Don't forget to join in the conversations on twitter
Tweet us at #quality2019

C4 #qfc4

Institute for Healthcare Improvement

BMJ

Institute for Healthcare Improvement

The International Forum on Quality and Safety in Healthcare, Glasgow, 2019

March 28, 2019

What Makes "Quality Improvement" an "Improvement Science?"

The "Science" Matters

Don Goldmann, MD
Chief Scientific Officer, Emeritus, IHI
Professor of Pediatrics
Harvard Medical School
Twitter: @DAGoldmann
dgoldmann@ihi.org

Sara Riggare, MSc
Karolinska Institutet
Twitter: @SaraRiggare
sara.riggare@ki.se

Roadmap

- What do we mean by “science” in general and “improvement science” specifically?
- How can we be confident that the outcomes we observe are attributable to the changes we make and the programs we implement? How can our claims be more credible?
  - Association *versus* causality
    - “Hill’s criteria” for causality
    - How we can get it wrong: bias, confounding, and counterfactuals
  - Levels of evidence and belief – what we know, how we learn
- Rigorous design and evaluation of improvement initiatives
- (Driver diagrams and logic models as expressions of “causal theory” (aka “program theory” or “change theory”))

Science

The scientific method in brief:

- Formulate a testable hypothesis (theory)
- Specify the outcome you expect if your hypothesis is correct
- Perform iterative experiments, collect data, and study results
- Update your confidence or "degree of belief" in your hypothesis
- Implement your idea, revise it, or abandon it based on your experimental results

Look familiar??

- *Scientific Improvement is based on hypothesis (change theory), prediction, iterative learning through experimentation and observation, and action based on analysis of the results*
  - The essence of Plan-Do-Study-Act (PDSA) cycles

A Personal Take on the Increasingly Confusing Terminology of "Improvement Science"

- Scientific regardless of name:
  - Science of improvement
  - Health care delivery science
  - Implementation science
  - Systems strengthening
  - Systems engineering
  - And now…."Learning Health Systems" journal and society and "Engagement Sciences"
- Scientific methods include
  - "Model for improvement" promulgated by IHI
  - Lean
  - Six Sigma
  - Lean Six Sigma
  - ….And other components of the scientific improvement toolbox

Implementation Science

- Now the “official” government language in the US
- A widely-used definition:
  - Scientific study of methods to promote uptake of research findings in real work practice settings to improve quality of care and population health (Eccles and Mittman, 2006)

Again, doesn’t this look familiar?

What, if Anything, is Different about Implementation Science?

- Emphasis on implementation outcomes of planned activities and interventions (e.g., fidelity, penetration, sustainability)
- Emphasis on context assessment
  - CFIR (Consolidated Framework for Implementation Science) https://cfirguide.org/
- Generally tends to favor randomized designs and to emphasize behavioral science

How to Make Academic and Research Colleagues Comfortable with Improvement Science

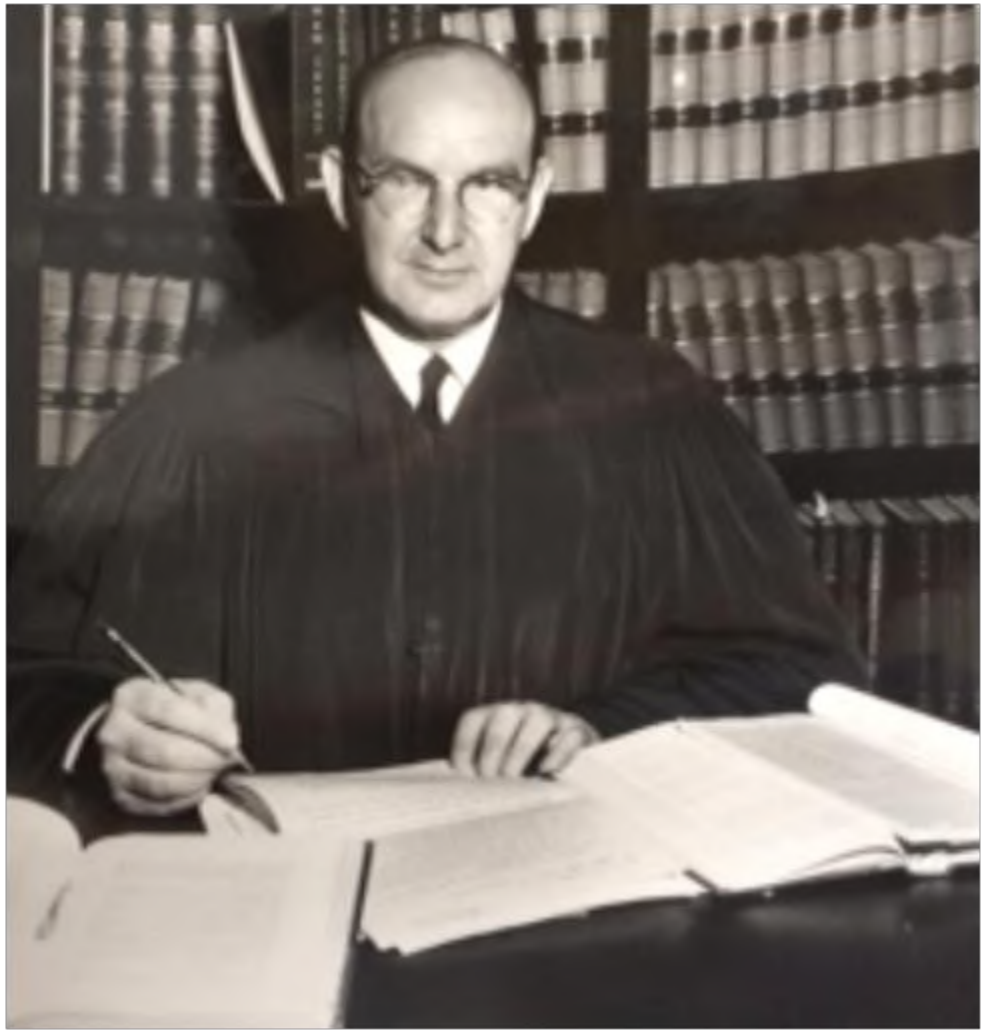
- Remind them it's basically just the experimental method + hefty doses of systems thinking, behavioral science, and insistence on graphic analysis of data over time
- My 10 years working with a PhD scientist to develop a staph vaccine…
  - Mice and PDSAs
  - Keeping a lab book

To Accelerate Improvement, Incorporate Other Scientific Disciplines in "Improvement Science"

- Behavioral science and behavioral economics
- Ethnography, anthropology, qualitative methods
- Information and data science
  - "Big data" analytics – machine learning and AI
- Network science
- Health economics
  - Determining "value" requires "running the numbers"
- And most of all, the broader field of health services research, especially epidemiology
  - What happened when I asked IHI improvement science students about bias and confounding….more in a minute about these terms

Building Evidence for What Works

- Improvers always should ask how confident they are that the change they are implementing will work in practice to improve quality
  - How strong is the evidence?
  - Or lacking strong evidence, what is the consensus degree of belief (confidence) in the change based on logic and experience
- Evidence and degree of belief are critical in both clinical medicine and in improvement and implementation programs

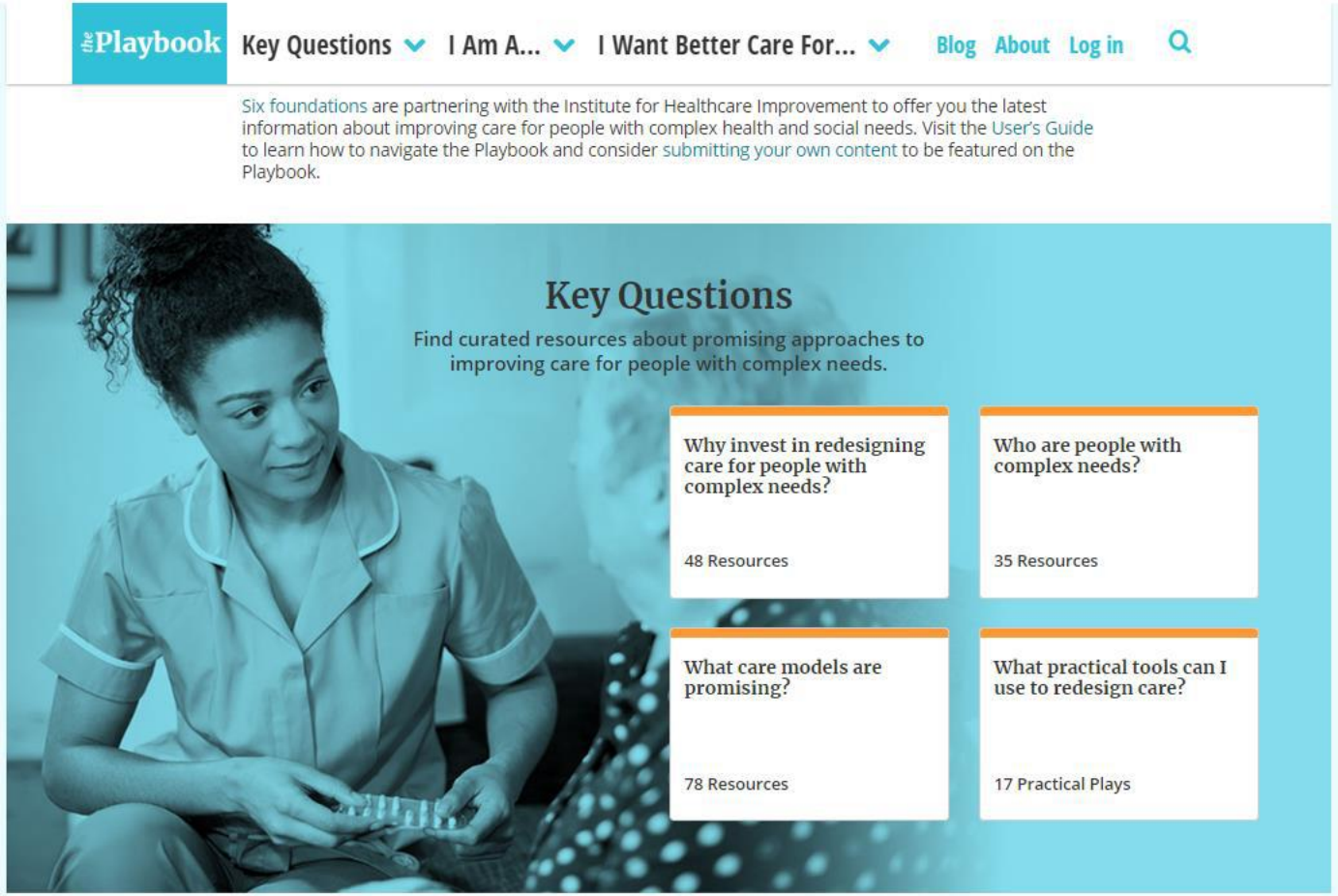


Judge Sidney Goldmann

Sidney Goldmann and Elizabeth

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**



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

We Can Do Better!

Critical Principle: Association is ***not*** the Same as Causation

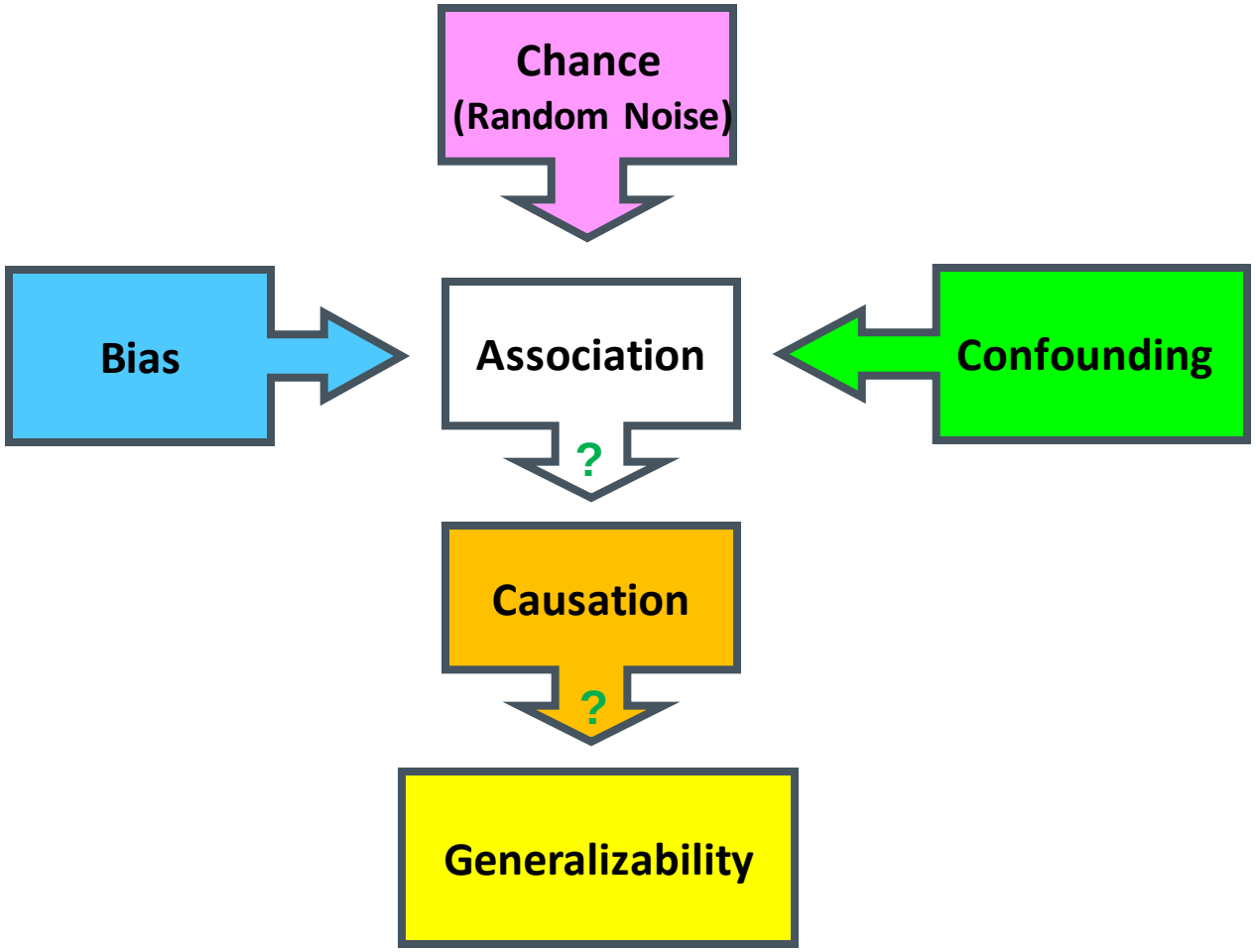
- When we act based on findings of our improvement programs, we tend to base our actions on association, not causation
- Examples of association, not causation:
  - Increased crime rates are statistically associated with ice cream sales, so does buying ice cream increase the crime rate?
  - More seriously, smoking is associated with a higher probability of getting lung cancer. Is this an association or an actual cause of lung cancer? (more in few minutes about this)

Association and Causation in Improvement Science

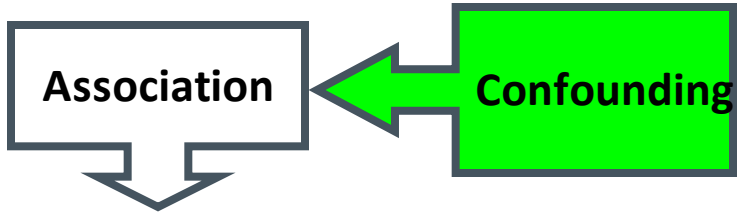
Be careful attributing improved outcomes to your interventions without formally considering how the improvements you are claiming may not be causal and could be misleading or even wrong

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…



Confounding



Confounder
(e.g., Patient Case Mix,
Severity of Illness)

Exposure
(QI Intervention
or treatment)

Outcome
(mortality,
readmission)

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
- The confounder must **not** be on the causal pathway between the exposure and outcome (not an "*effect modifier*" or "*mediator*").

Smoking and Lung Cancer: Association *versus* Causation

- Whether smoking was associated with or caused lung cancer used to be controversial
- R.A. Fisher (architect of the randomized controlled trial (RCT) as the "gold standard" of evidence), argued that a "confounding factor" – a lung cancer gene – explained the apparent effect of smoking
- A RCT was impossible, but a 1950 case-control observational study by Richard Doll and Austin Bradford Hill showed a clear association between smoking and cancer
- Surgeon General convened an advisory committee that developed 5 criteria suggesting causation in observational studies like Doll's
  - These criteria were refined by Hill in a 1965 paper: "Hill's 9 Criteria"
    - Hill regarded his "criteria" as "viewpoints:" "*None my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis, and none can be required as a sine qua non*"

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

Hill's Criteria – From Association to Causality

Plausibility	Does the postulated causal relationship make sense?
Strength	What is the size of the effect?
Consistency of association	Has the same effect been seen in other settings/studies?
Temporality	Did the exposure (intervention) precede the outcome?
Biological gradient	Does more exposure or higher fidelity of the intervention lead to a greater effect?
Coherence	Is the association consistent with existing theory and knowledge?
Experiment	Does modification of the intervention lead to a difference in outcome? (Was the experimental method used?)
Analogy	Did analogous interventions in other settings yield similar results?
Specificity of association	Could anything else have produced the observed result?

*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 Criteria in Improvement Science

- Plausibility (credible causal theory 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 (dose of an implementation activity delivered, dose received, 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)

Cross-Walk with Pronovost Catheter-Related Bloodstream Infection (CLABSI) Keystone Study

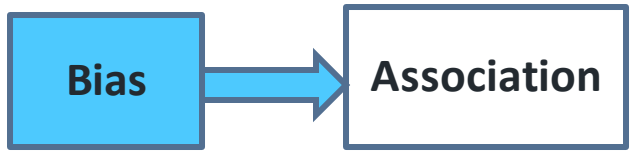
Plausibility	Plausible based on logic and evidence
Strength	Large (66% reduction in mean rate, median reduced to 0)
Consistency of association	Reductions seen in all 103 participating ICUs
Temporality	Yes, based on "generalized linear latent and mixed model" in six 3-month periods; data in table, **not displayed graphically over time**
Biological gradient	**Not specified**
Coherence	Consistent with available evidence; alternative theories not persuasive
Experiment	Natural experiment in which different centers activated components of intervention in different orders and times
Analogy	**When applied in England or SSIs in Michigan, no effect observed**
Specificity of association	**There was no comparison group or "*COUNTERFACTUAL*"**

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 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."
  - Consider whether there were "bright spots" where you did not intervene, not just where you did.
  - And were there comparable numbers of "dark spots" in intervention and non-intervention sites?

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?

- Selection bias (includes volunteer and “enthusiast” bias in QI)
- Performance bias (“trying harder” in QI)
- Detection/Ascertainment bias (enthusiasts look harder for supportive data)
  - Interviewer bias
- Attrition bias (people, organizations drop out along the way)
- Reporting/publication bias (just the good news, please!)
- Protopathic bias (disease already underway) (mitigated by inserting a lag time before the outcome)
- Indication bias (existing risk factors or conditions influence both decision to treat/intervene and, therefore, outcomes)
- Misclassification bias
- Lead time bias (early screening picks up a condition before it would be manifest clinically)(distort incidence estimates and outcomes)

Cochrane Handbook for Systematic Reviews of Interventions
http://handbook.cochrane.org/

Most published studies of QI interventions are low quality and do not meet publication standards promulgated by the Cochrane Collaboration or SQUIRE 2.0*

Are quality improvement collaboratives (QICs) effective? A systematic review. Wells S, Tamir O, Gray J, Naidoo D, Bekhit M, Goldmann D. BMJ Qual Saf. 2017 Oct 21. doi: 10.1136/bmjqs-2017-006926.

Conclusions:
QICs have been adopted widely as an approach to shared learning and improvement in healthcare. 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*, and publication bias is likely.

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

A Worthy Goal for Improvers:

Design studies that will determine if observed effects can be attributed to QI interventions

Diverse Study/Evaluation Designs from which to Choose

- Cluster randomized controlled trials (cluster RCTs)
- Stepped wedge randomized and non-randomized trials
- Interrupted time series trials
- Before/after trials with and without comparison groups
- Observational studies with attention to exposure and follow-up (including propensity scoring, instrumental variables to mitigate confounding)
  - "Big Data" Mining and advanced analytics
  - Large simple trials
- Pragmatic trials and action and community-based participatory research
- Context-sensitive mixed methods research/Realist Evaluation

http://www.academyhealth.org/evaluationguide

Randomized Controlled Trials (RCTs) and Cluster RCTs – "The Gold Standard"?

- Randomization should reduce or eliminate confounding by randomly distributing confounders between intervention and control groups
- However:
  - Randomization may not be practical or ethical
  - RCTs are expensive and take a long time
  - Generalizability limited due to strict inclusion/exclusion criteria (people who "get in" don't represent the population in which the intervention will be applied)
  - The control group may be exposed to the intervention ("bleeding" into the comparison group) (MERIT rapid response team trial)
  - Retention of participants may differ between intervention and control groups (attrition bias)
  - Fixed-protocols limit adaptation of the interventions (the essence of QI)
  - Rate of outcomes of interest may be less than predicted, making the study futile due to low statistical power (MERIT trial)

Real-Time Learning and Adaptation in Randomized Studies

- Considerable current emphasis on moving from fixed protocol project designs to adaptive designs that inform agile revision of the implementation approach
  - Lessons from ICU trial evaluating MRSA and VRE screening in reducing MRSA and VRE transmission
  - Statistical power issues

Stepped-Wedge Trials

- Preferably randomized

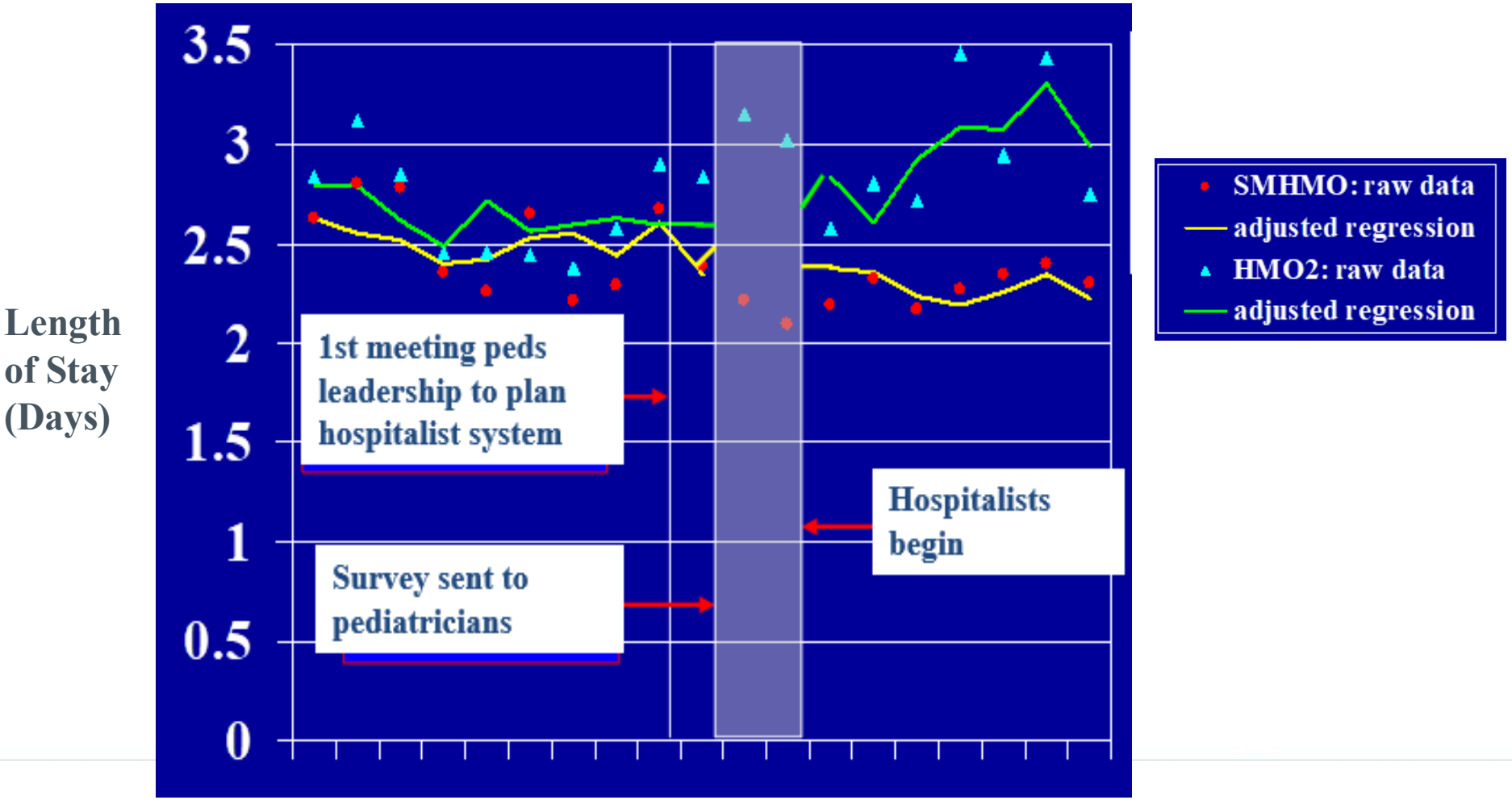
Group/Participant	Group	Time Point						
		1	2	3	4	5	6	7
	1							
	2							
	3							
	4							
	5							
	6							

Legend:

Control/wait-listed condition

Intervention condition

Interrupted Time Series with Non-Random Comparison Group (My First Try!)



Embedded versus External Evaluation

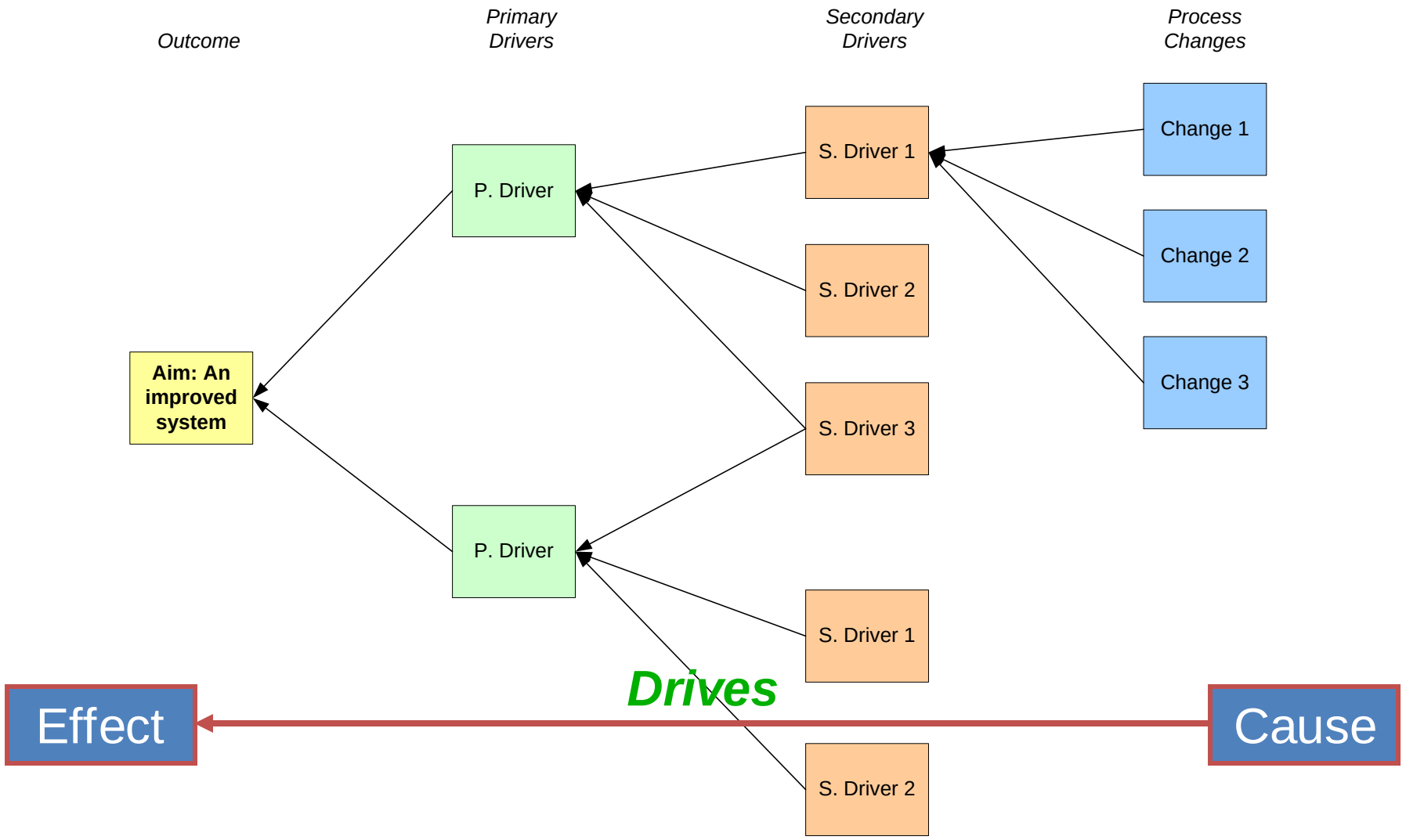
Conceptual Models, Logic Models, and Driver Diagrams All:

- Specify the causal theory for achieving outcomes
- Insure that everyone is on the same page
- Provide a framework for measurement
- Inform evaluation, whether embedded or external
- Usually are required for competitive grants and contracts
- Allow other organizations or researchers to compare their project/study design to what others have used

Driver Diagrams

- Clear, intuitive, visual demonstration of the most highly leveraged interventions and factors that are believed to promote the desired outcome
  - For example: Improved screening for hypertension → BP medication → BP control → Long term reduced AMI and stroke risk
  - It's still a theory and predictive model
- Easy to "hang" measures on each key driver
- Promotes specificity regarding the impact of specific changes on the key "drivers" and the causal pathway to the desired outcome
- May be useful to construct an "anti-driver" diagram or force field analysis to explicitly call out important barriers along the causal pathway
- Not ideal for showing complexity and interactions
- Ignore "less important" and unmeasured factors in determining the outcome

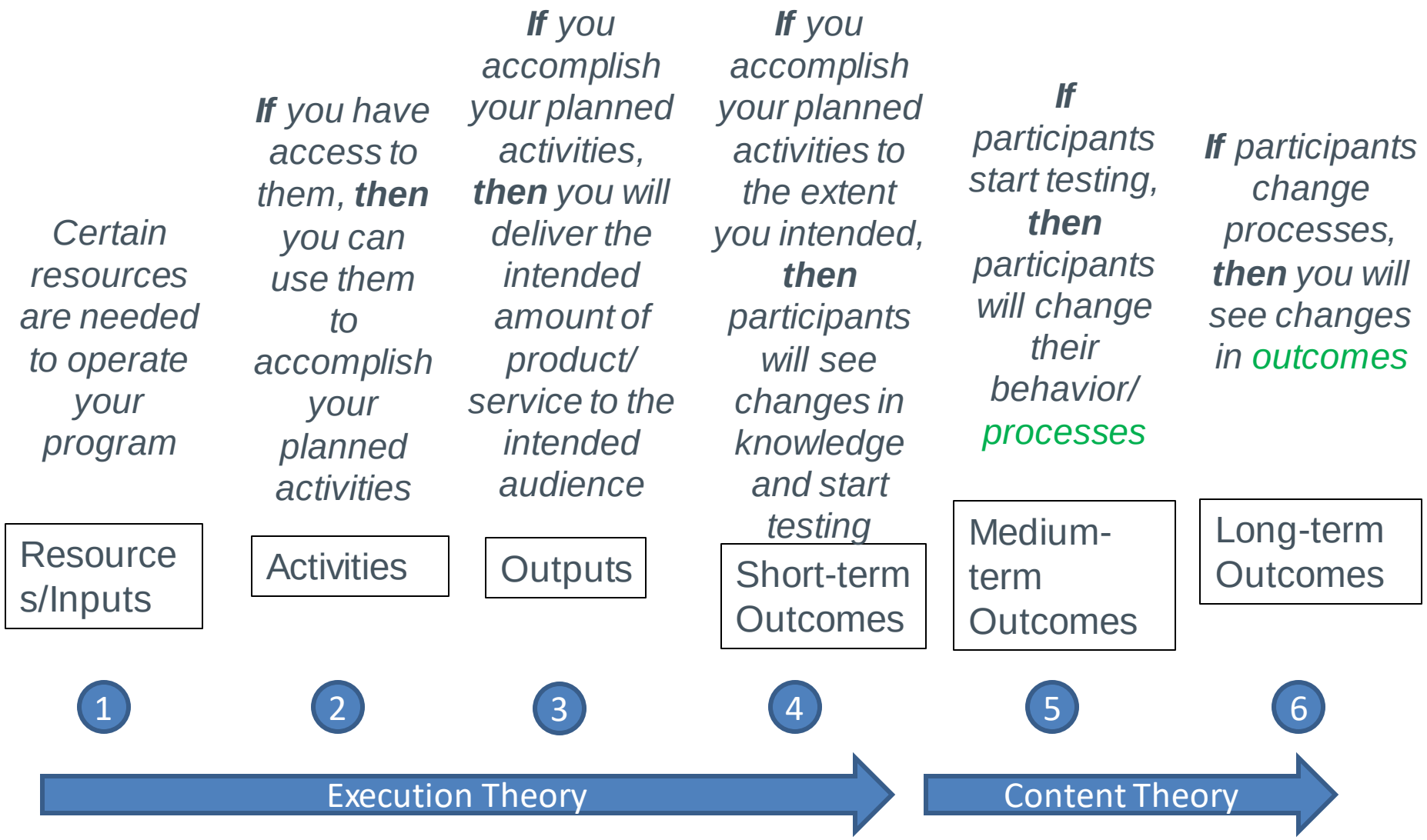
Cause-Effect Driver Diagram



Logic Models

- Guide and focus project planning, resource allocation, tasks, and work over time
- Clear delineation of exactly what will determine a given outcome, including
  - Inputs and resources
  - Internal and External factors and assumptions
  - Activities
  - Outputs
  - Outcomes (short, medium, long term)
  - Include implementation theory, activities, and outcomes
- Should enable evaluators to see exactly what was done, the intended "dose," and the received "dose"
- Support testing and replication elsewhere

The Basic Logic Model: "If-Then"



Project Title *(Month Year – Month Year)*

	Inputs	Activities	Outputs	Short term outcomes	Medium term outcomes	Long term outcomes
Context What is the 2-3 bullet elevator speech? What background info is necessary to understand this project ?	What are you investing to get the work done? *e.g., staff, funding, space, external partners, materials, tools*	What are you doing? *e.g., training in QI, monthly data analysis calls*	What are the reach and products of the activities? *e.g., 3 QI teams trained in improvement methods*	What changes in learning do you expect to come from your activities and outputs? *e.g., teams complete PDSA cycles*	What changes in processes do you expect as a result of your short-term outcomes? *e.g., blood protocol implemented with 80% reliability*	What changes in organizational and patient outcomes do you expect ultimately? *e.g., 20% decrease in under-5 mortality*

Assumptions

What are you assuming about the ability to deliver the program in the above planned way?

e.g., leadership is on board and the will is strong, surgeons will make the time to attend trainings and lead their QI team

External factors

What factors outside of the project may be a barrier or facilitator to reaching your desired outcomes?

e.g., new incoming government that prioritizes x, high staff turnover in health centers

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)