Introduction and aims of the Learning Lab

Pierre Barker

1:00 to 1:20pm
Introduction and Aims of the Minilab

● Introduction
  – Your IHI Faculty
  – The importance of evaluation in the improvement field
    – History of evaluation in QI field is poor
    – QI has been strong on using internal data (for improvement) but weak on designing data collection needed for rigorous evaluation.
    – Credibility of science of improvement is threatened by lack of attention to evaluation design
    – Lack of applicability of RCT not an excuse for weak evaluation design
    – Good evaluation designs exist for QI work.

Aims of the Learning Lab

● To describe what types of evaluations are most suited for improvement initiatives.
  – Formative Theory-Driven Evaluation

● To describe the questions an evaluator can ask to clarify how initiatives aim to bring about improvement.

● To describe the data an evaluator can collect to understand the progress of an improvement initiative.

● To describe the analytic approaches an evaluator can use to identify what was learnt during an improvement imitative.

● To complete an IHI Evaluation form
### Overview of the Learning Lab

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 1:20pm | Background to Evaluation  
  *Gareth Parry* |
| 1:50pm | Why most improvement initiatives are found to “fail”  
  *Kedar Mate* |
| 2:35pm | Break |
| 2:55pm | Program Theory  
  *Gareth Parry* |
| 3:35pm | How will we know?  
  *Kedar Mate* |
| 4:20pm | Wrap-up |

- Each Session will include an exercise aimed at completing the IHI Evaluation form

### The IHI Evaluation Form

- A blank form – to be completed by you!
- Example forms for you to refer to.
Background to evaluation

Gareth Parry

1:20 to 1:50pm

Hiraeth

- A Welsh word that has no direct English translation.
- Homesickness tinged with grief or sadness over the lost or departed, a mix of longing, yearning, nostalgia and wistfulness.
  - Some concepts can be expressed in a single word in one language while others require a detailed description.
  - Some things are just lost in translation
- What do we mean by:
  - Quality Improvement
  - Evaluation
  - Policy
Penicillin

- The Nobel Prize for Medicine was awarded in 1945 to:

Sir Alexander Fleming  Ernst B. Chain  Sir Howard Florey

1928
- From a single petri dish he had discarded, Alexander Fleming identified a mold and named it Penicillum.

1930s
- Fleming tried with limited success to produce and test the effects of Penicillin as a surface antiseptic.

1939
- Ernst Chain took an interest in Fleming’s work and with Howard Florey identified the active ingredient and produced it in larger quantities.
- Tested it by injecting Penicillin into two infected mice – who recovered.
Penicillin

1940s
- With Norman Heatley, produced sufficient quantities to test penicillin in fifty mice.
- Tested it in three people who were dying from bacterial infections.
- Subsequent successful clinical trials led to funding to mass produce Penicillin.

1945
- Sir Alexander Fleming, Ernst Chain & Sir Howard Florey awarded Nobel Prize for Medicine

Penicillin

- **Sir Henry Harris** said at the Florey Centenary lecture in 1998:
  - "Without Fleming, no Chain; without Chain, no Florey; without Florey, no Heatley; without Heatley, no penicillin."

- **Or:**
  - Without Fleming, no innovation; without Chain and Florey, no testing, without Heatley, no wide scale use of penicillin
Degree of belief

Degree of Belief in Change Ideas

High

Moderate

Low

Innovation Phase
(set design targets, develop ideas and predictions, and draft an initial conceptual model and change package)

Pilot Phase
(test and revise/amend conceptual model and change package)

Adapt and Spread
(Implement and disseminate a successful change package)
Exercise

- What is the overall purpose and goal of your project?

- What stage is your project at?
  - Innovation
  - Testing
  - Spread

- At your table, on the IHI Evaluation form, please complete questions 1, 2 and 3.

- Refer to the example form for ideas.

Why most improvement initiatives are found to “fail”

*Kedar Mate*

1:50 to 2:35pm
Donald T. Campbell

“The United States and other modern nations should be ready for an experimental approach to social reform…in which we learn whether or not these programs are effective, and in which we retain, imitate, modify or discard them on the basis of their apparent effectiveness on the multiple imperfect criteria available.”


Jonah Lehrer: The Truth Wears Off?

“In 2001, Michael Jennions, a biologist at the Australian National University, set out to analyze ‘temporal trends’ across a wide range of subjects … He looked at hundreds of papers … and discovered a consistent decline effect over time, as many of the theories seemed to fade into irrelevance.”

“… the data Ioannidis found were disturbing: of the thirty-four claims that had been subject to replication, forty-one per cent had either been directly contradicted or had their effect sizes significantly downgraded.”

Innovation to Prototyping: Small Number of Settings

Initial Testing: Small Number of Settings
More Settings as Range of Contexts Begins to Expand

Wide Range of Contexts
Reduction in Effectiveness from Applying the Same Fixed-Protocol Program in Different Contexts

Reduction in Effectiveness from Applying the Same Fixed-Protocol Program in Different Contexts
Effectiveness May Be Maintained If We Can Learn in What Contexts the Protocol Can Be Amended to Work

Learn which contexts it can be amended to work in as we move from Innovation to Prototype to Test and Spread

20,000+ Study Aim & Hypotheses

- **Study Objective:** Reduce perinatal transmission of HIV to <5% across a population of 5.5 M (220 health care facilities)
  - **Hypothesis 1:**
    - QI activities could achieve 90% or greater performance on six key process measures
  - **Hypothesis 2:**
    - A BTS Collaborative design would more rapidly improve processes in addition to routine use of QI methods
Study Design & Methods

- Study design: Cluster-RCT
- Intervention: QI methods and BTS Collaborative
- Unit of randomization: clinic supervisor and their associated 4-8 clinics (cluster)
- Planned randomization sequence:
  - 3 “waves” separated by 6 month intervals

Problems

- Unit of randomization (clinic supervisor) was eliminated in one district and reassigned in another
- The intervention could not be replicated in this new context (a traditional BTS Collaborative)
- Willing clinics could not be engaged
- Districts sought “contamination” (spread of best practices)
- Effects on process measures observed, but calculated outcomes not changing

- The focus was on whether our intervention worked
- Not on how and under what circumstances did it work
Types of Evaluation

- **Summative Evaluation:**
  - Aims to assess program outcomes or impacts
  - A study whose results will be used to decide whether the program can be generalized widely
    - Enumerative study
  - Will often involve quantitative data, hypothesis and theory testing and estimation

---

**Annals of Internal Medicine**

**Effects of a Quality Improvement Collaborative on the Outcome of Care of Patients with HIV Infection: The EQHIV Study**

Bruce E. Landon, MD, MBA; Ira B. Wilson, MD, MSc; Keith McInnes, MS; Mary Beth Landrum, PhD; Lisa Hirschhorn, MD, MPH; Peter V. Marsden, PhD; David Gustafson, PhD; and Paul D. Cleary, PhD

**Conclusions:** This prospective, matched study of almost 10,000 patients found that a quality improvement collaborative did not significantly affect the quality of care. Additional research is needed to improve methods of teaching and implementing quality improvement programs to achieve better results.

Types of Evaluation

Formative Evaluation:
- Aims to improve programs
- A study whose results will be used to amend or replicate the program in the future
  - Analytic study
  - Implementation evaluation
  - Mixed-methods studies
  - Realistic evaluation
- Will often involve hypothesis generation, theory forming, and qualitative methods

Conclusions: Updating program theory in the light of experience from program implementation is essential to improving programs’ generalizability and transferability, although it is not a substitute for concurrent evaluative fieldwork. Future iterations of programs based on the Michigan project, and improvement science more generally, may benefit from the updated theory present here.
What Type of Evaluation?

<table>
<thead>
<tr>
<th>Summative</th>
<th>Formative</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

Innovation Phase
(set design targets, develop ideas and predictions, and draft an initial conceptual model and change package)

Pilot Phase
(test and revise/amend conceptual model and change package)

Adapt and Spread
(implement and disseminate a successful change package)

Exercise

- What are the change concepts you are focusing on?
- What is your degree of belief that the change concepts can be amended to work in many settings?
- At your table, on the IHI Evaluation form, please complete question 4.
- Refer to the Example forms for ideas.
Break

2:35 to 2:55pm

Program theory

Gareth Parry

2:55 to 3:35pm
An Approach to Evaluation

- Assumptions from a Science of Improvement Perspective:
  - Learn what is takes to bring about improvement.
  - Improvement requires social change and that people are more likely to act if they believe.
  - Interventions need to be amended to local settings (contexts).
  - Concepts rather than fixed protocols are a good starting point for people to test and learn whether improvement interventions can be amended to their setting.

---

Ovretveit, Leviton, Parry, BMJ Qual Saf 2011
A Framework for Evaluation

The Kirkpatrick Evaluation of Learning Framework has four levels:

1. What was the participants’ experience?
   - Did the participants have an excellent experience working on the improvement project?

2. What did the participants learn?
   - Did they learn improvement methods and begin testing?

3. Did they modify their behavior?
   - Did they work differently and see change in their process measures?

4. Did the organization improve their performance?
   - Did they improve their outcomes?

Content Theory

WHAT CHANGES WILL TEAMS MAKE THAT WILL RESULT IN IMPROVEMENT?

Explains how we predict that the change concepts and improvement drivers applied in the project will lead to improved outcomes.
Execution Theory

**WHAT WILL YOU DO THAT WILL LEAD TEAMS TO ADOPT THE PROCESS CHANGES?**

Explains what IHI and other partners are doing that will lead front-line teams to adopt the changes (improvement drivers) described in the content theory.

- IHI Activities
- Participant Experience
  - Level 1
- Process/Behavior Changes
  - Level 3
- External Partner Activities
- Learning
  - Level 2

---

**Conceptual Model**

*What's the theory?*

- For example:
  - What are we trying to achieve?
  - Conceptual framework
  - Predicted improvement in what time?

**Execution**

- Experience
- Learning (testing)

---

**New Conceptual Model**

*What's the new theory?*

- Process
- Outcomes

---

**Who needs to know the new Conceptual Model?**

---

<table>
<thead>
<tr>
<th>What was the aspiration goal?</th>
<th>Experience</th>
<th>Learning (testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What would be considered a success?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Results over time.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What enabled or became a barrier for improvement?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What worked for whom over what time period?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Exercise

- What is the execution theory for your project?
- What measures are you using?

- At your table, on the IHI Evaluation form, please complete questions 5 & 6
- Refer to the Example forms for ideas.

---

How will we know?

Kedar Mate

3:35 to 4:20pm
Developing Evidence for True Effectiveness – “What Works” in Diverse Contexts and Populations

Degree of Belief in Change Ideas

High

Moderate

Low

Innovation/Prototype Phase

Pilot Phase

Adapt and Spread
Innovation Phase Evaluation

- Objectives:
  - Clarify the changes/interventions (content theory)
  - Did they work
  - Inform the next test

- Assess whether learning has been achieved (K2) and behaviors have been changed (K3)

- Focus on generating sufficient knowledge to inform the next test

Innovation Phase Evaluation

- Innovators may evaluate their models themselves
- External evaluators may check to see that the innovators are interpreting their data correctly

- Qualitative can be very helpful at nuancing the innovation
Prototyping Phase Evaluation

- Objectives:
  - Can a model from the innovation phase work in a new context;
  - Does the model need to be amended;
  - Update the degree of belief

- Strong reliable measurement system needed

- Qualitative approaches useful to refine the change/intervention (content theory) and the implementation approach (execution theory)

---

Prototyping Phase Evaluation

- Assess whether the model produced behavior changes (K3)
- Assess whether the model produced improvements in the system (K4)

- Counterfactuals help to increase degree of belief:
  - Historical (same organization) using time series
  - Non-historical (matched organizations)
Stepped Wedge Design

<table>
<thead>
<tr>
<th>Parallel</th>
<th>Stepped Wedge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Time</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0 1 1 1 1</td>
</tr>
<tr>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>2</td>
<td>2 0 0 1 1</td>
</tr>
<tr>
<td>3</td>
<td>3 0 0 0 1</td>
</tr>
<tr>
<td>4</td>
<td>4 0 0 0 0</td>
</tr>
</tbody>
</table>

Hussey, MA 2007

Prototyping Phase Evaluation

- Data from the evaluation must be fed back to encourage learning (formative)
- Requires close relationship between implementers and evaluators during design phase
- …and ongoing during the QI effort to allow mid-course changes in the design
Spread Phase Evaluation

- Objectives: increasing degree of belief in impact across a broad range of contexts
- The changes/interventions are well-tested at this stage
- The implementation approach (execution theory) is under investigation
- Measurement system looks at K1-4 levels
- Scale complicates qualitative approach—require sampling
- Findings fed back to the implementing sites (formative)

What Methods?
Rigor and Complexity: View #1

Simple
Linear
Cause-and-Effect

Complex
Non-Linear
Chaotic

Formal Evaluative Designs
RCT’s
“Rigorous”

Informal Narrative Study
Stories and Anecdote
“Non-Rigorous”

The “Evaluation Project” – Add Rigor to the Complex
What Methods?
Rigor and Complexity – View #2

“Rigorous” Learning

Simple Linear Cause-and Effect

Complex Non-Linear Chaotic

Poor Learning

Evidence in Simple Systems

“Rigorous” Learning

Fisher’s Tests Traditional RCT’s

Simple Linear Cause-and Effect

Complex Non-Linear Chaotic

Poor Learning

Case Study Anecdote
Diverse Study/Evaluation Designs

- Examples include:
  - Cluster randomized trials and meta-analytic trials
  - Bayesian and Adaptive trials
  - Pragmatic trials
  - Action and community-based participatory research
  - Quasi-experimental designs (e.g., factorial, time series designs, step wedge)
  - Context-sensitive mixed methods research/Realist Evaluation
  - Observational studies with attention to exposure and follow-up (including propensity scoring, instrumental variables)
    - Data Mining and advanced analytics

Exercise

- How will you know what the overall impact of your project is?
- How will you learn what is working and what needs to be amended?
- How will you update your degree of belief in the program theory?
- How will you develop an updated program theory?

- At your table, on the IHI Evaluation form, please complete question 7
- Refer to the Example forms for ideas.
Wrap up

4:20 to 4:30pm