SL3: Design and Evaluation of Improvement initiatives

Gareth Parry, Amrita Dasgupta, Don Goldmann

December 10, 2017

These presenters have nothing to disclose
IHI Faculty

Gareth Parry
Senior Scientist

gparry@ihi.org
@GJParry03

Amrita Dasgupta,
Senior Research Associate

adasgupta@ihi.org

Don Goldmann
Chief Medical and Scientific Officer

dgoldmann@ihi.org
@DAGoldmann
<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00 – 1:05</td>
<td>Welcome &amp; Introductions</td>
</tr>
<tr>
<td>1:05 – 1:20</td>
<td>Background</td>
</tr>
<tr>
<td>1:20 – 1:30</td>
<td>Aims</td>
</tr>
<tr>
<td>1:30 – 2:00</td>
<td>The What: Content Theory</td>
</tr>
<tr>
<td>2:00 – 2:30</td>
<td>The Where: Context</td>
</tr>
<tr>
<td>2:30 – 2:45</td>
<td>Break</td>
</tr>
<tr>
<td>2:45 – 3:15</td>
<td>The How: Execution Theory</td>
</tr>
<tr>
<td>3:15 – 3:45</td>
<td>Measurement</td>
</tr>
<tr>
<td>3:45 – 4:00</td>
<td>Evaluation Questions</td>
</tr>
<tr>
<td>4:00 – 4:30</td>
<td>Evaluation Designs</td>
</tr>
</tbody>
</table>
Talk to your neighbor:

What have you tried to improve this week?
Background

Gareth Parry

December 10, 2017
An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU

Peter Pronovost, et al December 2006

Conclusions: An evidence-based intervention resulted in a large and sustained reduction (up to 66%) in rates of catheter-related bloodstream infection that was maintained throughout the 18-month study period.
Conclusions: A multipayer medical home pilot, in which participating practices adopted new structural capabilities and received NCQA certification, was associated with limited improvements in quality and was not associated with reductions in utilization of hospital, emergency department, or ambulatory care services or total costs over 3 years. These findings suggest that medical home interventions may need further refinement.

Conclusions: Implementation of surgical safety checklists in Ontario, Canada, was not associated with a significant reductions in operative mortality or complications.
“..described in the 1980s by American program evaluator Peter Rossi as the “Iron Law” of … arguing that as a new model is implemented widely across a broad range of settings, the effect will tend toward zero.”
Reduction in Effectiveness from Applying Same Fixed-Protocol Program in Different Contexts

Where Can A Protocol Be Amended to Work?

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

Core Concepts & Detailed Tasks

**Core Concepts**

*Use a reliable method to identify deteriorating patients in real time.*

*When a patient is deteriorating, provide the most appropriate assessment and care as soon as possible.*

**Detailed Tasks and Local Adaptations**

- **MEWS >=5**
- **MEWS >=4**
  - 2 Nurses 1 Physician
  - 1 Nurse 1 Physician
  - 1 Physician
Degree of Belief

Evidence

Act
Degree of Belief in Change Ideas

- **Innovation**: Generate/discover new models of care with evidence of improvement in a small number of settings.
- **Testing**: Test whether a model works or can be amended to work in specific contexts.
- **Scale up and Spread**: Implementation of models shown to apply in a broad range of contexts.
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.
### From an Improvement Perspective:

<table>
<thead>
<tr>
<th>Learning</th>
<th>Empower those at the point of care to test, predict, fail forward and <em>learn</em> what is takes to bring about improvement.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Change</td>
<td>Improvement requires <em>social change</em> and that people are more likely to act if they believe. Work with, rather than doing to.</td>
</tr>
<tr>
<td>Context Matters</td>
<td>Interventions need to be <em>amended</em> to local settings (contexts).</td>
</tr>
<tr>
<td>Initial Concepts</td>
<td><em>Concepts</em> rather than fixed protocols are a good starting point for people to test and learn whether improvement interventions can be amended to their setting.</td>
</tr>
</tbody>
</table>

*We need: ‘Theory-driven rapid-cycle formative evaluation’*
Salzburg Global Seminar Session 565
Improvers and Evaluators as best friends

Evaluability Assessment

With all key stakeholders:

- Agree the Theory of Change
  - Five Core Design Components

- Agree the evaluation design, including:
  - Agreeing on the evaluation questions
  - Formative and/or summative approaches
  - Availability and Use of Data
  - Available human and financial resources

Five Core Components

<table>
<thead>
<tr>
<th>Core Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
</tr>
<tr>
<td>Aim Statement</td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
</tr>
<tr>
<td>Driver Diagram or Change Package</td>
</tr>
<tr>
<td>3) The Where: Context</td>
</tr>
<tr>
<td>Where will the improvement occur</td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
</tr>
<tr>
<td>Logic Model</td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
</tr>
<tr>
<td>Measurement Plan</td>
</tr>
</tbody>
</table>

That will maximize the chances that

The results and learning derived from the evaluation of an improvement initiative can be clearly communicated.
Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
<td>Aim Statement</td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
<td>Driver Diagram or Change Package</td>
</tr>
<tr>
<td>3) The Where: Context</td>
<td>Where will the improvement occur</td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
<td>Logic Model</td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
<td>Measurement Plan</td>
</tr>
</tbody>
</table>

Model for Improvement

- What are we trying to accomplish?
- How will we know that a change is an improvement?
- What change can we make that will result in improvement?
Five Core Components
1) Aims

Amrita Dasgupta
## Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
</tr>
<tr>
<td><em>Aim Statement</em></td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
</tr>
<tr>
<td><em>Driver Diagram or Change Package</em></td>
</tr>
<tr>
<td>3) The Where: Context</td>
</tr>
<tr>
<td><em>Where will the improvement occur</em></td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
</tr>
<tr>
<td><em>Logic Model</em></td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
</tr>
<tr>
<td><em>Measurement Plan</em></td>
</tr>
</tbody>
</table>

![Model for Improvement Diagram](image-url)
Hope is not a plan.

Some is not a number.

Soon is not a time.

-Don Berwick, MD
Aims in Practice

Supports planning backwards (what needs to get done when)
"Aspirational aims" are stretch goals used to inspire.
- Examples: End preventable maternal and newborn mortality for all.

"Achievable goals" are measurable targets believed to be achievable during a project’s timeframe.
- Answers “What are we trying to accomplish?”
- Captured in “aim statements”
A strong aim statement includes:

- **Numeric goals** for outcomes (how much)
- **Timeframe** (by when)
- **Boundaries** of the system being improved (where, who)

**Example:** Achieve a 30% reduction in facility-based maternal and neonatal deaths in participating facilities, within 30 months.
Building an Aim

**Pre-Work**
- Protect time to develop an attainable and informed aim
- Review what has been achieved in the past in similar work and settings
- Consider voices needed to set the aim and build buy-in

**Creating the Aim**
- Understand the current state in your system, answer a need in your community

**Ongoing**
- Check progress as you go and refocus aim as needed
Reflection: Building an Aim

- Which of these steps in building an aim have you done well?

- Which could you do better?
Exercise: What’s Your Aim Statement?

1. Own Project: Reflect and discuss in pairs
   - What are you trying to accomplish (your aim?)
   - What is the outcome measure that best captures the aim of your project?
   - What is the baseline level of performance on the outcome? How much does the outcome need to improve?

2. Share with the person next to you and give feedback.
2) Content Theory

Amrita Dasgupta
### Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
</tr>
<tr>
<td>Aim Statement</td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
</tr>
<tr>
<td>Driver Diagram or Change Package</td>
</tr>
<tr>
<td>3) The Where: Context</td>
</tr>
<tr>
<td>Where will the improvement occur</td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
</tr>
<tr>
<td>Logic Model</td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
</tr>
<tr>
<td>Measurement Plan</td>
</tr>
</tbody>
</table>

![Model for Improvement Diagram](image)
Content Theory (The What)

What changes will be made that result in improved outcomes?

- The processes we predict will improve patient/community outcomes.
- A driver diagram is a visualization of this theory.
Conceptual view of a driver diagram

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Primary drivers</th>
<th>Secondary drivers</th>
<th>Specific change ideas</th>
<th>Change concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim or outcome</td>
<td>Primary driver 1</td>
<td>Secondary driver 1</td>
<td>Ideas: 1</td>
<td>Concept 1</td>
</tr>
<tr>
<td></td>
<td>Primary driver 2</td>
<td>Secondary driver 2</td>
<td>2</td>
<td>Concept 2</td>
</tr>
<tr>
<td></td>
<td>Primary driver 3</td>
<td>Secondary driver 3</td>
<td>3</td>
<td>Concept 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary driver 4</td>
<td>4</td>
<td>Concept 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary driver 5</td>
<td>5</td>
<td>Concept 5</td>
</tr>
</tbody>
</table>

Key leverage points in the system

Specific ideas, concepts and bundles that could generate the desired state
Improving Newborn Mortality Initiative

Aim: Reduce newborn mortality by 20% across six district service areas by December 2018

Primary Drivers:
- Expand access to pre-natal care
- Activate community members to promote health and connect peers with health services
- Reliable delivery of pre-natal care bundle

Secondary Drivers:
- Supportive community structures
- Attractiveness of Health Services
- Communication & transportation
- Remove cost barriers
- Spread information to communities
- Identify and train community leaders
- Open lines of communication
- Community-based education
- Adequate supply of necessary materials
- Staff clinical knowledge and skills
- Use of data for decision making
Exercise: Prioritizing Drivers

- Brainstorm a list of drivers that impact your aim.
  - Which drivers do we believe will deliver the biggest impact?
  - Which ones will be easiest to work on? (Factors include personnel, culture, resources)
  - What is our current level of performance on these drivers?

- Select 3 as your “primary drivers.”
3) The Where: Context

Gareth Parry
## Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
<td><strong>Aim Statement</strong></td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
<td><strong>Driver Diagram or Change Package</strong></td>
</tr>
<tr>
<td>3) The Where: Context</td>
<td><strong>Where will the improvement occur</strong></td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
<td><strong>Logic Model</strong></td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
<td><strong>Measurement Plan</strong></td>
</tr>
</tbody>
</table>

### Model for Improvement

- **What are we trying to accomplish?**
- **How will we know that a change is an improvement?**
- **What change can we make that will result in improvement?**

The model is represented in a circular flow with steps labeled **Act**, **Plan**, **Study**, and **Do**.
Improvement of Any Kind Requires:

**Context:** Supportive macro environment and organizational culture.

**Improvement Capability:** Individual and team knowledge and skills of QI.

**Will**
To do what it takes to change the system

**Execution**
Improvement approaches to test the change

**Ideas**
Evidence on which to base the redesign of a system
Assumptions

**Will**
- Teams have chosen to participate.
- Leaders of the adopter communities are supportive and engaged.
- The initiative is a strategic priority – Policies exist to support this change at a national level.

**Ideas**
- High degree of belief in the changes being tested.
- The change package has been used with similar adopter communities with positive results.

**Execution**
- Teams have sufficient resources to participate in QI activities.
- Teams have protected time to test changes.
- Adopter communities have prior experience working on improvement initiatives
- Team have sufficient QI knowledge/skills to lead change
**DIRECTIONS FOR USE**

I. For each of the six areas, place an “X” below the level of capability that you think best fits your hospital’s current improvement capability and briefly describe the data/evidence you used to inform your choice. Descriptions for each level of capability can be found on pages 3-5.

<table>
<thead>
<tr>
<th>Levels of Capability</th>
<th>Just Beginning</th>
<th>Developing</th>
<th>Making Progress</th>
<th>Significant Impact</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Please provide a brief description of the type of data or other evidence you used to inform your choice.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1) Leadership for Improvement

2) Results

3) Resources

4) Workforce and Human Resources

5) Data Infrastructure and Management

6) Improvement Knowledge and Competence

II. Reflect on the results of your assessment:

- Does your assessment suggest one or more specific actions you can take soon to increase your hospital’s capability? Note these actions and who you would need to collaborate with to move ahead.

- Does your assessment suggest a need for more information to help you determine specific actions to increase your hospital’s capability? Note these needs.
Improvement Capability Self-Assessment

1. Leadership for Improvement
2. Results
3. Resources
4. Workforce and Human Resources
5. Data Infrastructure and Management
6. Improvement Knowledge & Competence
How can we use this tool?

- To stimulate discussion about areas of strength and weakness;
- To better understand your health system’s improvement capability; and
- To help you reflect on and evaluate specific improvement efforts.
**LEVELS OF CAPABILITY**

The levels below are intended to provide a basic indication of the improvement capability of your hospital in a number of domains that are associated with overall improvement success. This information is confidential; the more honest the assessment, the more likely the initiatives selected will be aligned with current ability and probability of success.

## Leadership for Improvement:

The capability of the leadership of the hospital to set clear improvement goals, expectations, priorities, and accountability and to integrate and support the necessary improvement activities within the organization.

<table>
<thead>
<tr>
<th>Just Beginning</th>
<th>Developing</th>
<th>Making Progress</th>
<th>Significant Impact</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are no clear organizational level improvement goals, expectations, and priorities. Improvement is seen as a department or service responsibility rather than requiring overall organizational leadership. Leadership for improvement is not coordinated across departments or services. Very little, if any, learning from improvement activities is shared across the hospital.</td>
<td>The hospital leadership has set clear improvement goals, expectations, and priorities through discussions with department and service leadership. Department or local leaders are held accountable for achieving the established goals without the support required for them to bring about improvement. Hospital leadership does not fully facilitate improvement activities across departments. Some learning from improvement activities is shared across the hospital.</td>
<td>Hospital leadership has prioritized some organizational level improvement goals to actively monitor and support. Hospital leadership focuses on the system of care and supports some local leaders to facilitate coordination of improvement activities across the services involved. Hospital leadership has established a system for sharing the learning from some improvement activities across the hospital.</td>
<td>Hospital leadership is actively engaged in monitoring and supporting all improvement goals. Hospital leadership focuses on the system of care and supports all local leaders in integrating and supporting improvement activities across the hospital. Hospital leadership has established a system for sharing the learning from all improvement activities across the hospital. Hospital leadership continually sets clear improvement goals, expectations, priorities, and accountability.</td>
<td></td>
</tr>
</tbody>
</table>

## Results:

The capability of a hospital to demonstrate measurable improvement across all departments and areas.

<table>
<thead>
<tr>
<th>Just Beginning</th>
<th>Developing</th>
<th>Making Progress</th>
<th>Significant Impact</th>
<th>Exemplary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some programs or services in the hospital can demonstrate measurable improvement, but this is not sustained over time and no sustained improvement can be demonstrated in any whole system organization-level measures.*</td>
<td>Although some programs or services in the hospital can demonstrate sustained and measurable improvement over time, very few if any of the whole system organization-wide measures can demonstrate improvement over time.</td>
<td>The hospital has demonstrated sustained improvement over time for a few whole system organization-wide measures.</td>
<td>The hospital has demonstrated sustained improvement over time for most whole system organization-wide measures.</td>
<td>The hospital can demonstrate sustained improvement over time for all whole system organization-wide measures.</td>
</tr>
</tbody>
</table>
Complete the assessment

Choose an organization you are working with in your improvement initiative:

1. Assess where the organization falls on one of the domains of the assessment.
2. Reflect on the following:
   - How does your assessment impact on the design of your improvement imitative?
   - What activities may you need to include and in what order?
Break
4) Execution Theory

Amrita Dasgupta
Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
<td>Aim Statement</td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
<td>Driver Diagram or Change Package</td>
</tr>
<tr>
<td>3) The Where: Context</td>
<td>Where will the improvement occur</td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
<td>Logic Model</td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
<td>Measurement Plan</td>
</tr>
</tbody>
</table>
Execution Theory (The How)

What will the improvement initiative do that will lead teams to adopt process changes?

- The rationale for how the experience, skills taught, activities and learning will lead to improvement.

- A Gantt chart or logic model is a visualization of this theory.
Why describe your execution theory?

- Clarifies your strategy
- Gets everyone on the same page
- Increases intentionality and purpose, sets priorities
- Identifies measures that matter
- Helps identify standard work & comparison across programs
What do we need to manage?

- Improvement Teams
- Data Systems
- Learning Events
- Coaching
- Leadership for Improvement
- Results Review
- Planning for Spread & Scale
Using Logic Models

**Execution Theory**

1. **Resources/Inputs**
   - Certain resources are needed to operate your program

2. **Activities**
   - If you have access to them, then you can use them to accomplish your planned activities

3. **Outputs**
   - If you accomplish your planned activities, then you will hopefully deliver the amount of product and/or service that you intended

4. **Outcomes**
   - If you accomplish your planned activities to the extent you intended, then your participants will benefit in certain ways

5. **Impact**
   - If these benefits to participants are achieved, then certain changes in organizations, communities, or systems might be expected to occur

**Content Theory**

Source: WK Kellogg Foundation, Logic Model Development Guide
How to draft a logic model

- Fill out your long term outcome
- Consider what medium and short term changes will be needed to reach that outcome
- In the inputs column, write in the resources you have to accomplish your long term outcome
- Have each person think silently first, then discuss with a team: what are the key activities you will do with those resources in order to achieve the aim
- Write the audiences you will reach and products you will create given those activities
- Follow a thread to check your reasoning
- Send around for people to review
- Meet to discuss changes that need to be made

GET YOUR FIRST BAD DRAFT OUT!
### Getting Started
- Develop a charter
- Expert meeting
- Organize QI teams
- Early improvement capability
- Getting leadership on board

### Learning Session 1:
- Introduce the Change package
- Model for Improvement (testing)

#### Action Period
- Review data and provide feedback
- Coaching calls

#### Outputs
- Change Package and Driver Diagram
- Measurement strategy
- Targeted number of sites enrolled
- Sites form improvement teams

#### Short term outcomes
- 75% Teams start testing

#### Medium term outcomes
- 60% Teams improve process measures

#### Long term outcomes
- 40% Teams improve outcome measures

### Learning Session 2:
- Highlight successful teams
- Model for Improvement (Implementation)

#### Action Period
- Review data and provide feedback
- Coaching calls

#### Outputs
- Site leadership engaged
- Teams have knowledge of how to apply MFI to the change package.

#### Short term outcomes
- 40% Teams start implementing

#### Medium term outcomes
- 80% Teams improve process measures

#### Long term outcomes
- 20% Teams improve outcome measures

### Learning Session 3:
- Teams present the successes
- Model for Improvement (Spread)

#### Action Period
- Review data and provide feedback
- Coaching calls

#### Outputs
- Teams learn from each other
- Teams have knowledge of how to apply MFI to implementation.

#### Short term outcomes
- 30% Teams start spreading within their organization

#### Medium term outcomes
- 90% Teams improve process measures

#### Long term outcomes
- 40% Teams improve outcome measures

### Inputs
- Staffing
  - Director
  - Expert faculty
  - Improvement Advisor
  - Administrative Support
- Capacity at site and improvement team level
- Leadership support at site level
- Funding
  - $X per site per year
Exercise: Logic Model Draft

- Write your aim in “impact” on the far right.
- Select 1 driver and write it in “mid term outcomes”
- Outline the inputs, activities, outputs and short-term outcomes that will lead to that system change.
- Share with a neighbor and give feedback.
5) Data, Measurement & Learning

Gareth Parry
## Five Core Components: The Model for Improvement

<table>
<thead>
<tr>
<th>Core Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Goals</td>
<td>Aim Statement</td>
</tr>
<tr>
<td>2) The What: Content Theory</td>
<td>Driver Diagram or Change Package</td>
</tr>
<tr>
<td>3) The Where: Context</td>
<td>Where will the improvement occur</td>
</tr>
<tr>
<td>4) The How: Execution Theory</td>
<td>Logic Model</td>
</tr>
<tr>
<td>5) Data Measurement &amp; Learning</td>
<td>Measurement Plan</td>
</tr>
</tbody>
</table>

### Model for Improvement

- **Act**: What are we trying to accomplish?
- **Plan**: How will we know that a change is an improvement?
- **Study**: What change can we make that will result in improvement?
- **Do**: Where will the improvement occur?
Data, Measurement & Learning

You cannot fatten a cow by weighing it.
## Example: The Surgical Sight Infection Family of Measures

<table>
<thead>
<tr>
<th>Types of Measures</th>
<th>Description</th>
<th>The Surgical Sight Infection Family of Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
<td>The voice of the customer or patient. How is the system performing? What is the result?</td>
<td>Surgical Sight Infection Rate</td>
</tr>
<tr>
<td><strong>Process</strong></td>
<td>The voice of the workings of the process. Are the parts or steps in the system performing as planned.</td>
<td>Percentage of appropriate prophylactic antibiotic selection. Percentage of on time administration of prophylactic antibiotics. Percentage of a safety climate score great than 4.</td>
</tr>
<tr>
<td><strong>Balancing</strong></td>
<td>Looking at a system from different directions or dimensions. What happened to the system as we improved the outcome and improvement measures?</td>
<td>Patient satisfaction Cost per case</td>
</tr>
</tbody>
</table>

System of Feedback

Measure Types
O = Outcome Measures
P = Process Measure
S = Process Step
PDSA = Learning cycle measure

Source: Brandon Bennett, Improvement Science Consulting
The Kirkpatrick Evaluation of Learning Framework

KP1  What was the participants’ experience?
   – Did the participants have an excellent experience working on the improvement project?

KP2  What did the participants learn?
   – Did they learn improvement methods and begin testing?

KP3  Did they modify their behavior?
   – Did they work differently and see change in their process measures?

KP4  Did the organization improve their performance?
   – Did they improve their outcomes?
System of Feedback

Measure Types
O = Outcome Measures
P = Process Measure
S = Process Step Measure
PDSA = Learning cycle measure

KP4: Outcome
KP3: Process
KP2: Testing
KP1: Experience

Source: Brandon Bennett, Improvement Science Consulting
Safety Culture Program Study

Mixed Methods Evaluation

**Quantitative: What was the impact?**

**Reaction & Learning (Kirkpatrick Level 1 & 2)**
- Staff support the Safety Culture Program
- Staff find the training favourable, engaging & relevant to their job
- Increased staff awareness, desire, knowledge, abilities & confidence following training

**Behaviour (Kirkpatrick Level 3)**
- Improved compliance with care processes designed to increase safety & reliability
- Change in the frequency of alleged unprofessional behaviors

**Outcomes (Kirkpatrick Level 4)**
- Improved perceptions of Safety Climate
- Improved quality and safety of patient care
- Improved patient experience
- Improved staff safety, satisfaction & engagement

**Qualitative: Why did it have this impact?**

- Pre-intervention interviews
  - January 2016
- Mid-intervention interviews
  - September 2017
- Post-intervention interviews
  - March 2019

*Take this out of the slides we upload.*
Exercise: Drafting Measures

- Review your Logic Model & Driver Diagram
- List up to 3 Outcome and Process measures you want to prioritize
- Brainstorm some ideas for how you will measures progress in how teams are ‘Testing’
Evaluation Questions

Amrita Dasgupta
Agreeing on the Theory of Change

The What

- **Aim:** Reduce newborn mortality by 20% across six districts by December 2018
- **Primary Drivers:**
  - Expand access to pre-natal care
  - Activate community members to promote health and connect peers with health services
  - Reliable delivery of pre-natal care bundle
- **Secondary Drivers:**
  - Supportive community structures
  - Accessibility of Health Services
  - Communication & transportation
  - Remove cost barriers
  - Spread information to communities
  - Identify and train community leaders
  - Open lines of communication
  - Community-based education
  - Adequate supply of necessary materials
  - Staff clinical knowledge and skills
  - Use of data for decision making

The How

- **Inputs:**
  - Strategies
  - Expert faculty
  - Improvement Advisors
  - Administrative Support
  - Capacity at site and improvement teams
- **Activities:**
  - Identifying the change
  - Developing the change strategy
  - Targeting number of site requests
  - Identifying teams
- **Outputs:**
  - Identifying teams
  - Identifying solutions
  - Identifying the change package
- **Short term outcomes:**
  - Examples of teams taking action
  - 90% teams start taking action
  - 80% teams improve process measures
- **Medium term outcomes:**
  - Teams learn from each other
  - Teams have knowledge of how to apply MVP to implementation
  - 60% teams implement
  - 50% teams improve process measures
  - 25% teams improve outcome measures
- **Long term outcomes:**
  - Teams have knowledge of how to apply MVP to spread changes
  - 50% teams implement
  - 50% teams improve process measures
  - 20% teams improve outcome measures

The Context
A Learning System

- Testing predictions is the heart of improvement
## Evaluation Design

<table>
<thead>
<tr>
<th></th>
<th>The What</th>
<th>The How</th>
<th>The Context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation phase:</strong></td>
<td>Model development typically takes place in a small number of settings, and evaluation questions should focus largely on <strong>The What</strong>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Testing Phase:</strong></td>
<td>The aim is to identify where the model works, or can be amended to work. Hence, although refining <strong>The What</strong> will occur, developing <strong>The How</strong> and <strong>The Context</strong> will also be important.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spread and Scale-up Phase:</strong></td>
<td>The aim is to spread or scaling up the model in contexts earlier work has indicated it is likely to work or can be amended to work. Here, the <strong>What</strong> and the <strong>Context</strong> should be well developed, and the focus will be primarily on <strong>The How</strong>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- To what extent can all the changes be implemented?
- What are barriers and facilitators to implementing the changes?
- What is the overall impact of the model on patient outcomes?
- Which elements of the model had the greatest impact on patient outcomes?
- To what extent can all the changes be implemented locally?
- What are barriers and facilitators to implementing the changes locally?
- Within what settings does the model work, or can be amended to work?
- To what extent did the implementation of the model vary across settings?
Available Resources

- What resources are available?
- What happens if we cannot demonstrate success?
A Learning System

Design
Initial Core Components

Improvement
Are we where we predicted we would be? If not, how do we amend the Core Components?

Close-out
Updated Core Components

Begin
End

Review A
Review B
Review C

Time
Exercise: Evaluation Questions

- Brainstorm evaluation questions that fit your project.
- Do the measures you outlined in the previous exercise answer these questions?
  - How might you change or add to these, if not?
- What resources do you have available to you for evaluation?
Evaluation Design

Don Goldmann
A Worthy Goal for Improvers: Design studies that will determine if observed effects can be attributed to QI interventions.
Quality of QI studies

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 effective? A systematic review. Wells S, Tamir O, Gray J, Naidoo D, Bekhit M, Goldmann D.

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.
Choosing a Project Design

- Select a design that answers the specific question you want to answer ("fit-for-purpose")
- Select a design that will allow you to attribute the effect you hope to see to the interventions you plan to implement
  - Account for the “counterfactual” (what would have happened if you did nothing – the secular trend)
  - Mitigate bias and confounding
Confounding variables **distort** an apparent association between an intervention and an outcome.
Properties of Confounders

- A confounder must be a cause (or a risk factor for) the outcome.
- The confounder also must be related to the exposure.
- The confounder must not be on the causal pathway between the exposure and outcome (not an intermediate variable).
Effect Modification

Effect modifiers reflect underlying differences in the mechanism of effect on a 3\textsuperscript{rd} variable/outcome.

Effect Modification Diagram:
- Exposure
- Modifier Present (genotype XX) → Outcome
- Modifier Absent (genotype YY) → Outcome
Bias

- 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
    - Interviewer bias
  - Reporting/publication bias (just the good news, please, preferably with a small p-value)
  - 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
  - Attrition bias (people drop out) (not fully mitigated by “intention to treat” analysis)
  - Lead time bias (early screening picks up a condition before it would be manifest clinically) (distort incidence estimates and outcomes)
Association is **Not** the Same as Causality.
## Sir Austin Bradford Hill – From Association to Causality

<table>
<thead>
<tr>
<th>Plausibility</th>
<th>Does the postulated causal relationship make sense?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strength</strong></td>
<td>What is the size of the effect?</td>
</tr>
<tr>
<td><strong>Consistency of association</strong></td>
<td>Has the same effect been seen in other settings/studies?</td>
</tr>
<tr>
<td><strong>Temporality</strong></td>
<td>Did the exposure (intervention) precede the outcome?</td>
</tr>
<tr>
<td><strong>Biological gradient</strong></td>
<td>Does more exposure or higher fidelity of the intervention lead to a more effect?</td>
</tr>
<tr>
<td><strong>Coherence</strong></td>
<td>Is the association consistent with existing theory and knowledge?</td>
</tr>
<tr>
<td><strong>Experiment</strong></td>
<td>Does modification of the intervention lead to a difference in outcome? (Was the experimental method used?)</td>
</tr>
<tr>
<td><strong>Analogy</strong></td>
<td>Are there analogous interventions in other setting that had similar results?</td>
</tr>
<tr>
<td><strong>Specificity of association</strong></td>
<td>Could anything else have produced the observed result?</td>
</tr>
</tbody>
</table>
## Cross-Walk with Pronovost Catheter-Related Bloodstream Infection Keystone Study

<table>
<thead>
<tr>
<th>Plausibility</th>
<th>Plausible based on logic and evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strength</strong></td>
<td>Large (66% reduction in mean rate, median reduced to 0)</td>
</tr>
<tr>
<td><strong>Consistency of association</strong></td>
<td>Reductions seen in all 103 participating ICUs</td>
</tr>
<tr>
<td><strong>Temporality</strong></td>
<td>Yes, based on “generalized linear latent and mixed model” in six 3-month periods; quarterly data in table, but not displayed graphically</td>
</tr>
<tr>
<td><strong>Biological gradient</strong></td>
<td>Not specified</td>
</tr>
<tr>
<td><strong>Coherence</strong></td>
<td>Consistent with available evidence; alternative theories not persuasive</td>
</tr>
<tr>
<td><strong>Experiment</strong></td>
<td>Natural experiment in which different centers activated components of intervention in different orders and times</td>
</tr>
<tr>
<td><strong>Analogy</strong></td>
<td>When applied in England, no effect observed</td>
</tr>
<tr>
<td><strong>Specificity of association</strong></td>
<td>There was no comparison group</td>
</tr>
</tbody>
</table>

Rigorous Learning in Complex Systems

- Diverse designs (quasi-experimental and adaptive trials)
- Diverse learning and evaluation methods (mixed methods)
- Diverse analytic methods (interrupted time series, SPC, multi-variable models)

Adapted from Berwick
Approaches to assess attribution

http://www.academyhealth.org/evaluationguide
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
Randomized Controlled Trials (RCTs)

- Considered the “Gold Standard” by evidence experts
- Randomization allows for difference in outcomes between the intervention and control groups to be attributed to the intervention since factors that might distort (confound) the association between intervention and outcome should be distributed randomly between the study groups

However:
- Randomization may not be practical or ethical
- RCTs are expensive and take a long time
- Limited “external validity” since inclusion and exclusion criteria are strict, and people who “get in” do not represent the population in which the intervention eventually will be applied
- The control group may be exposed to the intervention (“bleeding” into the comparison group)
- Retention of participants may differ between intervention and control groups
- 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
Bayesian Adaptive Trials - Apologies for Talking about Statistics

- In improvement, the “degree of belief” differs from the “degree of evidence” by incorporating Bayesian statistical inference.

- Bayesian approaches combine prior (existing) “degree of belief” or probability with new data, to calculate an updated “degree of belief” in the form of a new posterior probability.

- In other words, an existing probability of an outcome in a particular setting can be updated in the light of new data.

- This is critical when evaluating health care innovations using standard statistical methods.

- Constantly updating the degree of belief in one’s theory is the essence of quality improvement.
Real-Time Learning and Adaptation

Considerable current emphasis on moving from fixed protocol project designs to adaptive and Bayesian adaptive designs that inform agile amendment of the implementation approach.

- Note that adaptive trials are a relatively recent advance in design originating in pharma trials
- When we talk about adaptive trials in improvement we mean something slightly different…
- Lessons from ICU trial evaluating MRSA and VRE screening in reducing MRSA and VRE transmission and the Australian Rapid Response Team (MET) study
- Statistical power issues
Cluster Randomized Stepped-Wedge Trials

- Site 1
- Site 2
- Site 3
- Site 4
- Site 5
- Site 6
- Site 7
- Site 8
- Site 9

Site not allocated to the intervention
Site is allocated to the intervention

Time Period

1 2 3 4 5 6 7 8 9 10
A structural multidisciplinary approach to depression management in nursing-home residents: a multicentre, stepped-wedge cluster-randomised trial

A stepped wedge design assessed the effectiveness of the Act in Case of Depression (AiD) approach to the management of depression among nursing home residents in the Netherlands.

All five groups received the intervention in phases (steps).

The authors concluded that while the AiD approach can reduce depression, screening needs to be addressed in dementia units as these units had lower adherence to screening.

Ethical and Statistical Issues in Stepped-Wedge Designs

- Stepped-wedge trials allow “learning” between the “steps,” leading to helpful adaptations in implementation
  - This raises statistical power and analysis issues
- QI stepped-wedge trials generally do not have data safety and monitoring boards and tend to be carried to conclusion for continued learning rather than recognizing significant improvement and allowing all patients to benefit more quickly
Interrupted Time Series, No Comparison Group – Slope and Level Change

Massachusetts Chapter 24 Cost-Containment
Interrupted Time Series – No Comparison Group
Interrupted Time Series with Non-Random Comparison Group (My First Try!)

Length of Stay (Days)

Quarters

1st meeting pediatrics leadership to plan hospitalist system

Survey sent to pediatricians

Hospitalists begin

SMHMO: raw data
adjusted regression
HMO2: raw data
adjusted regression
Before/After Design with Comparison Group

- Pre Intervention
- Intervention
- Post Intervention

Effect

Risk-adjustment

Impact

Intervention
Control
Predicted
Before and After Design with No Comparison Group

![Graph showing the comparison between Pre, Intervention, and Post Intervention stages, with observed and expected effect measurements. The graph illustrates the impact of the intervention on the effect measure.]
<table>
<thead>
<tr>
<th>LEVEL OF EVIDENCE</th>
<th>TYPE</th>
<th>EXPLANATION/DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strong Evidence</strong></td>
<td>For individual studies</td>
<td>Concept tested in rigorous, randomized effectiveness studies (including cluster randomized controlled trials and randomized step wedge, factorial, or quasi-experimental designs) <em>(Note: Studies with these designs offer very strong evidence and should be ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td></td>
<td>Randomized Control Trial (RCT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-randomized Trial with Comparison Group</td>
<td>Concept tested in effectiveness studies with comparison groups (including non-randomized trials, before-after studies, interrupted time series studies, and repeated measures studies) <em>(Note: may be ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td></td>
<td>For reviews of multiple studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Meta-analysis</td>
<td>One or more meta-analyses of studies of an intervention showing effectiveness <em>(Note: these studies offer very strong evidence and should be ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td></td>
<td>Systematic Review (with evidence grading)</td>
<td>A systematic review of studies, giving a level of evidence for each study included <em>(Note: likely ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td><strong>Moderate Evidence</strong></td>
<td>For individual studies</td>
<td>Concept demonstrated to be effective in one or more rigorous observational studies with comparison groups and appropriate adjustment for bias and confounding <em>(Note: may be ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td></td>
<td>Rigorous Observational Study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For reviews of multiple studies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Systematic Review (without evidence grading)</td>
<td>A systematic review of studies, not giving a level of evidence for each study included <em>(Note: may be ready to implement with local adaptation.)</em></td>
</tr>
<tr>
<td><strong>Promising Evidence</strong></td>
<td>For individual studies</td>
<td>One or more rigorous real-world case studies</td>
</tr>
<tr>
<td></td>
<td>Case study</td>
<td></td>
</tr>
</tbody>
</table>
Exercise

• How will you design the improvement initiative in such a way that you can answer your evaluation questions?
• Brainstorm some ideas for how you will choose your evaluation design.


Bibliography


