Perspectives on Measurement for Improvement

The Microsystem Festival
Scientific Day
Jönköping
3\textsuperscript{rd} March 2017
Dr Tom Woodcock
Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Northwest London

• Translate research evidence into practice...
• ...to improve patient care, outcomes and experience
• Conduct world class research in improvement science
• Build capacity and capability for improvement
• Attract funding and industry partnerships

• Partnership between healthcare and academia
• Over 25 partner healthcare organisations
NIHR CLAHRC
Northwest London

Team Projects
Evidence to Practice

Fellowship
Clinicians, managers, patients etc

Other activities...
e.g. E-Learning
International Consultancy

A systematic approach to achieving successful improvements in healthcare
Based on QI methodologies

Research Inquiry – How effective are QI methods at supporting improvement?
What, if anything, needs to change to make them more effective?
Successful Healthcare Improvements From Translating Evidence into Practice (SHIFT-Evidence)

A Framework for Practice and Research
Enumerative vs Analytic Study

On Probability As a Basis For Action,
W E Deming, The American Statistician, Vol. 29 No. 4 1975, pp. 146-152

Analytical studies: a framework for quality improvement design and analysis,

“Because of the temporal nature of improvement, the theory and methods for analytical studies are a critical component of the science of improvement.”

Enumerative or Analytic?

1. Establish new patient wait times for appointment for each GP practice in the region

2. Do practices with a full time nurse practitioner have shorter waits than those without?

3. Will introduction of nurse practitioners in practices without one decrease their wait times?
## The 3 reasons for measurement

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Research</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>New knowledge</td>
<td>Improvement of service</td>
</tr>
<tr>
<td>Testing Strategy</td>
<td>One large test</td>
<td>Sequential tests</td>
</tr>
<tr>
<td>Sample Size</td>
<td>“Just in case” data</td>
<td>“Just enough” data, small sequential samples</td>
</tr>
<tr>
<td>Type of hypothesis</td>
<td>Fixed hypothesis</td>
<td>Hypothesis is flexible, changes as learning takes place</td>
</tr>
<tr>
<td>Variation (Bias)</td>
<td>Design to eliminate unwanted variation</td>
<td>Accept consistent variation</td>
</tr>
<tr>
<td>Determining if a change is an improvement</td>
<td>Statistical tests (t-test, chi square), p-values</td>
<td>Run charts, Shewhart control charts</td>
</tr>
</tbody>
</table>

Source: Solberg et al 1997
Improving the Quality of Quality Improvement Projects


“Case Example

At a recent patient safety meeting, the presenter suggested that a QI intervention in the presenter’s health system improved compliance with appropriate prophylaxis for deep venous thrombosis/pulmonary embolism (DVT/PE), reduced the incidence of DVT/PE, and, consequently, reduced patient complications and saved lives. […]

When an audience member questioned the validity of the results, the presenter clarified that the data were for ‘quality improvement’ not ‘research,’ implying [...] that QI projects are exempt from the rigorous methodological standards required of other research projects. In our experience, such views are widely promulgated among QI practitioners. …”
Some (common?) problems...

- Not knowing why we are measuring
- Measuring wrong/too many/too few things
- The denominator problem
- The baseline problem
- The feedback problem
- The rule-hacking problem
- The reporting problem
- The methodology problem
How is it supposed to work?

1. Decide aim
2. Choose measures
3. Confirm collection
4. Collect data
5. Analyse & present
6. Take appropriate action
7. Review measures
8. Repeat steps 4-6

The feedback problem

Adapted from a slide by Mike Davidge
Web Improvement Support for Healthcare

Plan Do Study Act cycles
Comments - context

Quantitative measure and SPC
The baseline-hacking problem
Another approach?

• Fix a minimum baseline period in advance of making any changes

• Decide and fix on rule-based criteria for starting a new “period” – 8 points in a row + identified special cause + no reverting 8 point rule-break

• Collect data for that baseline

• IF the pre-agreed criteria are met at some point after the end of the baseline; start new period
The rule-hacking problem

“Non-random patterns (special cause variation) were determined according to standard definitions (see bmj.com).”[citation]

Cited article:
“[...] Several other tests can also detect signals of special cause variation based on patterns of data points occurring within the control limits.8–11 Although there is disagreement about some of the guidelines, three rules are widely recommended:

• A run of eight (some prefer seven) or more points on one side of the centre line.
• Two out of three consecutive points appearing beyond 2 SD on the same side of the centre line (ie, two-thirds of the way towards the control limits).
• A run of eight (some prefer seven) or more points all trending up or down.

Lee and McGreevey recommended the first rule and the trend rule with six consecutive points either all increasing or all decreasing.”

... and the reporting problem
Towards Improved Reporting

Beginning a process of developing “standards” for reporting statistical process control analyses

1. Article in submission highlighting issue
2. Seek funding and interest
3. Formal consensus process
4. Standards
5. Evaluation of progress made
Improving Planning for Measurement in QI Initiatives

Aim: To develop a Measurement Plan Assessment Tool (MPAT) to help healthcare QI teams plan measurement effectively.
Methods

Current Project

Stage 1:
Identify evidence to support structure development and content generation
- Existing Literature
- Practical experience of supporting over 60 healthcare QI projects
- Previous evaluations of healthcare quality improvement initiatives such as The Health Foundation Safer Clinical Systems Programme

Stage 2:
Select and refine content
- 2-round modified Delphi survey
  (Formal expert consensus building process)

Stage 3:
Address emergent themes from Delphi survey and critically appraise draft tool
- Consensus Meeting

Future Work

Stage 4:
Create online resource with a bespoke graphical user interface and data repository
- Consultation with web designers and users
- Completed

Stage 5:
Assess validity and user acceptability
- Pilot testing
- Pending
Results

**Design**
- Aim
- Measure Set
- Operational Definitions

**Data Collection and Management**
- Data Collection Process
- Training in and Embedding of Consistent Data Collection
- Database Design
- Outliers and Missing Data

**Analysis**
- Planning the Analysis

**Action**
- Planning for Action

**Embedding**
- Planning for Sustainability

E.g. Have operational definitions been written for all the selected improvement measures?
# Results

## Table 1: Total number of questions per subsection, and % that reached the 75% consensus level at the end of the Delphi Survey

<table>
<thead>
<tr>
<th>Section</th>
<th>Subsection</th>
<th>Total number of questions in subsection</th>
<th>Total number of questions reaching consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design</strong></td>
<td>Aim</td>
<td>10</td>
<td>8 (80%)</td>
</tr>
<tr>
<td></td>
<td>Measure Set</td>
<td>13</td>
<td>8 (62%)</td>
</tr>
<tr>
<td></td>
<td>Operational Definition</td>
<td>27</td>
<td>18 (67%)</td>
</tr>
<tr>
<td><strong>Data Collection and Management</strong></td>
<td>Data Collection Process</td>
<td>13</td>
<td>8 (62%)</td>
</tr>
<tr>
<td></td>
<td>Training in and Embedding of Consistent Data Collection</td>
<td>5</td>
<td>2 (40%)</td>
</tr>
<tr>
<td></td>
<td>Database Design</td>
<td>4</td>
<td>3 (75%)</td>
</tr>
<tr>
<td></td>
<td>Outliers and Missing Data</td>
<td>3</td>
<td>2 (67%)</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>Planning the Analysis</td>
<td>16</td>
<td>9 (56%)</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>Planning for Action</td>
<td>4</td>
<td>4 (100%)</td>
</tr>
<tr>
<td><strong>Embedding</strong></td>
<td>Planning the Sustainability</td>
<td>9</td>
<td>8 (89%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>104</td>
<td>70 (67%)</td>
</tr>
</tbody>
</table>

*Table 1: Total number of questions per subsection, and % that reached the 75% consensus level at the end of the Delphi Survey*
Learn the tools, then internalize them.

Charlie Parker
“Designs that are better suited to the evaluation of clearly defined and static interventions may be adopted without giving sufficient attention to the challenges associated with the dynamic nature of improvement interventions and their interactions with contextual factors.”

The Methodology Problem

COPD Bundle: process and outcome

[Graph showing readmission rate and COPD care bundle compliance over time.]
Summary

• Statistical process control chart analysis is the right tool for the improvement job (analytic)
• It is just as possible to apply this tool rigorously as it is other statistical tools
• We don’t always do it rigorously at present, which impacts on perception of QI
• There are things we can use and do to improve! Planning, tools, transparency, ...
Thank you!
Additional slides
Overview

1. Introduction – CLAHRC NWL, Improvement Science Fellowship

2. CLAHRC NWL Approach to Measurement
   - AED, Process Mapping, Checkpoints
   - Web Improvement Support for Healthcare (WISH)
   - Evaluation

3. Improvement Science Fellowship work
   - Support for planning measurement
   - Support for executing measurement
   - Evaluation
   - SPC Publication Guidelines
Measurement in, and of, Improvement

• Are changes happening in the way health care is delivered?
• Are these associated with improvements for patients, carers & the public?
• Are these changes causally linked?
• How can we reproduce this improvement elsewhere?
Measuring the wrong things

“After careful consideration of all 437 charts, graphs, and metrics, I’ve decided to throw up my hands, hit the liquor store, and get snookered. Who’s with me?!”

“So things are good, stuff is OK, and I reiterate my request for more specific data.”
From ideas to measurement

Action Effect Diagrams provide ideas for measurement

Defined process for refining ideas into live measurement
Action Effect Diagram

Overall aim
To improve the health, quality of life and experience of care for patients from Hospital X who have been discharged following acute exacerbation of COPD

Self-management post-exacerbation
- Whether the patient smokes
- Correct use of inhalers out of hospital
- Fitness and exercise

Attendance at and engagement with smoking cessation programme
- Patient awareness of their condition and management plan

Factors proposed to be influenced by care bundle
- Referral and information exchange regarding smoking cessation if patient is eligible
- Quality of provision and patient experience of smoking cessation programme
- Education on and observation of correct inhaler technique
- Provision of correct inhaler device
- Information pack given to patient prior to discharge
- Information exchange regarding pulmonary rehabilitation assessment and potential benefits, referrals if appropriate
- Quality of provision and patient experience of pulmonary rehabilitation programme
- Arrangement of outpatient appointment for 4 weeks prior to discharge

Follow up care from GP

Implementation activities
- Design COPD Care Bundle (Rand/Boss)
- Staff education session(s) on COPD Bundle
- Word Champion
- Patient Video

Measure concepts
1. Length of stay
2. Readmission Rate
3. Patient Experience Survey
4. Two-week follow-up telephone survey
5. Number of smoking cessation attendances
6. Number of pulmonary rehabilitation attendances
7. Number of completed bundles
8. Number of smoking cessation referrals
9. Number of pulmonary rehabilitation referrals
10. % eligible patients on bundle
11. % eligible patients referred to pulmonary rehabilitation
12. Staff awareness survey
13. Proportion of people designated as inhaler technique providers who are competent providers

Key:
A = Oxford Centre for Evidence-Based Medicine scale
B = Level of evidence: Oxford Centre for Evidence-Based Medicine scale
C = Measurement concept:
D = Proportion of people designated as inhaler technique providers who are competent providers

Definition or documented evidence of cause/effect
No documented evidence of cause/effect

GP Engagement Events

Design of patient information materials
Specialist staff training on Pulmonary Rehabilitation and Referral
- Team in place
- Overall aim agreed
- Action-effect diagram
- Process map(s)

- Work in progress on definitions.
- Data sources

- Database operational
- Outcome measures agreed & access to data established

- Interventions ID, Stakeholder agreement
- Measure concepts and names

- Sign off definitions
- Initial data collection PDSAs completed
- Baseline data

- Improvement Cycles

- Ongoing evaluation
Implementing in the dark

COPD CARE BUNDLE - HOSPITAL X

Overall Compliance

- First 6 months: 80.2%
- Last 6 months: 94.8%
Weekly percentage of patients on Ward X clerked within 4 hours of arrival (weekdays in-hours)

Graph showing the percentage of patients clerked within 4 hours of arrival from 2012-04-02 to 2013-10-02.

- **Percentage clerked in 4h**
- **Average for period**
- **Upper Control Limit**
- **Lower Control Limit**
Monthly % ED Attendances Admitted; England 2011-2014

Average and Range Chart for Monthly Seasonal Factors

- Percentage ED attendances admitted
- Seasonalised Average
- Seasonalised Lower Process Limit
- Seasonalised Upper Process Limit

Month

Average and Range Chart for Monthly Seasonal Factors

Average Chart

Range Chart

Month
Web Improvement Support for Healthcare

Plan Do Study Act cycles

Comments - context

Quantitative measure and SPC

Journal of Biomedical Informatics
Volume 52, December 2014, Pages 151-162
Special Section: Methods in Clinical Research Informatics

Model-driven approach to data collection and reporting for quality improvement
Vasa Curcin*, Thomas Woodcock*, Alan J. Poots*, Azeem Majeed*, Derek Bell*

doi:10.1016/j.jbi.2014.04.014

Open Access funded by Department of Health UK

Open Access

Highlights
• Addressing the challenge of the second translational gap is key to improving healthcare processes.
• Data-driven methodologies improve likelihood of success.
• We propose the Improvement Data Model (IDM) for data collection and reporting for local improvement.
• WISH, a prototype software tool based on IDM is used by over 600 users in 50+ improvement projects.
COPD Care Bundle Outcome Evaluation

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0116187

<table>
<thead>
<tr>
<th></th>
<th>7 day readmissions</th>
<th>28 day readmissions</th>
<th>90 day readmissions</th>
<th>Number of bed-days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual number for London COPD admissions, 2002 - 2012</td>
<td>454.1 (56.3)</td>
<td>1,353.1 (175.1)</td>
<td>2,472.2 (248.4)</td>
<td>63,277.9 (12,281.9)</td>
</tr>
<tr>
<td>Mean annual number for bundle COPD admissions, 2002 - 2012</td>
<td>272.2 (69.7)</td>
<td>727.3 (163.9)</td>
<td>1,335.5 (284.4)</td>
<td>38,021.2 (4,528.9)</td>
</tr>
<tr>
<td>Annual trend in London readmissions pre-implementation ¹</td>
<td>+1.4% (0.063)</td>
<td>+0.6% (0.173)</td>
<td>+0.2% (0.522)</td>
<td>-1.7 (&lt;0.001)</td>
</tr>
<tr>
<td>Annual trend in bundle readmissions pre-implementation ²</td>
<td>+1.9% (0.700)</td>
<td>+1.5% (0.238)</td>
<td>+1.0% (0.173)</td>
<td>-1.0 (&lt;0.001)</td>
</tr>
<tr>
<td>Annual trend in London readmissions post-implementation ²</td>
<td>-4.4% (0.058)</td>
<td>+4.8% (0.025)</td>
<td>+2.7% (0.105)</td>
<td>-1.2 (0.666)</td>
</tr>
<tr>
<td>Annual trend in bundle readmissions post-implementation ³</td>
<td>-6.9% (0.487)</td>
<td>-3.8% (0.001)</td>
<td>-0.5% (0.093)</td>
<td>-1.4 (0.671)</td>
</tr>
<tr>
<td>Effect size would need for p≤0.05</td>
<td>-19.3%</td>
<td>-2.2%</td>
<td>-0.6%</td>
<td>-6.5</td>
</tr>
</tbody>
</table>

Table S2: Bundle trusts vs. other London trusts for COPD admissions, using ICD-10 codes J40-44

¹ P-value refers to difference of this trend from zero
² P-values refer to difference between these trends and the trend in London comparison trusts
³ P-value refers to difference between this trend and trend in London comparison trusts, adjusted for baseline trends
Measurement Planning Assessment Framework

- Design
- Collection and Management
- Analysis
- Action
- Sustainability
Measurement Planning Assessment
Examples

Is it clear how the measures are linked to the aim?

Will there be quality assurance reviews of data entry?

Are specific statistical methods outlined in the plan?

Who will receive reports/review the measures regularly?

Is any aspect of the measurement process dependent on an individual?
Measurement for Improvement Assessment

• Framework to assess how well a team is using data to inform implementation
  E.g. Are the team picking up on special causes of variation? Are they then acting accordingly?
• Provides recommendations as to next steps
• Self assessment / observation
Measuring for improvement

Check your understanding

Tick all the statements you think are true, then Confirm to check your answers.

1. A common cause of variation dominates over other causes of variation.

2. If a process shows only common cause variation, no special cause variation, then it is said to be "in statistical control".

3. A process that is in statistical control is predictable, at least within limits.

4. If a process is in statistical control, I should look for high and low points on the chart and seek explanations of this variation.

5. If a process is out of statistical control, the only way to improve it is to redesign the whole process.

CONFIRM

ZOOM
4. Evaluation that hides variation

Evaluations by leading experts writing in top journals often select quantitative statistical approaches more suited to *enumerative* studies.

E.g. “Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation”

*BMJ*. 2011; 342: d195 10.1136/bmj.d195

“Recording of respiratory rate increased to a greater degree in SPI1 than in control hospitals; in the second six hours after admission recording increased from 40% (93) to 69% (165) in control hospitals and from 37% (141) to 78% (296) in SPI1 hospitals (odds ratio for “difference in difference” 2.1, 99% confidence interval 1.0 to 4.3; P=0.008)”
“Designs that are better suited to the evaluation of clearly defined and static interventions may be adopted without giving sufficient attention to the challenges associated with the dynamic nature of improvement interventions and their interactions with contextual factors.”

Evaluation Framework

Objective: Develop and apply an evaluation framework for improvement initiatives

– Retains advantages of statistical process control
– Facilitates sustainability and spread
Consolidated Framework For Implementation Research

“Adaptability relies on a definition of the 'core components' [...] versus the 'adaptable periphery' [...] often the distinction [...] can only be discerned through trial and error over time as the intervention is disseminated more widely and adapted for a variety of contexts”

Theory driven evaluation

- Map out the programme theory
- Research evaluation to test out that theory
- When, how, why intervention works?
- Unpick the complex relationship between context, content, application and outcomes
- Develop a necessarily contingent and situational understanding of effectiveness
- **Seek theoretical generalisability**

doi 10.1093/intqhc/mzm004
Safer Clinical Systems II

“The evaluation team sought to identify the theory (concepts, rationale and assumptions) behind the Safer Clinical Systems approach, to determine how far the approach helped the sites to make their systems more reliable, and to explain how the approach might work (the mechanisms of change), while also considering contextual factors.”

- Mixed-method longitudinal study design
- SPC – did the sites made systems more reliable?
- Combining with qualitative findings gives a richer picture

- Identified need to improve skills and processes relating to measurement of quality and safety in the NHS
COPD Bundle: process and outcome

**Graph:**
- **X-axis:** Years relative to bundle implementation
- **Y-axis:** Readmission rate (%)
- **Legend:**
  - Medical
  - COPD

**COPD Care Bundle - Hospital X Overall Compliance**
- **Y-axis:** Percentage Compliance
- **Graph Elements:**
  - Percentage Compliance
  - Average
  - Lower Control Limit

**Data Points:**
- Values indicating compliance over time from January 2002 to December 2002.
Context, adaptability and reproducibility
Theory-driven SPC evaluation
Theory-driven SPC evaluation
Theory-driven SPC evaluation
Testing a theory

Assume each measure either: improves, remains unchanged, deteriorates (SPC)
With 2 factors there are 9 possible scenarios.
E.g.

This result would not support the hypothesis represented by the diagram – investigating why should promote learning
With one outcome measure, and n process measures hypothesized to influence it, and two time periods only; $3^{n+1}$ possible “results”.
Connections to explore

- Stepped-wedge designs
- Factorial designs (cf. Quality Improvement Through Planned Experimentation; Roen, Nolan, Provost)
- Structural equation modelling
- (Dynamic) Bayesian networks

Can these approaches be deployed to harness the AED in a theory driven evaluation that retains the analytic benefits of statistical process control?
Guidelines for SPC reporting

Standardised presentation and specification of minimum accompanying information for publication of control charts

For example:

• Specify the type of chart used (e.g. p-chart, Xmr chart)
• Specify precisely which set of “rules” were chosen in advance (cf. specifying significance level for a hypothesis test)
Discussion

1. Within your network, how do you decide and communicate what parts of interventions must be the same across sites, and what can vary?

2. How do you ensure data from across different sites is consistent and comparable?

3. How do you decide what constitutes improvement: within sites, within networks, between networks?

4. How do you decide what should be spread across the network, and beyond?

What helps and what hinders you in doing these things?
Summary

Some challenges in improvement work:
• Measure the right things
• Use high-quality data
• Learn from data during implementation: best decision making
• Conduct evaluations that are useful for sustainability and spread

Some Interventions:
• Action Effect Diagram
• Process for measure development
• Measurement Planning Assessment Framework
• WISH + Measurement Assessment Tool
• SPC reporting guidelines
• Evaluation framework