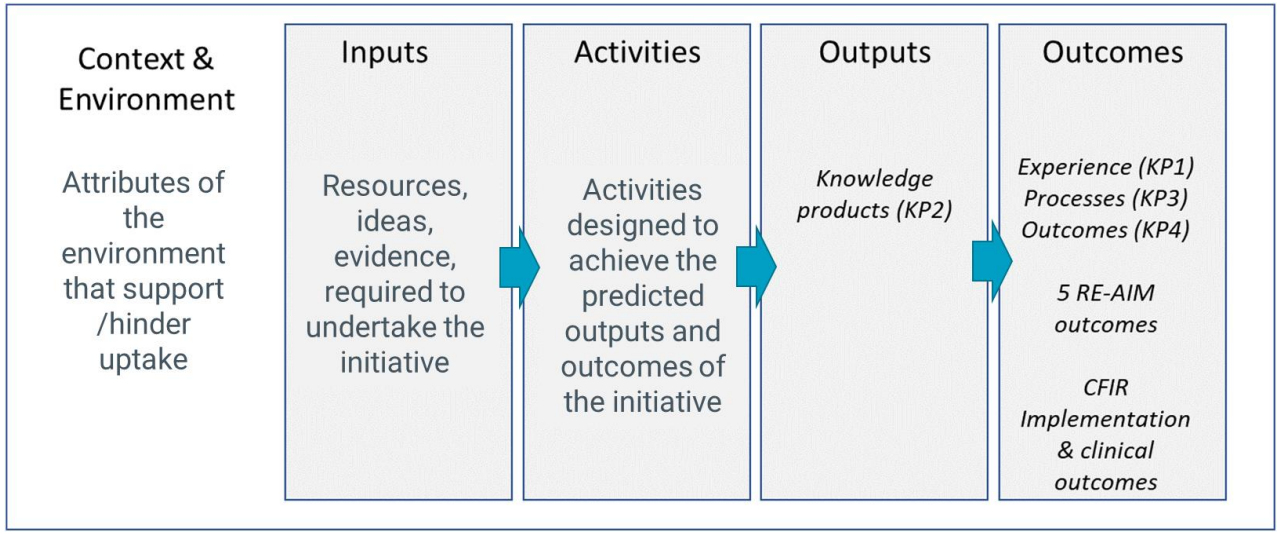
2) **Content Theory:**

3) **Execution Theory:**

4) **Results & Learning:**

5) **Sharing & Communication:**

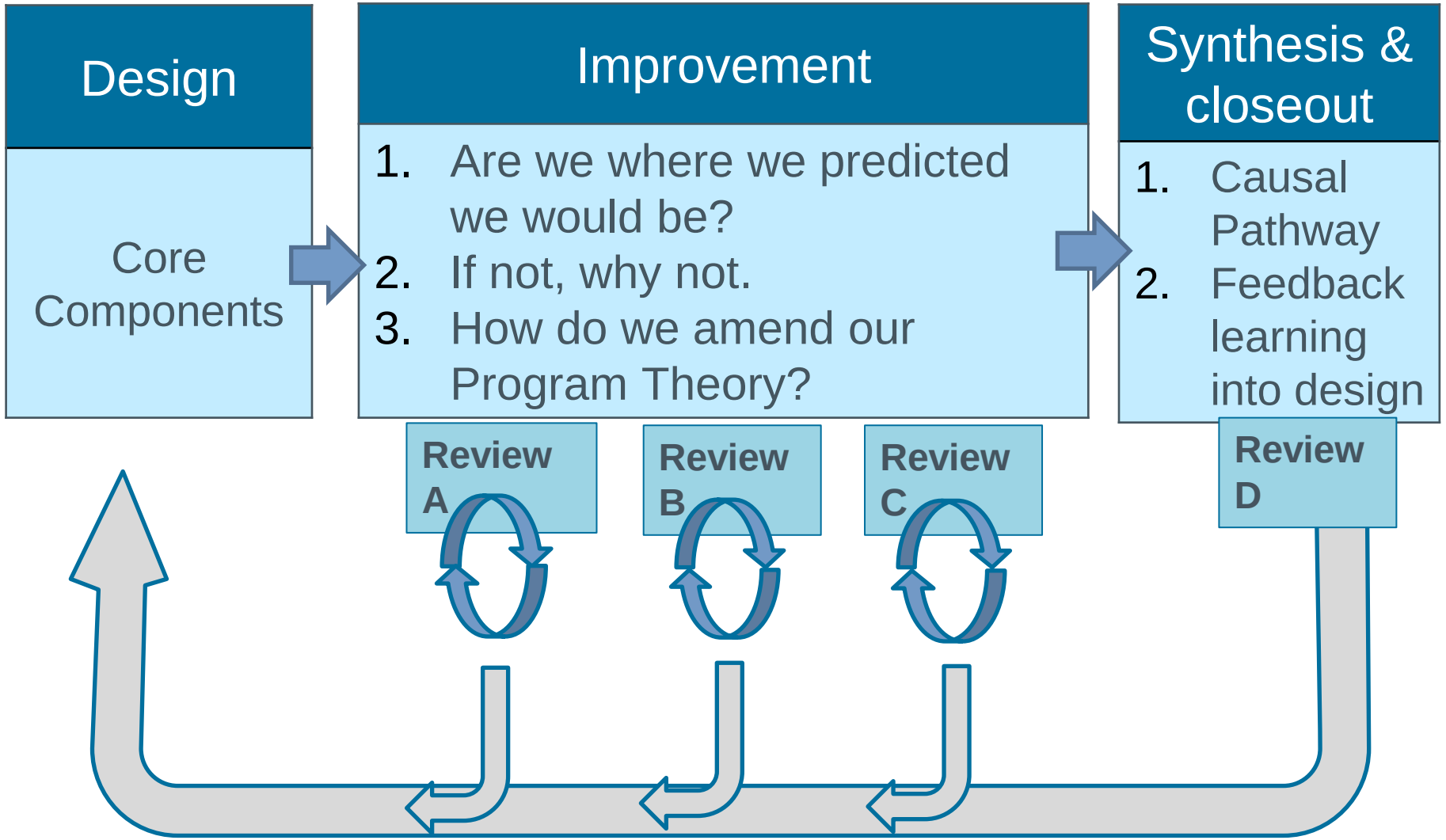
Logic Model



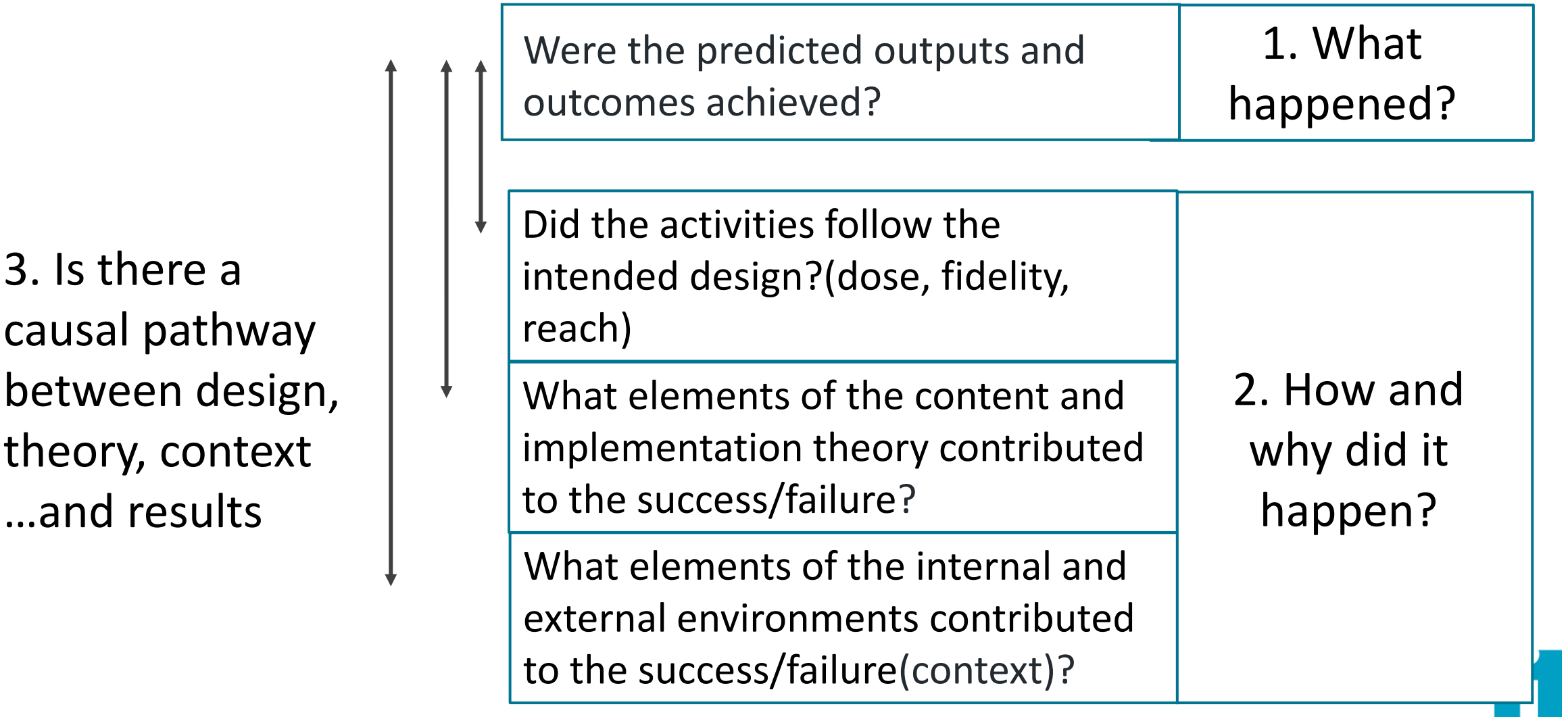
3 Ideas to Improve Deployment of QI Science

1. Establish your Project Theory: Core Components of Design
2. Use a Framework for Improvement Research and Evaluation (FIRE)
3. The “Holy Grail”: Use study designs that support the causal pathway

Evaluation is a continuous Learning Process

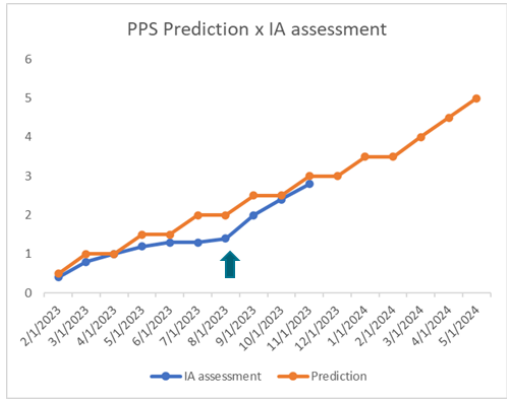


3 Key Questions: Framework for Improvement Research and Evaluation (FIRE)



1. What Happened? Were the predicted outputs/outcomes achieved?

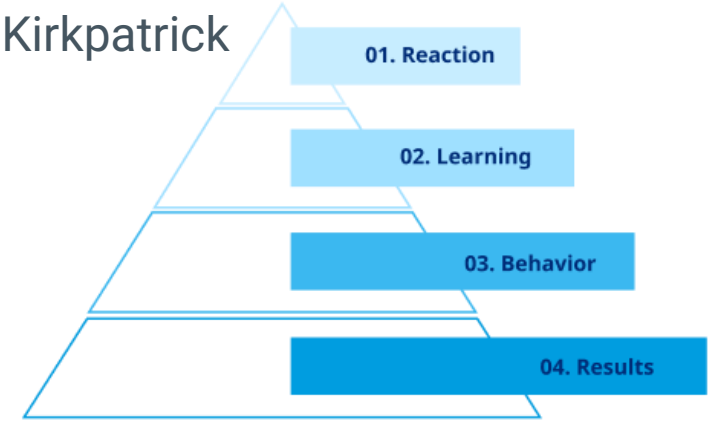
IHI Project Progress Score (0-5): Predict progress over time, allows for adaptive design to ensure progress to goal



3. Is there a causal pathway between design, theory, context …and results

Were the predicted outputs and outcomes achieved?	1. What happened?
Did the activities follow the intended design?(dose, fidelity, reach)	2. How and why did it happen?
What elements of the content and implementation theory contributed to the success/failure?	
What elements of the internal and external environments contributed to the success/failure(context)?	

Kirkpatrick



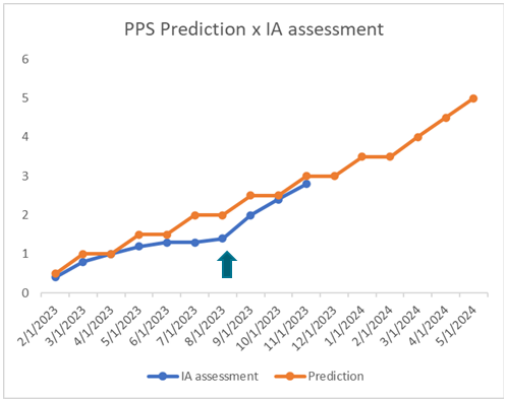
Ref: Assessment Scale for Collaboratives: https://www.ihi.org/resources/Pages/Tools/AssessmentScaleforCollaboratives.aspx

1. What Happened? Were the predicted outputs/outcomes achieved?

Were the predicted outputs and outcomes achieved?	1. What happened?
Did the activities follow the intended design?(dose, fidelity, reach)	2. How and why did it happen?
What elements of the content and implementation theory contributed to the success/failure?	
What elements of the internal and external environments contributed to the success/failure(context)?	

3. Is there a causal pathway between design, theory, context …and results

IHI Project Progress Score (0-5): Predict progress over time, allows for adaptive design to ensure progress to goal



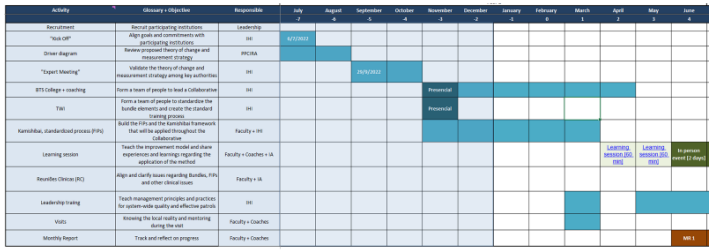
RE-AIM

1. Did the activities follow the intended Design?

3. Is there a causal pathway between design, theory, context …and results

Were the predicted outputs and outcomes achieved?	1. What happened?
Did the activities follow the intended design?(dose, fidelity, reach)	2. How and why did it happen?
What elements of the content and implementation theory contributed to the success/failure?	
What elements of the internal and external environments contributed to the success/failure(context)?	

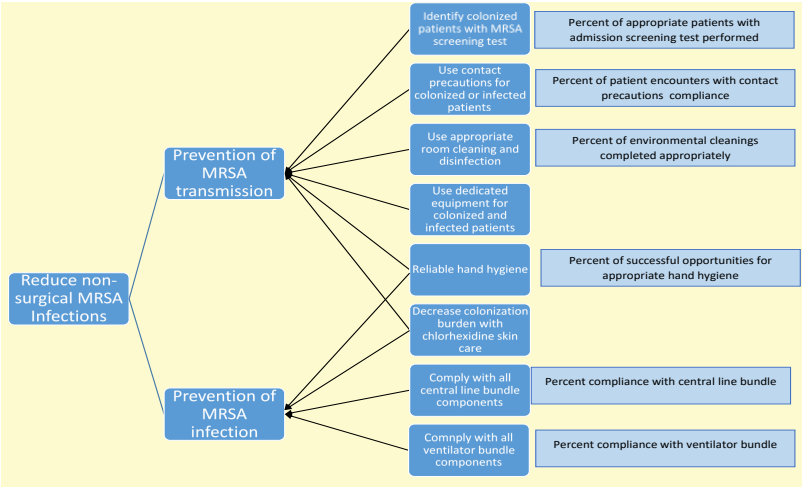
Gantt Chart

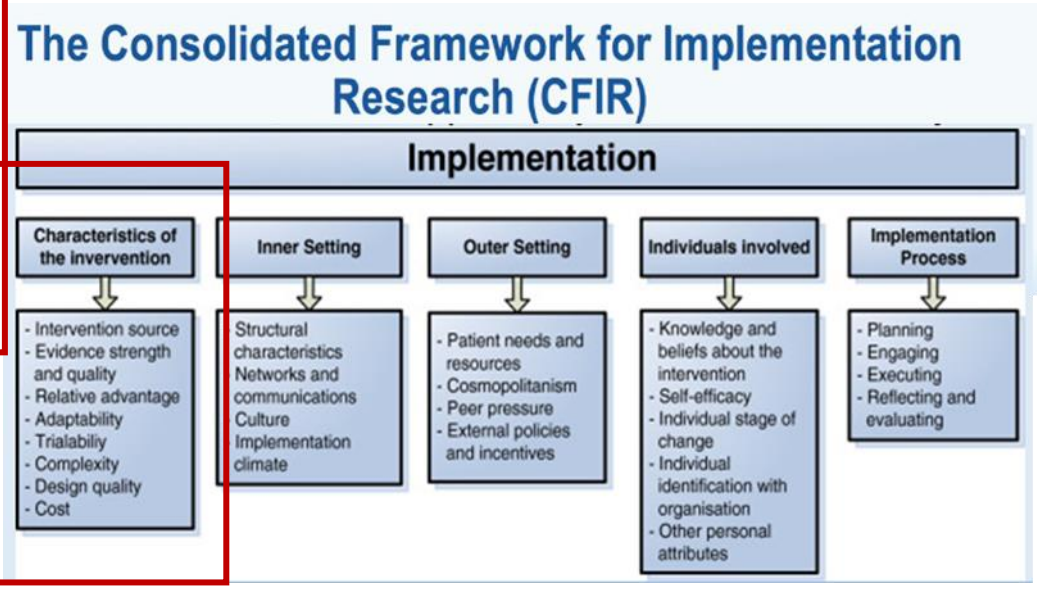


2. How did content and implementation theory affect result?

3. Is there a causal pathway between design, theory, context …and results

Were the predicted outputs and outcomes achieved?	1. What happened?
Did the activities follow the intended design?(dose, fidelity, reach)	2. How and why did it happen?
What elements of the content and implementation theory contributed to the success/failure?	
What elements of the internal and external environments contributed to the success/failure(context)?	

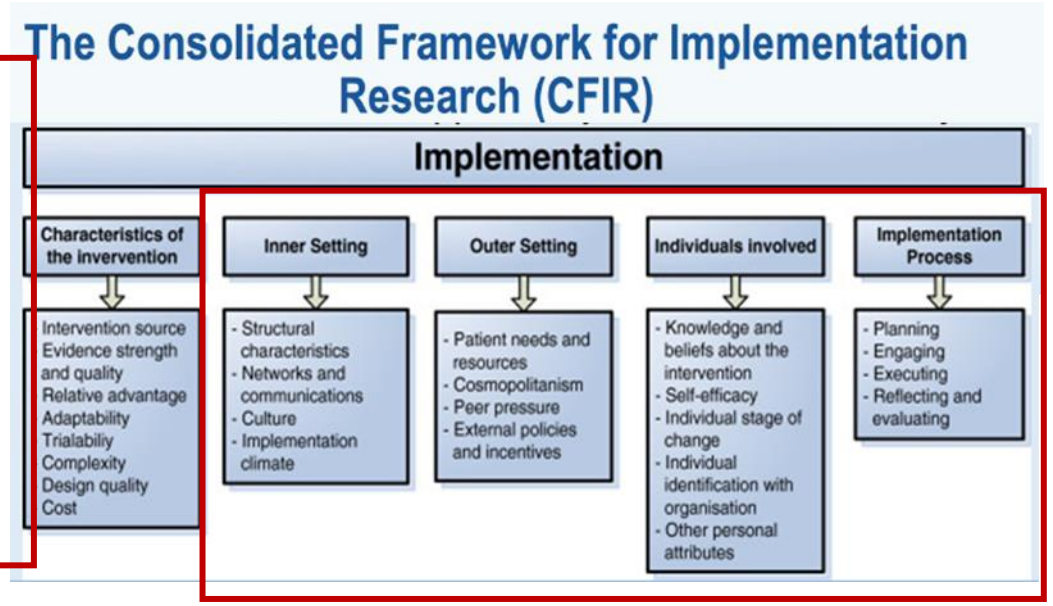




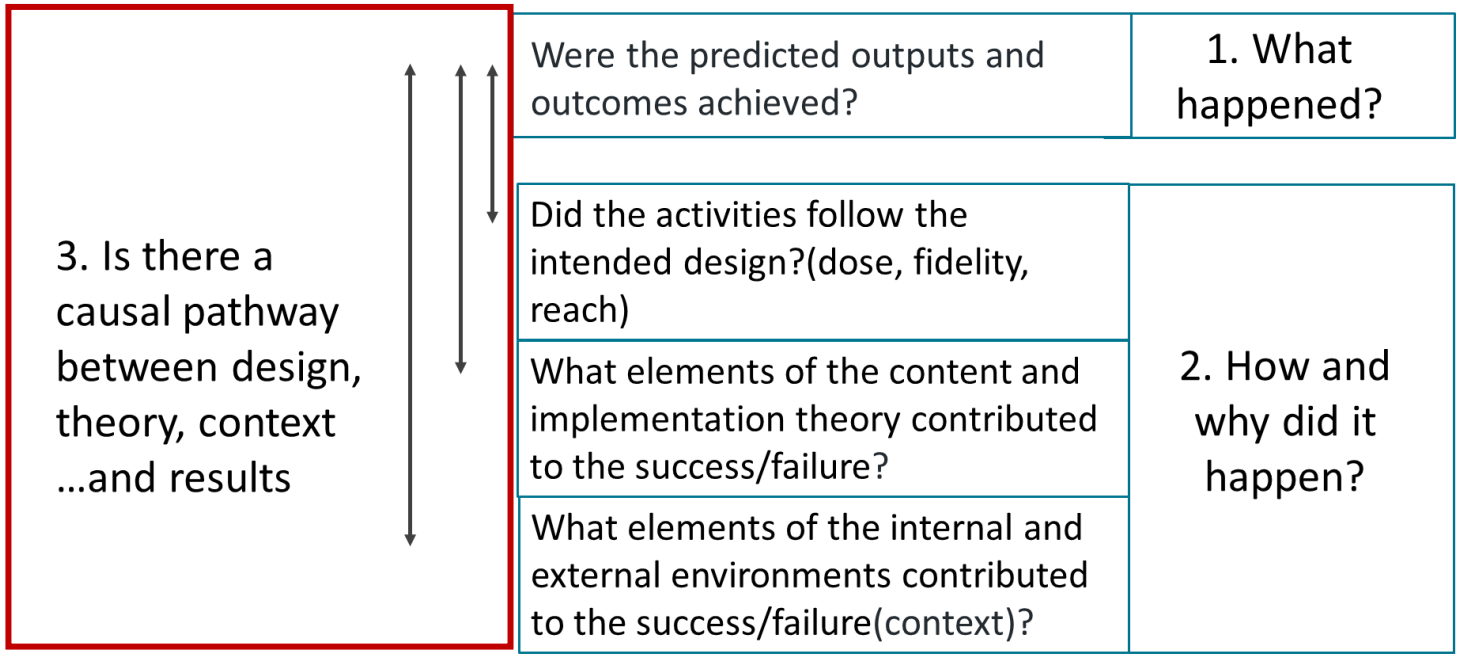
3. What role did the environment/context play?

3. Is there a causal pathway between design, theory, context …and results

Were the predicted outputs and outcomes achieved?	1. What happened?
Did the activities follow the intended design?(dose, fidelity, reach)	2. How and why did it happen?
What elements of the content and implementation theory contributed to the success/failure?	
What elements of the internal and external environments contributed to the success/failure(context)?	



1. Did the activities follow the intended Design?



Study Types to Infer Causality and Learning

- Quasi- experimental, mixed methods studies
- Randomized studies (cluster, stepped wedge, adaptive)

3 Ways to Improve Deployment of QI Science

1. Establish your Project Theory: Core Components of Design
2. Use a Framework for Improvement Research and Evaluation (FIRE)
3. The “Holy Grail”: Use study designs that support the causal pathway

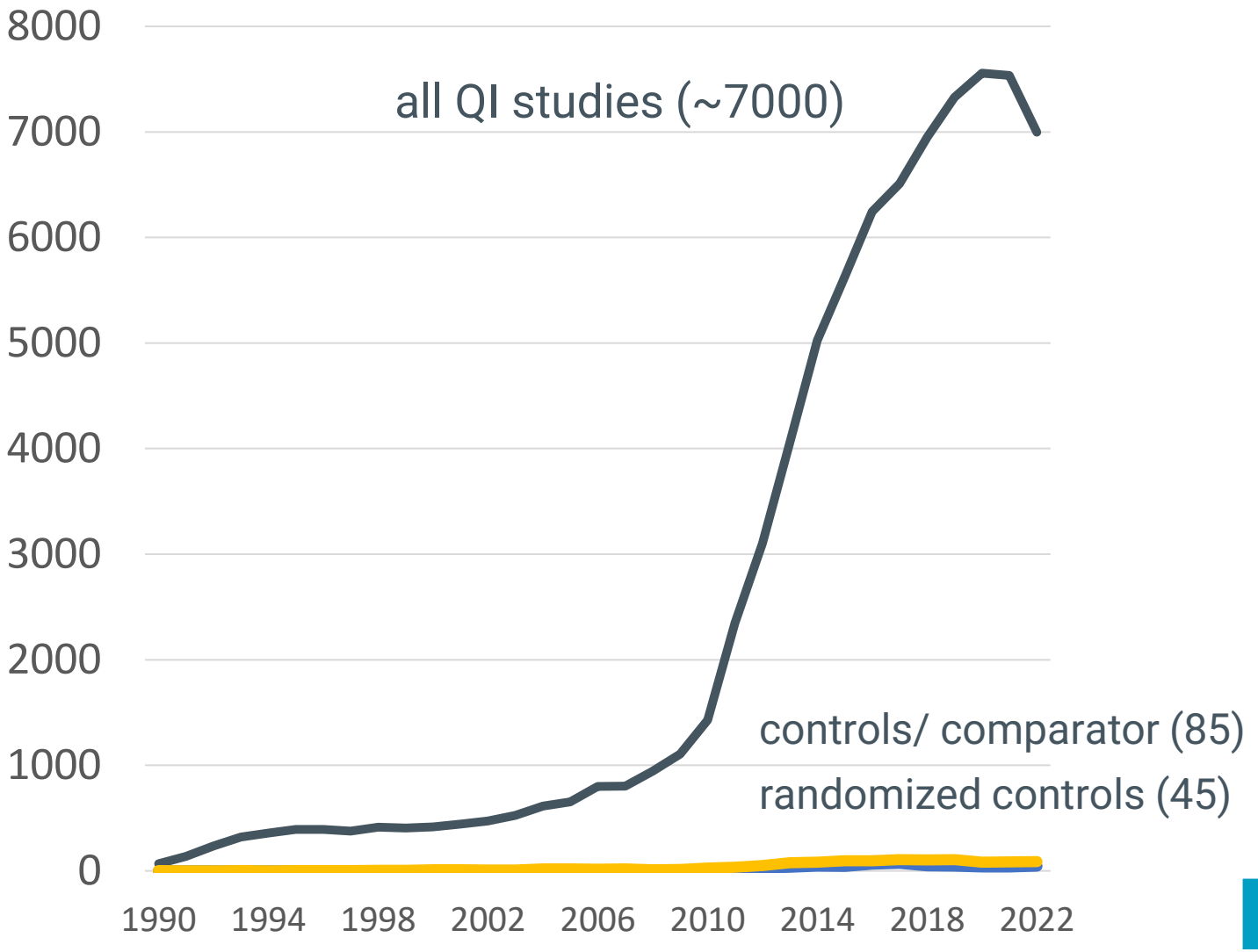
Five Study Designs to Choose From

1. Quasi-experimental/interrupted time series (may include a comparison group)
2. Stepped wedge randomized and non-randomized trials
3. Cluster randomized controlled trials (cluster RCTs) (preferably adaptive, not fixed protocol)
4. Before/after trials with and without comparison groups
5. Observational studies with advanced analytics to get closer to a RCT (e.g., “target trials”)

Use of Counterfactuals in QI

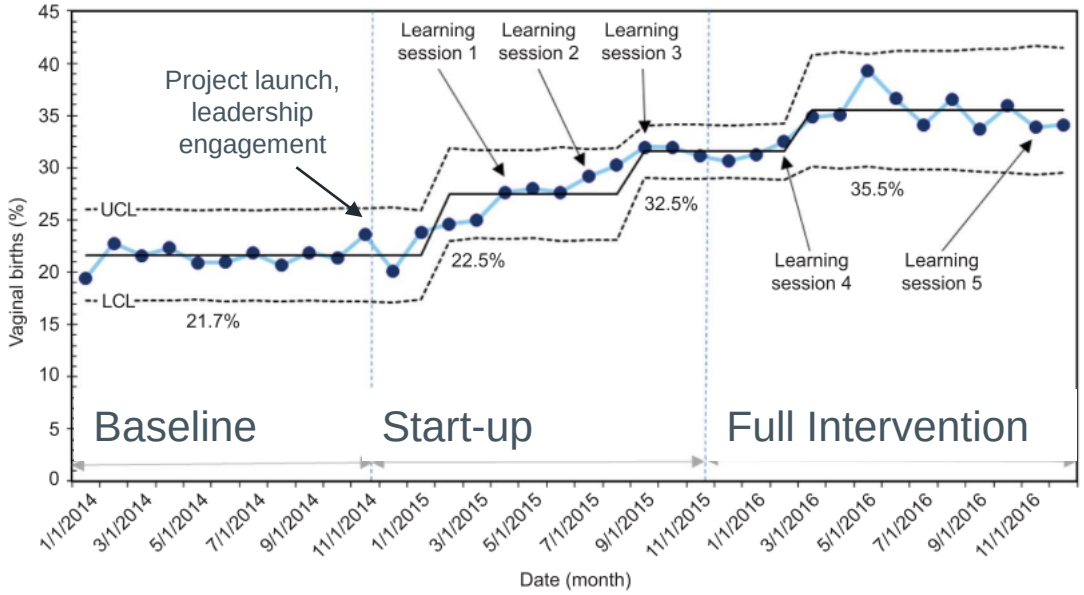
Miguel A. Hernán, MD, DrPH

The goal is to have a suitable comparison group to test the counterfactual – what would have happened had we not intervened due to secular trend or other interventions

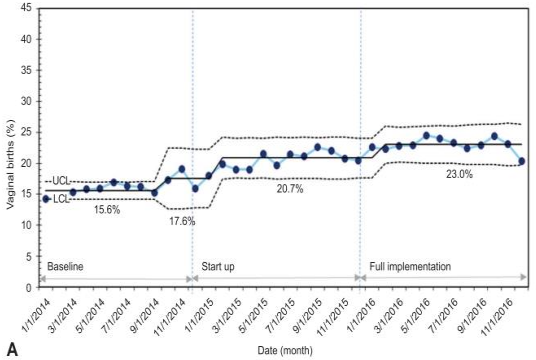


Testing the Counterfactual in Quasi-experimental Settings

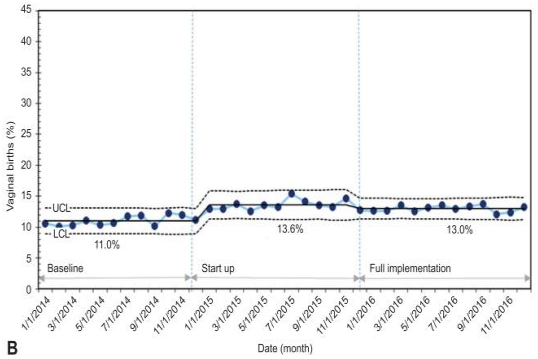
CS Study in Brazil: QI intervention to increase vaginal delivery rate in low risk pregnant women. 27 hospital Collaborative



Adding a Comparator: Non-intervention hospitals in Sao Paulo



5 hospitals in Sao Paulo that received the QI intervention (+8%)

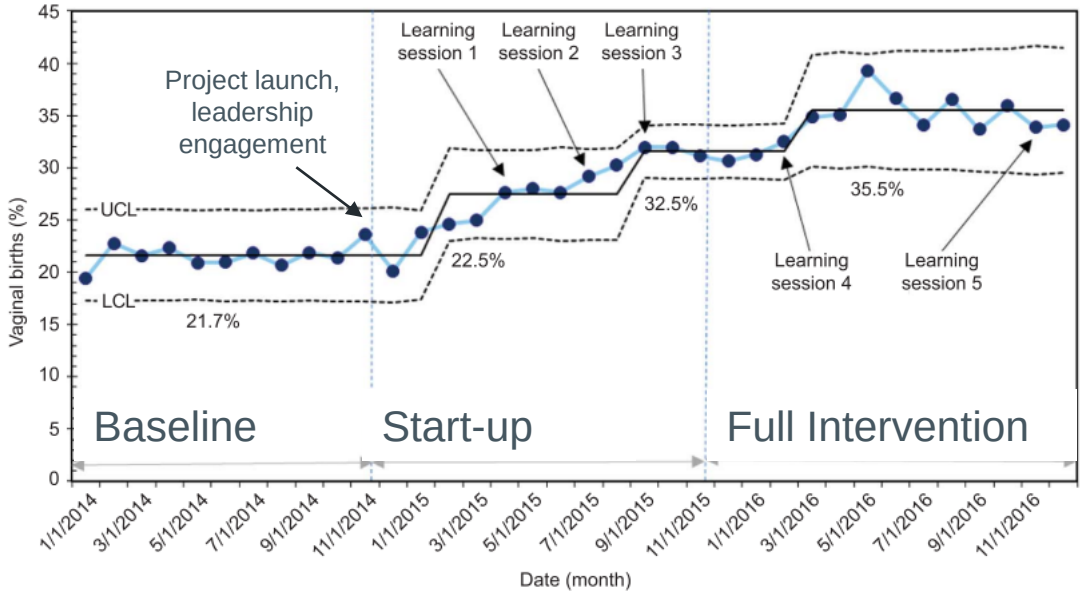


Remaining 8 hospitals in Sao Paulo that DID NOT receive the QI intervention (+2%)

Borem P, et al. Obstet Gynecol. 2020 Feb;135(2):415-425

Improving the Causal Path in Quasi-experimental Settings

Careful annotation of Shewhart charts



Borem P, et al. Obstet Gynecol. 2020 Feb;135(2):415-425

Planned Experimentation: Factorial Analysis

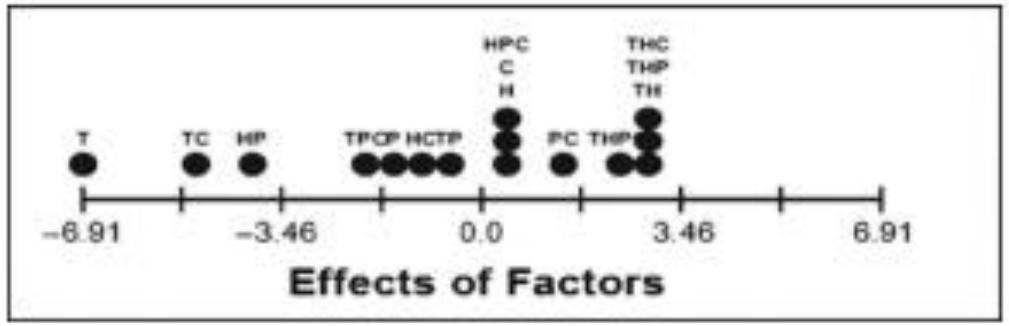
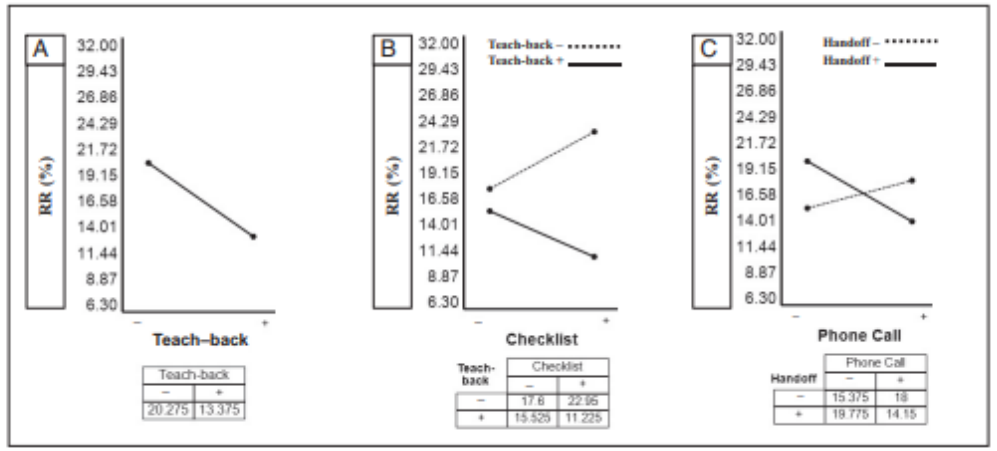


FIGURE 2
Dot diagram revealing important effects on reducing RR (left to right): teach-back (T), interaction effects T plus a checklist (C), and handoff (H) plus a phone call (P). Cluster of bundle elements around 0 represents "noise."



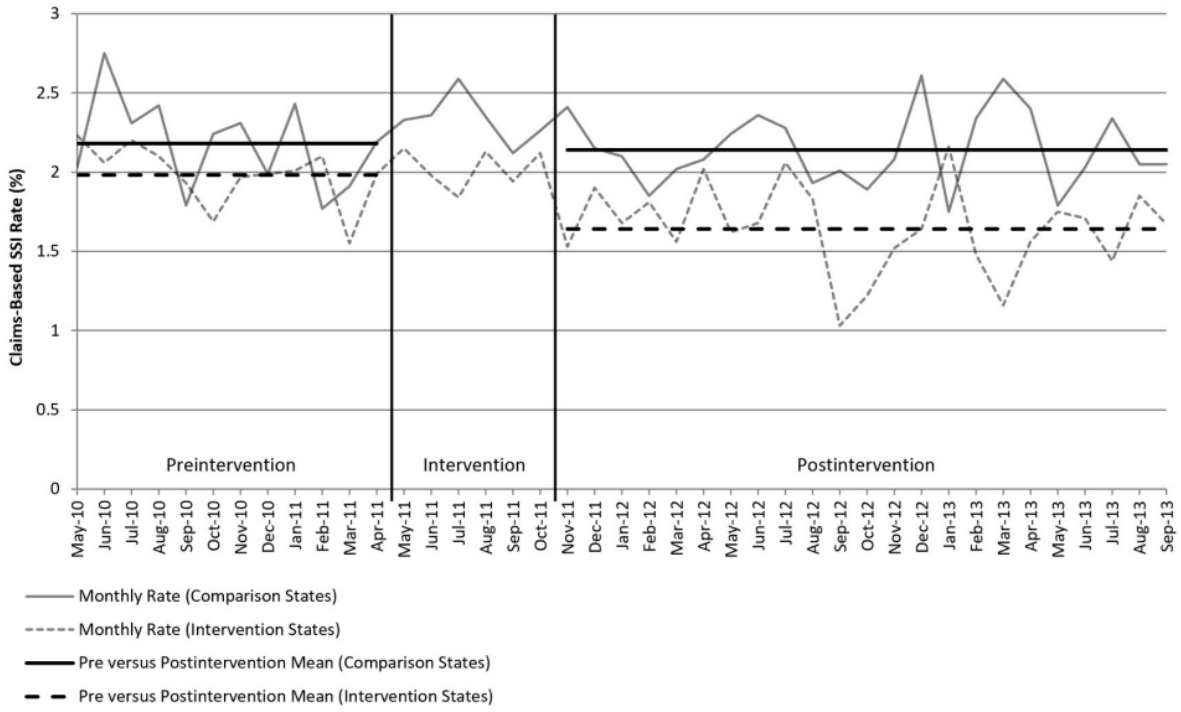
Nena Osorio et al. Factorial Analysis Quantifies Effects of Pediatric Discharge Bundle on Hospital Readmission, Pediatrics. 2021;148(4):

Testing the Counterfactual using matched controls

Project Joints

- a multistate US QI campaign that used QI to promote prevention of surgical site infection (SSI)
- 5 early adopter states received campaign intervention and five matched comparison states
- 23% greater decline in the risk-adjusted SSI rate for hip arthroplasty (not for knee surgery)

Hip Surgery SSI rate

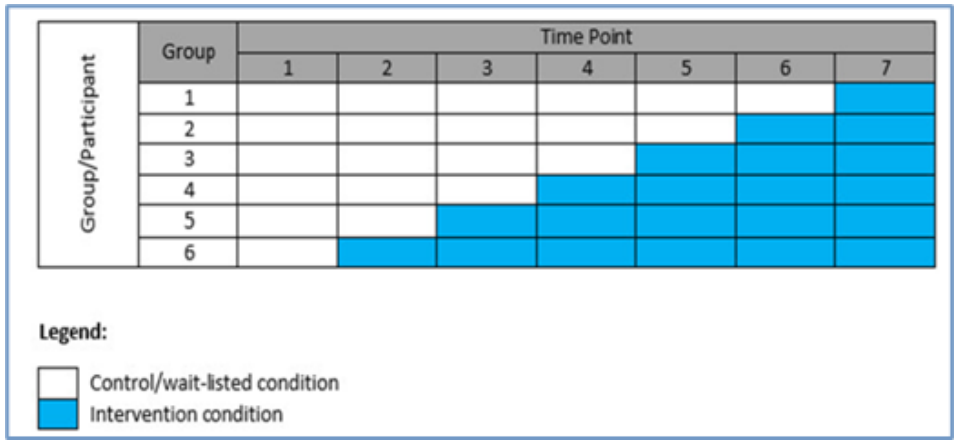


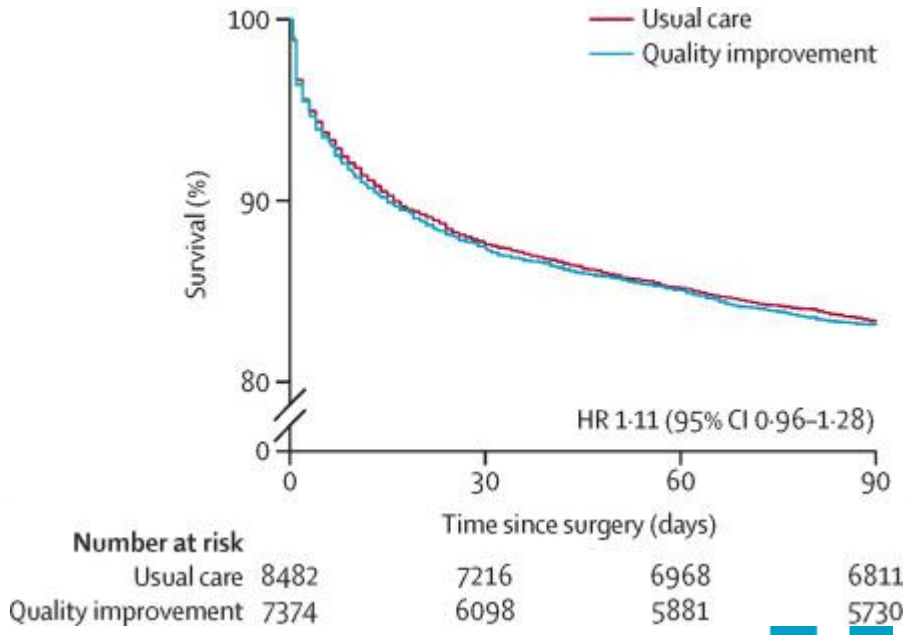
Calderwood MS, Yokoe DS, Murphy MV, et al. BMJ Qual Saf 2019;28:374–381.

Testing the Counterfactual in Step-Wedge trial (with or without Randomization)

NHS QI Programme to Improve Survival After Emergency Laparotomy (EPOCH)

- Stepped-wedge, cluster randomized controlled trial (cRCT)
- 15 geographical clusters of trusts, 16 “steps” of 5 weeks each (longest exposure to intervention 80 weeks)
- No difference between QI and “usual care”





Peden CJ, et al. Lancet 2019;393:213-21

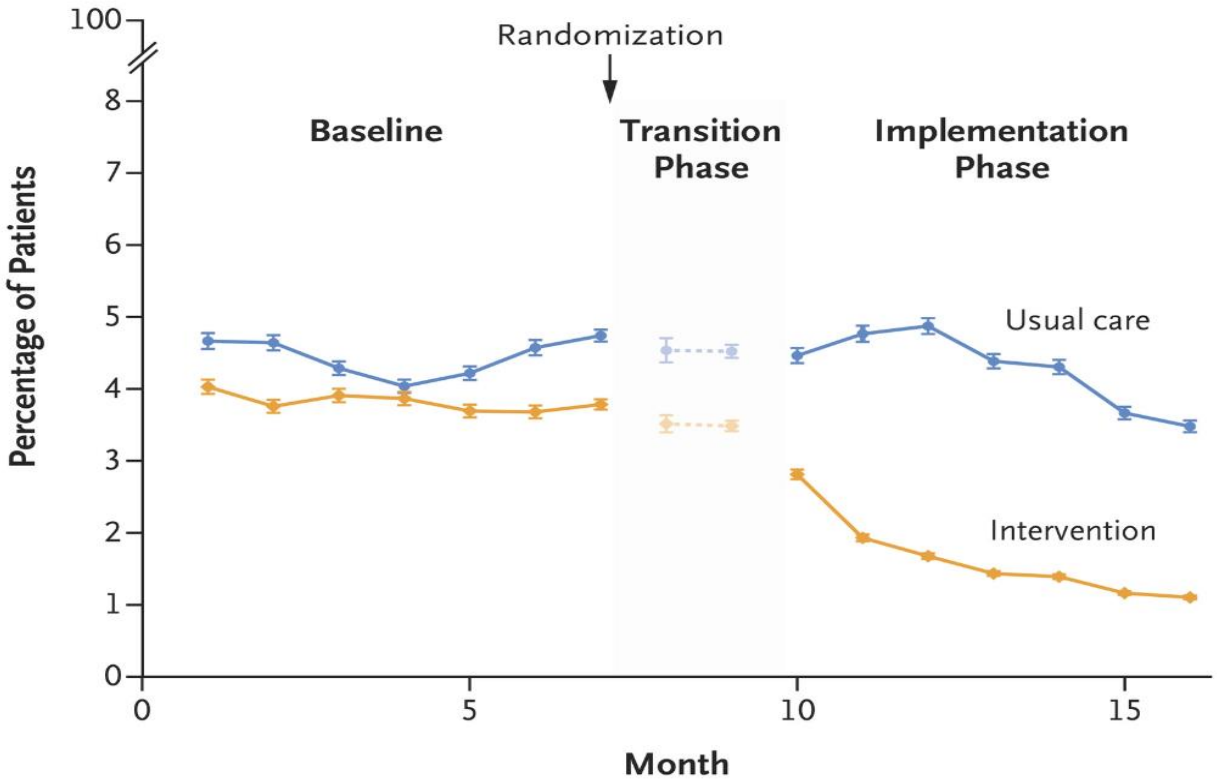
Testing the Counterfactual in cluster RCT: Maternal Hemmorhage

4 countries, 40 hospitals randomized (20/20) to receive/not receive hx prevention bundle to prevent and treat maternal Hx

4 primary interventions to implement Hx (E-MOTIVE) bundle

- Training
- Standard medicine/kit “trolley”
- Local champion
- Regular feedback

% patients with Hemorrhage, laparotomy, or maternal death from bleeding



Gallos et al, NEJM May 9, 2023

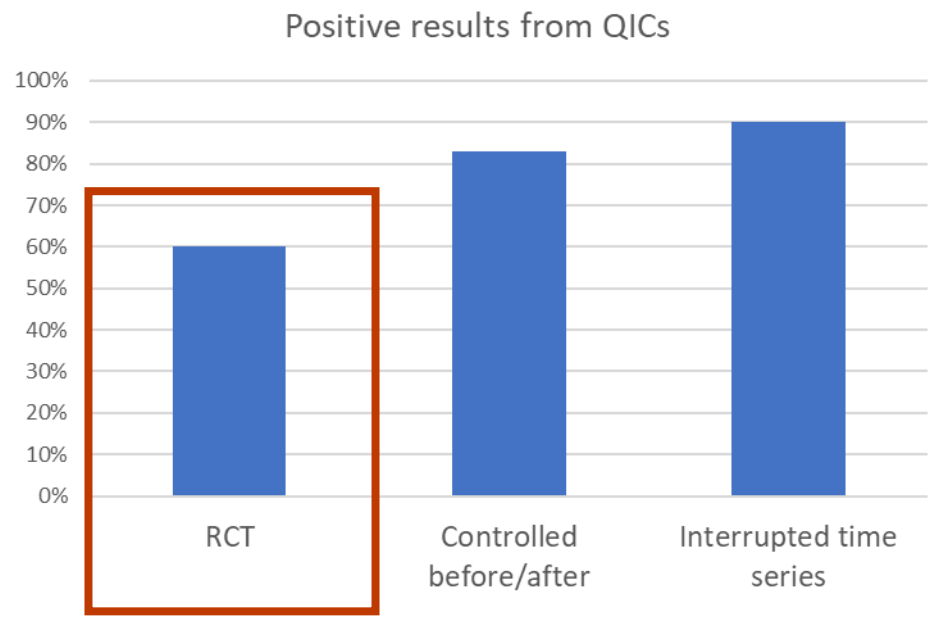
QI project perform less well in RCTs

Improvement collaboratives in health care. Health Foundation 2014
7 RCTs and Reviews
167 uncontrolled studies

Evidence about collaborative effectiveness

Impact	% of trials or reviews that found benefit	% of other studies that found benefit
Processes	33% of 3 studies	72% of 136 studies
Patient outcomes	20% of 5 studies	77% of 43 studies
Service use or costs	100% of 1 study	89% of 9 studies

Wells et al. Are quality improvement collaboratives effective? A systematic review. BMJ Qual Saf. 2018.
64 studies met EPOC study design standards for inclusion.
Positive results in 73% of the studies



Why do QI projects perform poorly in RCTs?

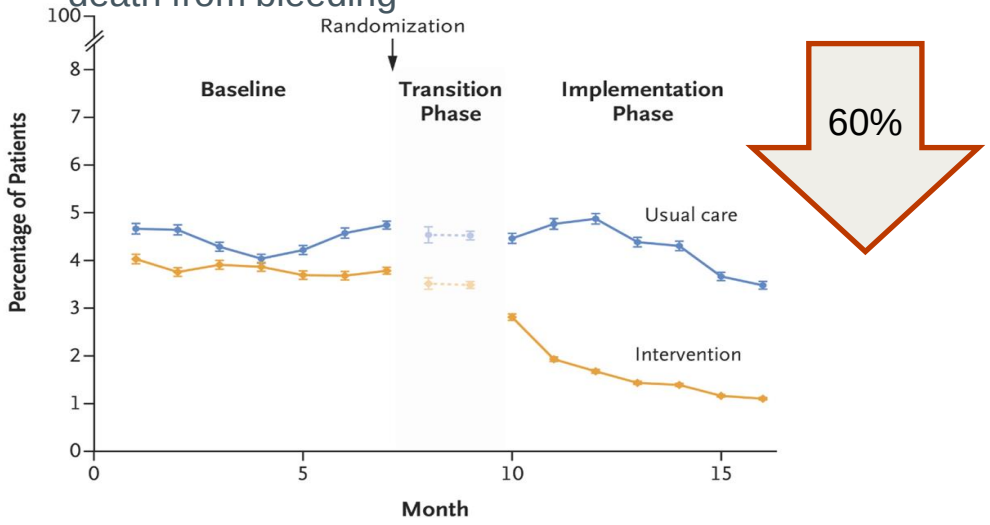
RCTs….

- Uncover bias
- Use fixed vs adaptive protocols
- Often lack fidelity to program theory: leave out key elements (e.g. learning networks, coaching)

Tale of 2 Studies:

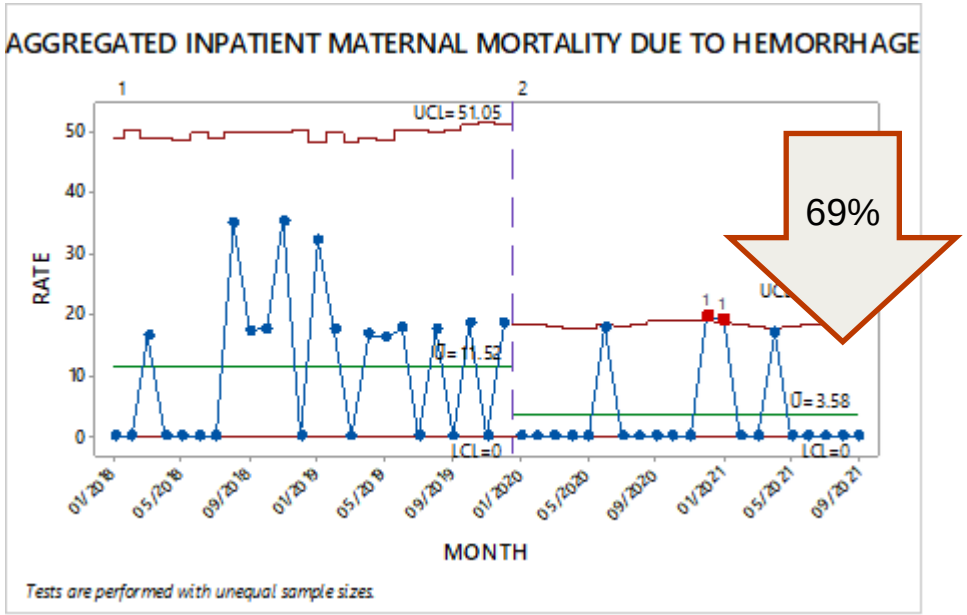
RCT – 20 intervention hospitals in 4 African countries (plus 20 controls)

% patients with Hemorrhage, laparotomy, or maternal death from bleeding



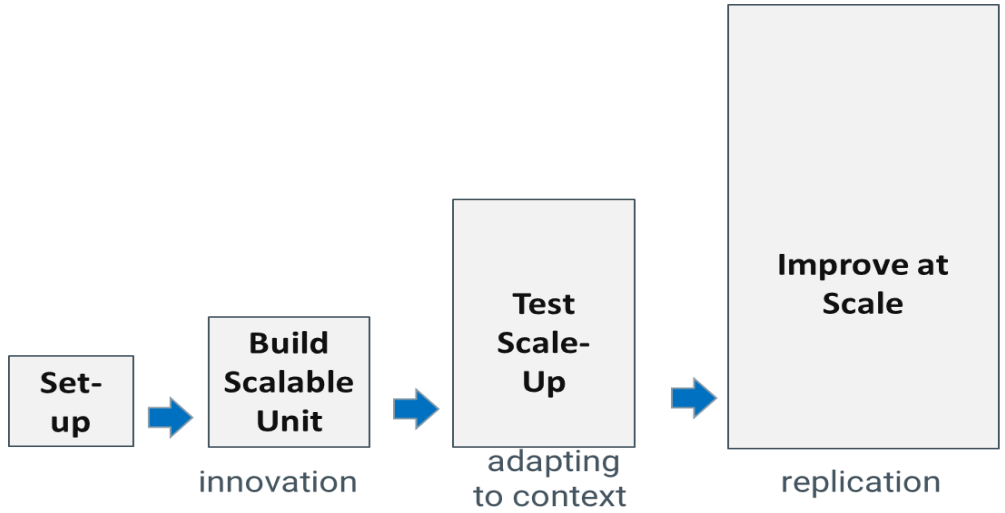
- 2 x 1h sessions in large conf rooms
- Co-session with WHO
- NEJM publication

Quasi-exp. QI – 19 intervention hospitals in Brazil (no comparators) Paulo Borem et al



- 10 min presentation in small IHI session
- Publication in QI journal anticipated

Can we enhance rigor of QI with Different Study Designs For Different Phases of Scale-up?



1. Phased approach to scale-up improvement

Use a strong core design with before/after counterfactual

Consider a comparator group

Consider randomization

Take-aways and Conclusions

- The current state of QI design, implementation, and evaluation is problematic
- Use epidemiological principles to mitigate bias and strengthen causal inference
- We can borrow from (and add to) evaluation and measurement frameworks commonly used in Implementation Science
- 3 ways to strengthen learning and the causal pathway
    - Establish your Project Theory: Core Components of Design
    - Use a Framework for Improvement Research and Evaluation (FIRE)
    - Use study designs that support the causal pathway