



eGEMs

Generating Evidence & Methods
to improve patient outcomes

Analytical Methods for a Learning Health System: 4. Delivery System Science

Michael Stoto, PhD;ⁱ Gareth Parry, MSc, PhD;ⁱⁱ Lucy Savitz, PhD, MBAⁱⁱⁱ

ABSTRACT

The last in a series of four papers on how learning health systems can use routinely collected electronic health data (EHD) to advance knowledge and support continuous learning, this review describes how delivery system science provides a systematic means to answer questions that arise in translating complex interventions to other practice settings. When the focus is on translation and spread of innovations, the questions are different than in evaluative research. Causal inference is not the main issue, but rather one must ask: How and why does the intervention work? What works for whom and in what contexts? How can a model be amended to work in new settings? In these settings, organizational factors and design, infrastructure, policies, and payment mechanisms all influence an intervention's success, so a theory-driven formative evaluation approach that considers the full path of the intervention from activities to engage participants and change how they act to the expected changes in clinical processes and outcomes is needed. This requires a scientific approach to quality improvement that is characterized by a basis in theory; iterative testing; clear, measurable process and outcomes goals; appropriate analytic methods; and documented results.

To better answer the questions that arise in delivery system science, this paper introduces a number of standard qualitative research approaches that can be applied in a learning health system: Pawson and Tilley's "realist evaluation," theory-based evaluation approaches, mixed-methods and case study research approaches, and the "positive deviance" approach.

Introduction

A learning health system is one in which “science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the care process, patients and families active participants in all elements, and new knowledge captured as an integral by-product of the care experience”¹ According to Fineberg,² these systems are committed to continuous improvement and as such are engaged in discovery, innovation, and implementation, which requires embedding research into the process of care. The last in a series of four papers on how learning health systems can use routinely collected electronic health data (EHD) to advance knowledge and support continuous learning (see Box 1), this review describes how

delivery system science provides a systematic means to answering questions about both organizational factors and design, infrastructure, policies, and payment mechanisms. This requires a scientific approach to quality improvement with the following attributes: clear, measurable process and outcomes goals; a basis in evidence; iterative testing; appropriate analytic methods; and documented results.³

In many health care organizations working on improving quality, the Institute for Healthcare Improvement’s (IHI) “Model for Improvement” (Figure 1) is one guiding approach. This approach, developed by Associates in Process Improvement,⁷ focuses on the analysis of existing systems to identify problems and possible changes to improve care, and on the development of indicators and

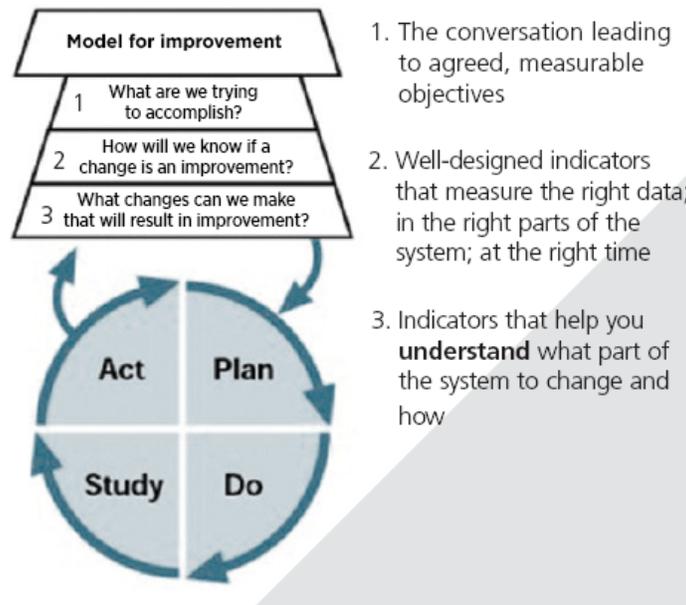
Box 1. Series on Analytic Methods to Improve the Use of Electronic Health Data in a Learning Health System

This is one of four papers in a series of papers intended to (1) illustrate how existing electronic health data (EHD) data can be used to improve performance in learning health systems, (2) describe how to frame research questions to use EHD most effectively, and (3) determine the basic elements of study design and analytical methods that can help to ensure rigorous results in this setting.

- Paper 1, “Framing the Research Question,”⁴ focuses on clarifying the research question, including whether assessment of a causal relationship is necessary; why the randomized clinical trial (RCT) is regarded as the gold standard for assessing causal relationships, and how these conditions can be addressed in observational studies.
- Paper 2, “Design of observational studies,”⁵ addresses how study design approaches, including choosing appropriate data sources, methods for design and analysis of natural and quasi-experiments, and the use of logic models, can be used to reduce threats to validity in assessing whether interventions improve outcomes of interest.
- Paper 3, “Analysis of observational studies,”⁶ describe how analytical methods for individual-level electronic health data EHD, including regression approaches, interrupted time series (ITS) analyses, instrumental variables, and propensity score methods, can be used to better assess whether interventions improve outcomes of interest.
- Paper 4, this paper, addresses translation and spread of innovations, where a different set of questions comes into play: How and why does the intervention work? How can a model be amended or transported to work in new settings? In these settings, causal inference is not the main issue, so a range of quantitative, qualitative, and mixed research designs are needed.



Figure 1. The Model for Improvement



Source: Langley and colleagues.⁷

performance measures that help practitioners understand what part of the system to change and how. For instance, Psek and colleagues⁸ describe the implementation of learning health system principles in the Geisinger Health System, identifying “evaluation and methodology” (activities and methodological approaches needed to identify, implement, measure, and disseminate learning initiatives) as one of nine learning health care system framework components. In a similar description of Kaiser Permanente’s approach to a learning health system, Schilling and colleagues⁹ identify “real-time sharing of meaningful performance data” as one of six building blocks.

In many systems, however, evaluation of innovations tends to straightforward, using tools such as control charts to monitor process and less frequently outcomes measures, often without randomization or even control groups. The study design and analytical

methods described in this paper provide a way to enhance the rigor of such evaluations. However, when there is evidence that the innovation can work, and the question is one of translation and spread, a different set of questions comes into play: How and why does the intervention work? What works for whom and in what contexts? How can a model be amended to work in new settings? In these questions, causal inference is not the main issue, so a broader range of quantitative, qualitative, and mixed methods designs are needed.

We begin this paper by describing the application of delivery system science in two different settings. We then describe how the research questions – and the relevant analytical tools – vary as evaluations move through innovation, testing, and improvement phases. We conclude with a discussion of the qualitative and mixed-methods research designs that are needed in these settings.

Examples: Delivery System Science in Learning Health Systems

In order to motivate the approaches discussed in this paper, we begin with two examples of delivery system science approaches to evaluation. The first example, from Intermountain Healthcare, illustrates the evaluation of a specific intervention – team-based care – in one delivery system. The second example discussed the general approach to evaluation adopted by the Center for Medicare and Medicaid Innovation (CMMI) for a wide variety of system-level interventions.

Evaluation of Team-Based Care at Intermountain Health

Delivery systems are increasingly interested in understanding what works as they pilot, take to scale, and then externally spread an intervention. Consider a recent, unsponsored study conducted at Intermountain Healthcare. With a \$24 million investment in the system's phased implementation of the medical home model with team-based care (TBC), leadership was interested in exploring the degree to which the complex intervention worked. The primary research aim was: "Do clinics with high performing, team-based care provider greater value compared to other clinics operating under a more traditional patient management approach—as measured by quality/clinical outcomes, cost, utilization, patient/family service, and staff outcomes?" Team-based care was defined as those clinics that had routinized the medical home and mental health integration model. Patient cohorts were carefully identified based on age (adults, >18 years) and stability of relationship within the system. With over 185 clinics, four categories emerged: no TBC (n=171,912 patient years); planning TBC (23,164 patient years); adoption of TBC (n=155,486 patient years); and routinized TBC (n=163,226 patient years). Rogers¹⁰ theory on stage of adoption was

used to classify clinics based on interview data and the Pawson and Tiley¹¹ model for realist evaluation was applied in the retrospective, longitudinal evaluation. Analysis focused on internal delivery system together with owned health plan data, limiting the patient cohort to those covered by SelectHealth insurance such that total cost of care in this case are assessed in terms of allowable charges or reimbursement by the owned insurer. Generalized estimation equations (GEE) were used to estimate the marginal effects on the population and assessment of correlational matrix. Qualitative interview data with patients/family members and staff were used to complement quantitative analyses using an embedded mixed methods approach. Positive results that documented the success of the TBC pilot, provided trusted information for evidence-based decision making on the part of system leadership to support taking the intervention to scale.¹²

CMMI's Approach to Rapid-cycle Evaluation

The Affordable Care Act established CMMI to test innovative payment and service delivery models. These models require structural changes in care delivery that challenge traditional assumptions. In particular, providers and health care organizations participating in patient-centered medical homes or Accountable Care Organizations must adapt their practices to deliver on these new models in meeting the Triple Aim of improving the experience of care, improving the health of populations, and reducing per capita costs of health care.¹³

The plans for CMMI's new, rapid-cycle approach to evaluation, which aims to deliver frequent feedback to providers in support of continuous quality improvement, while rigorously evaluating the outcomes of each model tested,¹⁴ provides a good example of delivery system science at its best. CMMI's Learning System is a systematic



improvement framework designed to understand change across a mix of health systems and increase the likelihood of successful tests of change in payment policy and delivery of care. Using terms such as theory of change, context, and spread that will be explained later in this paper, the CMMI Learning System consists of seven major steps:

1. Establish clear aims
2. Develop an explicit theory of change
3. Create the context necessary for a test of the model
4. Develop the change strategy
5. Test the changes
6. Measure progress toward aim
7. Plan for spread

In this context, the evaluators are part of the solution by using CMS claims data to promote and support continuous quality improvement in the participating sites. This means that they must not only assess results of retrospective claims-based analysis, but also understand the context through site visits and surveys of involved organizations. CMMI, therefore, evaluates innovations regularly and frequently after implementation, allowing both rapid identification of opportunities for course correction and timely action on that information.¹⁵

Implications

CMMI's approach has not been without criticism. One challenge is that randomization is either infeasible (owing to voluntary participation, data-collection challenges, or multimodal interventions) or inappropriate (beneficiaries would receive different levels of care).^{16,17} Critics argue that without randomized trials it's unlikely to be able to provide decisive data on whether its largest quality-improvement programs are effective.¹⁸ To address this, Shrank¹⁹ notes that advanced statistical techniques to account for potential confounding related to providers' characteristics that might

influence outcomes independent of the intervention. Moreover, the providers who choose to participate in CMMI's models, and the populations they serve, may differ from nonparticipants. Evaluation, therefore, requires appropriate selection of comparison groups, taking environmental and policy characteristics into account, as well as statistical analyses such as propensity scores to control for bias and clarify specific causal mechanisms similar to those described in the second²⁰ and third²¹ papers in this series.

Furthermore, it is important to distinguish between evaluations of national programs such as the Partnership for Patients Program and the implementation of these programs at individual hospitals or other sites. The team-based care evaluation at Intermountain Health falls into the latter category. The criticisms of CMMI's evaluation strategy relate to assessing national program's overall impact. The focus of this paper, however, is on methods for translating and adapting methods for which there is already some evidence of efficacy to new settings. Once an intervention has been shown to work in some locations, evaluation in new settings is less a question of causal inference as one of translation and spread. Harking back to the first paper in this series,²² the question is not so much whether the model works, but how it can be amended to work in new settings.

Even with a local focus, translating complex, multilayered interventions to other practice settings is far more difficult than spreading simple interventions. Both the interventions and the contexts in which they are deployed are complex and difficult to control through normal means. Effective evaluation in this situation requires judgment. Deciding what works where requires understanding the key characteristics of the settings, organizations, and environments where the intervention has been tested, the timing of

interventions and measurement of intermediate and long-term outcomes, how outcomes differ across the testing sites, and how the contexts in these sites compare to a broader universe where the intervention (or an adapted version of it) might be implemented. Researchers need to capture essential dimensions of variation to tease out what they imply about the robustness of the intervention components (which things are critical and which can be modified) and the kinds of settings and environments where they work. A sound and carefully articulated theory of change and qualitative research methods to understand the details of context and implementation is also needed.

Evaluation of Health Care Improvement Initiatives

Addressing some of the criticisms of the CMMI approach, Parry and colleagues²³ write that fixed protocol RCTs are ill-suited for health care quality improvement (QI) initiatives, which are complex and context sensitive and iterative in nature, and vary depending on innovative stage (i.e., initial stage, more developed testing stage, or the wider “spread” stage). The question that must be addressed is not “Does it work?” but rather “How and in what contexts does the new model work or can be amended to work?” This question requires a theory-driven formative evaluation approach that considers the full path of the intervention from activities to engage participants and change how they act to the expected changes in clinical processes and outcomes.

Figure 2 illustrates schematically why new improvement ideas so often fail. Ideas associated with a low degree of belief are likely to require thorough exploration and amendment in a small number of settings before they are considered ready for wider testing or spread in more settings. Because they are tested in only a few settings, the full range of complexity of the innovation may not

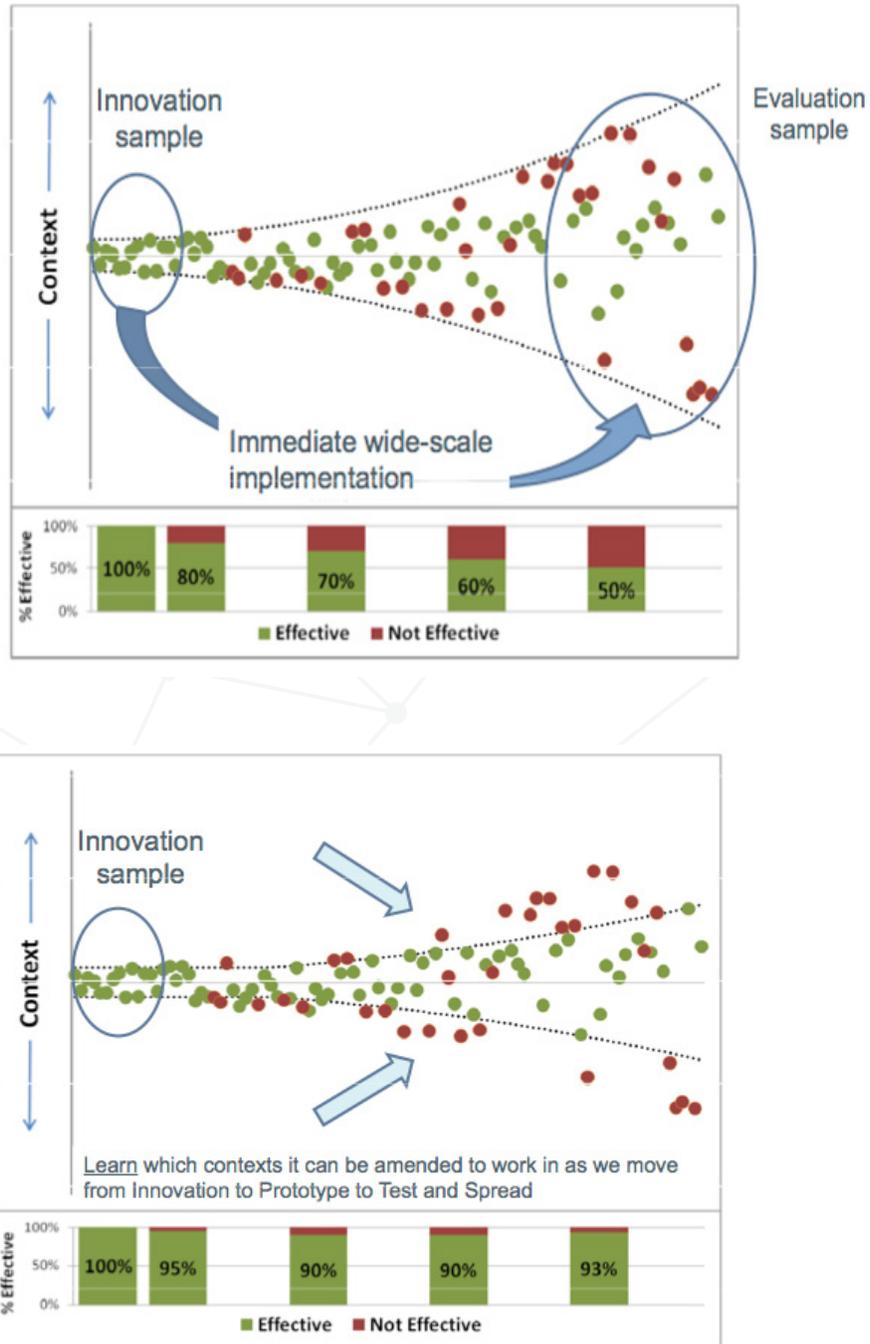
be fully understood, yet a simple summary model of the change is formed into a fixed study protocol. Replicating the intervention in organizations similar in context may be effective in 80 percent of organizations, but after further attempts to replicate in an increasingly broad range of contexts, the intervention may work in only 50 percent of these organizations. For example, in early studies, surgical site checklists were associated with improved patient outcomes.²⁴ However, a recent study in 101 acute hospitals in Ontario, Canada, found no significantly improved outcomes.²⁵

Alternatively, Parry and colleagues²⁶ propose a strategy for amending models that has greater potential to enable spread to other organizations. This calls for evaluation that is theory driven and formative. Evaluators must understand the core concepts that underpin the detailed tasks undertaken as part of a new model because these core concepts are more likely to be generalizable to other organizations than the detailed tasks.²⁷ Then as organizations introduce a new model, they should start with the core concepts using examples of what has worked in other settings, test approaches for introducing detailed tasks tailored to their local context. For example, in the checklist literature, there is emerging evidence that the activities associated with forming a multi-disciplinary team, and reviewing the evidence that will form a checklist at a local is an important core concept.²⁸

In this approach, the guiding evaluative question for health care improvement is “How and in what contexts does the new model work or can be amended to work?” Addressing this question requires a program theory, or a chain of reasoning from the activities involved in an initiative through to the change in processes and outcomes expected. The program theory may also be broken down into an activity-focused execution theory and a clinical-focused content theory. Execution theory is defined



Figure 2. Why New Improvement Ideas Fail so Often



Source: Parry and colleagues.²³

as the rationale for how the experience provided by the improvement initiative, the implementation, and the learning accomplished leads to improvement in the process measures. Content theory is defined as the rationale for how improvement in process measures associated with applying the new model leads to improvement in organizational performance or patient outcomes.

A logic model provides a framework for clarifying and illustrating how the activities and inputs associated with a project lead to short-term process improvement (execution theory) and then on to mid- and long-term outcome improvement (content theory). For example, the execution theory may illustrate how the activities associated with a 'collaborative' are predicted to result in a new model being implemented, and the content theory provides an illustration of how the new model is likely to impact on patient or organizational outcomes. The program theory guides evaluators in developing appropriate research questions and data collection methods, using both qualitative and quantitative methods. Qualitative data indicate what improvement teams are doing and why, where they meet barriers or facilitators to change, and the contexts where success is or is not being achieved and why. Quantitative data illustrate progress toward the goals. These results, if fed back to the implementation or improvement teams, as the program roles out can be used to update the initial program theories.

As summarized in Table 1, Parry and colleagues²⁹ describe such an evaluation strategy, accounting for the phase of improvement as well as the context and setting in which the improvement concept is being deployed. In the innovation phase, the goal is to discover a new model of care based on evidence from a small number of settings, perhaps one or two teaching hospitals. The aim of the evaluation is to estimate the improvement achieved and update the

degree of belief that the content theory will apply in similar contexts. In this phase, evaluation approaches may include: (1) quantitative measurement systems to estimate the impact of variations of the content theory; (2) qualitative interviews with model developers and those who have tested the model to describe the underlying concepts and indicate how they impact the results obtained; and (3) regular, rapid-cycle feedback of the findings to the leads of the innovation phase. This approach, for example, may lead to an understanding of the core components of a surgical-checklist, including an understanding of the pre-implementation multi-disciplinary evidence review and team development described above.

In the testing phase, the goal is to engage organizations and enable them to test whether a model works or can be amended to work in their context. For example, a broader range of teaching hospitals and some large general hospitals might be involved. The aims of the evaluation are to describe an amended content theory, to estimate the improvement achieved from applying the amended content theory in specific contexts, and the degree of belief that the amended content theory will apply in specific contexts. Another aim is to describe an amended theory for engaging with organizations in specific contexts to test and amend the new content theory, and to estimate the likely application of testing and amendment of content theory. In addition to those needed for the innovation phase, the testing phase calls for additional evaluation approaches such as randomized and observational studies, as well as qualitative interviews to identify how teams did or did not learn and apply their learning. This approach, can lead to a better understanding of the likelihood a surgical checklist will lead to improved outcomes with specific contexts such as orthopedic surgery, emergency surgery, and within training hospitals.



Table 1. Summary of Evaluation Aims and Approaches by Improvement Phase

PHASE	INNOVATION	TESTING	SCALE-UP AND SPREAD
<i>Aim of the improvement phase</i>	Generate or discover a new model of care with evidence of improvement in small number of settings	Engage organizations and enable them to test whether a model works or can be amended to work in their context	Engage organizations to adopt models with a high degree of belief in applicability and impact in a broad range of contexts
<i>Aim of the evaluation</i>	<p>From a small group of organizations with limited context to:</p> <ul style="list-style-type: none"> • describe a new content theory • estimate the improvement achieved from applying the theory • update degree of belief that the content theory will apply in similar contexts from where it was developed 	<p>From an initial content theory, with moderate degree of belief,</p> <ul style="list-style-type: none"> • describe an amended content theory • estimate the improvement achieved from applying the amended content theory in specific contexts • update degree of belief that the amended content theory will apply in specific contexts <p>From an initial execution theory,</p> <ul style="list-style-type: none"> • describe amended theory for engaging organizations in specific contexts to test and amend new content theory • estimate the likely application of testing and amendment of content theory in the future 	<p>From an initial content theory, with high degree of belief that it will apply in specific contexts, to</p> <ul style="list-style-type: none"> • Describe any amendments identified in the spread phase <p>From an initial execution theory,</p> <ul style="list-style-type: none"> • describe amended theory for engaging organizations in specific contexts to test and amend the new theory • estimate the likely application of testing and amendment of content theory in the future

Source: Parry and colleagues.²³

Table 1. Summary of Evaluation Aims and Approaches by Improvement Phase (Cont'd)

PHASE	INNOVATION	TESTING	SCALE-UP AND SPREAD
<i>Evaluation approaches</i>	<ul style="list-style-type: none"> • Quantitative measurement system to estimate the impact of variations in the development of the content theory • Clarification of content theory <ul style="list-style-type: none"> - through qualitative interviews with model developers and those who have tested the model - to draw out the underlying concepts, describe them and indicate how they impact on the results obtained <p>Regular, rapid-cycle feedback of the findings to the leads of the innovation phase</p>	<ul style="list-style-type: none"> • Quantitative measurement system to estimate the impact of amendments to execution and content theories • Longitudinal quantitative data analysis, including control chart and interrupted time series methods, to provide an estimate of the improvement associated with amended content and execution theories • Randomized cluster and stepped-wedge trials • Recommendations for how to amend content and execution theories through qualitative methods to identify how teams did or did not learn and apply their learning in their local context • Regular, rapid-cycle feedback of the findings to the leads of the testing phase 	<ul style="list-style-type: none"> • Measurement system to provide estimates of the impact of amendments to the execution theory • Randomized cluster and stepped-wedge designs • Longitudinal quantitative data analysis, including control chart and interrupted time series methods, to provide an estimate of the improvement associated with an amended execution theory • Recommendations for how to amend execution theory and point to issues with the content theory through qualitative methods to identify how teams did or did not learn and apply their learning, in their local context • Regular, rapid-cycle feedback of the findings to the leads of the scale-up and spread phase



In the scale-up and spread phase the goal is to engage organizations to adopt in a broad range of contexts such as all hospitals in a state or country. The aims of the evaluation are to describe an amended theory for engaging with organizations in specific contexts to content theory and to estimate its effects. This calls for all of the approaches used in the testing phase, as well as adaptive research designs allowing researchers to amend the execution theory and point to issues with the content theory through qualitative methods to identify how teams did or did not learn and apply their learning, in their local context. This approach can lead to a better understanding of how for example a surgical checklist can be implemented. For example, what, if any educational support is required, and what tools can be provided to adapt and tailor a checklist to a local setting.

The evaluation approach described by Parry and colleagues³⁰ uses data in three different ways. First, quantitative measurement systems are needed to estimate the impact of variations of the content theory in different settings. The methods for EHD described in the second³¹ and third³² papers in this series can be used to generate this information. Individual- and group-randomized studies that go beyond the observational methods discussed in this series may also be necessary, but EHD can be used in the conduct of such “real world” studies. Second, this evaluation approach requires qualitative interviews with model developers and those who have tested the model. Qualitative data of this sort are *not* likely to be available in most existing electronic data systems. This evaluation approaches requires rapid-cycle feedback of the findings to project leads so that changes can be made as necessary. The requirement for rapid feedback means that performance measures from existing EHD systems are necessary not only to evaluate impact of the innovation, but also as an integral part of the innovation.

Qualitative and Mixed Research Methods

Qualitative methods used in replicated case study designs offer an in-depth understanding of those factors operating in a particular place, setting, program, and/or intervention.³³ Qualitative research methods can also help probe how and why things happen by exploring how causal mechanisms are triggered in varying contexts. Thus, qualitative methods can be a useful complement to quantitative approaches, whose strength lies in identifying patterns of variation in and covariation among variables. Yet, qualitative methods are often subject to justifiable criticism as insufficiently rigorous and transparent. Fortunately, a well-established body of social science methods addresses this criticism. For example, drawing on discussions at an international symposium on health policy and systems research, Gilson and colleagues³⁴ summarize a series of concrete processes for ensuring rigor in case study and qualitative data collection and analysis (Table 2). With a focus on health systems rather than individuals, Yin’s³⁵ classic book on case study methods, now in its fifth edition, also provides relevant guidance.

The first item on Gilson and colleagues’³⁶ list – use of theory – deserves special attention in qualitative and mixed methods studies. This refers to a family of theory-oriented evaluation methods that use program theory to guide questions and data gathering and focus on explicating the theories or models that underlie programs, elaborating causal chains and mechanisms, and conceptualizing the social processes implicated in the program’s outcomes.³⁷

Patton³⁸ has led the way in qualitative evaluation. However, another well-known member of this family of theory-oriented evaluation methods called “realist evaluation” was developed by Pawson and Tilley³⁹ and introduced to health services research

Table 2. Processes for Ensuring Rigor in Case Study and Qualitative Data Collection and Analysis

Use of theory. Theory is essential to guide sample selection, data collection, analysis, and interpretive analysis.

Prolonged engagement with the subject of inquiry. Even though ethnographers may spend years in the field, health policy and systems research tends to draw on lengthy and perhaps repeated interviews with respondents and/or days and weeks of engagement at a case study site.

Case selection. Purposive selection allows earlier theory and initial assumptions to be tested and permits an examination of “average” or unusual experience.

Sampling. It is essential to consider possible factors that might influence the behavior of the people in the sample and ensure that the initial sample draws extensively across people, places, and time. Researchers need to gather views from a wide range of perspectives and respondents and not allow one viewpoint to dominate.

Multiple methods. For each case study site, best practice calls for carrying out two sets of formal interviews with all sampled staff, patients, facility supervisors, and area managers and conducting observations and informal discussions.

Triangulation. Patterns of convergence and divergence may emerge by comparing results with theory in terms of sources of evidence (e.g., across interviewees and between interview and other data), various researchers’ strategies, and methodological approaches.

Negative case analysis. It is advisable to search for evidence that contradicts explanations and theory and then refine the analysis accordingly.

Peer debriefing and support. Other researchers should be involved in a review of findings and reports.

Respondent validation. Respondents should review all findings and reports.

Clear report of methods of data collection and analysis (audit trail). A full record of activities provides others with a complete account of how methods evolved.

Source: Gilson and colleagues.³⁴



in the United States by Berwick.⁴⁰ Realist evaluation places the focus of research and evaluation less on relationships among variables and more on an exploration of the causal mechanisms that generate outcomes. The approach also recognizes that mechanisms that succeed in one context may not succeed in others, as summarized in Figure 3a. Figure 3b illustrates how evidence from individual evaluations can be synthesized over time. The reviewer's basic task is to sift through the mixed fortunes of the program (both solid and dashed lines) attempting to discover those contexts (C+) that have produced solid and successful outcomes (O+) from those contexts (C-) that induce failure (O-).⁴¹ Furthermore, Pawson and Tilley take the view that in the social world causality operates through the perceptions, incentives, capacities, and perspectives of the individuals involved in the system of interest. Thus they argue that learning about causality requires direct observation of these causal mechanisms by interviewing participants and by using other qualitative research methods (e.g., focus groups).⁴²

As an example of this theory-based approach, consider the Keystone Initiative study by a collection of Michigan intensive care units (ICUs) of the impact of a simple checklist on hospital acquired infections. When the authors found that the median rate of venous catheter bloodstream infections (CVC-BSIs) at a typical ICU dropped from 2.7 per 1,000 patients to zero after three months, which was sustained for 15 months of follow-up,⁴³ the question was not whether the intervention was responsible. Rather, Dixon-Woods and colleagues⁴⁴ sought to explore how and why the program worked. They did so by developing an *ex post* theory by (1) identifying program leaders' initial theory of change and learning from running the program, (2) enhancing this with new information in the form of theoretical contributions from social scientists, and (3) synthesizing prior and new information to produce

an updated theory. They found that the Michigan project achieved its effects by (1) generating pressures among rival ICUs to join the program and conform to its requirements, (2) creating a densely networked community with strong horizontal links that exerted normative pressures on members, (3) reframing CVC-BSIs as a social problem and addressing it through a professional movement combining "grassroots" features with a vertically integrating program structure, (4) using several interventions that functioned in different ways to shape a culture of commitment to doing better in practice, and (5) harnessing standardized performance measures that clinicians believe are valid as a disciplinary force.

In mixed methods studies, investigators intentionally integrate or combine quantitative and qualitative data rather than keeping them separate in a way designed to maximize the strengths and minimize the weaknesses of each type of data. Doing so allows researchers to view problems from multiple perspectives to enhance and enrich the meaning of a singular perspective. Standards for best practices in mixed methods research continue to coalesce. A recent report commissioned by the Office of Behavioral and Social Sciences Research brings together existing recommendations and criteria via a review, with recommendations for how mixed methods studies ought to be conducted and how funding proposals should be reviewed.⁴⁵ The integration of quantitative and qualitative data can be achieved by merging results together in a discussion section of a study, such as reporting first the quantitative statistical results followed by qualitative quotes or themes that support or refute the quantitative results. This can help to contextualize information, to take a macro picture of a system (e.g., a hospital) and add in information about individuals (e.g., working at different levels in the hospital). The schema proposed by Zhang

Figure 3a. Generative Causation Approach to Realist Evaluation

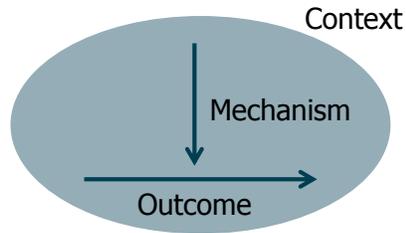
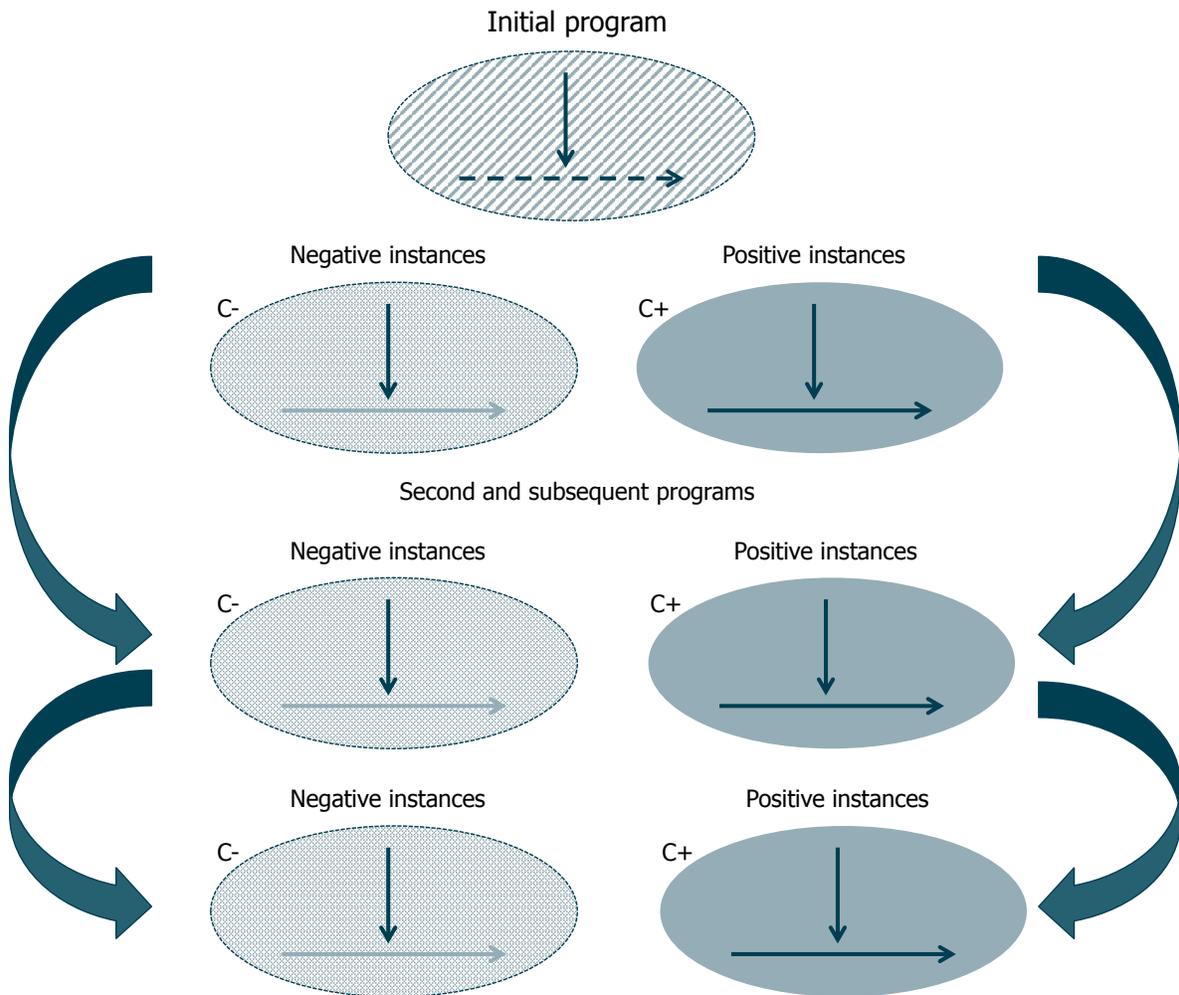


Figure 3b. Realist Synthesis Approach



Source: Pawson.⁴¹



and Creswell⁴⁶ provides categorizing strategies for mixing qualitative and quantitative data because these mixing strategies are critical to capitalizing on the value of mixed methods studies. They determine how the quantitative strengths of statistical validity, evidentiary deduction, and determination of associations and causality can be combined with the in-depth social and behavioral insight, contextual understanding, and theoretical grounding obtained through qualitative approaches. Briefly, connected mixed methods studies connect the qualitative and quantitative portions in such a way that one approach builds on the findings of the other approach. Integrated mixed methods studies analyze qualitative and quantitative data separately and integrate them during interpretation. Embedded mixed methods studies use data from one method (often qualitative) within another method (often quantitative portion).

The positive deviance approach⁴⁷ is another mixed methods technique that can be useful in learning health systems. This approach presumes that the knowledge about 'what works' is available in existing organizations that demonstrate consistently exceptional performance. As adapted for use in health care, this approach involves four steps: (1) identify 'positive deviants,' i.e., organizations that consistently demonstrate exceptionally high performance – compared to expectations – in the area of interest (e.g., proper medication use, timeliness of care); (2) study the organizations in-depth using qualitative methods to generate hypotheses about practices that allow organizations to achieve top performance; (3) test hypotheses statistically in larger, representative samples of organizations; and (4) work in partnership with key stakeholders, including potential adopters, to disseminate the evidence about newly characterized best practices. The approach is particularly appropriate in situations where organizations can

be ranked reliably based on valid performance measures, where there is substantial natural variation in performance within an industry, when openness about practices to achieve exceptional performance exists, and where there is an engaged constituency to promote uptake of discovered practices.

Conclusions

When the focus is on translation and spread of innovations, the questions are different than in evaluative research. Causal inference is not the main issue, but rather one must ask: How and why does the intervention work? What works for whom and in what contexts? How can a model be amended to work in new settings?

In these settings, organizational factors and design, infrastructure, policies, and payment mechanisms all influence an intervention's success, so a formative evaluation approach that considers the full path of the intervention from activities to engage participants and change how they act to the expected changes in clinical processes and outcomes is needed. Both the interventions and the contexts in which they are deployed are complex and difficult to control through normal means. Effective evaluation in this situation requires judgment. Researchers need to capture essential dimensions of variation to tease out what they imply about the robustness of the intervention components (which things are critical and which can be modified) and the kinds of settings and environments where they work. A sound and carefully articulated theory of change and qualitative research methods to understand the details of context and implementation is also needed.

Addressing these questions requires a scientific approach to quality improvement that is characterized by a basis in theory; iterative testing; clear, measurable process and outcomes goals; appropriate analytic methods; and documented

results. One approach, developed by Parry and colleagues⁴⁸ distinguishes between content theory (the rationale for how improvement in process measures associated with applying the new model leads to improvement in organizational performance or patient outcomes) and execution theory (the rationale for how an improvement initiative, the implementation, and the learning accomplished leads to improvement in the process measures). Proceeding through three phases (innovation, testing, and scale-up and spread), this approach identifies the improvement and evaluation aims at each phase along with the quantitative and qualitative evaluation approaches that are required.

This evaluation approach uses quantitative measurement systems to estimate the impact of variations of the content theory in different settings. The methods described in the second⁴⁹ and third⁵⁰ papers in this series can be used to generate this information from existing EHD. This evaluation also requires qualitative data from interviews with model developers and those who have tested the model that are *not* likely to be available in most existing electronic data systems. Finally, rapid-cycle feedback of the findings to project leads is required so that changes can be made as necessary. The requirement means that performance measures from existing EHD systems are an integral part of the innovation, not only an evaluation tool.

Because the focus is on how and why, delivery system science requires rigorous and qualitative and mixed-methods research techniques such as mixed-methods and case study research approaches, Pawson and Tilley's "realist evaluation" and other theory-based evaluation approaches, and the "positive deviance" approach. Good judgment and an in-depth understanding of the system in which the innovations are implemented are critical for applying these methods objectively and effectively.

References

1. Institute of Medicine, Best care at lower cost. Washington (DC): National Academy Press. 2012.
2. Fineberg H. Research and health policy: where the twain shall meet. Presentation at: AcademyHealth Annual Research Meeting; 2014 Jun 8-10; San Diego, CA.
3. Selker H, Grossman C, Adams A, Goldmann D, Dezii C, Meyer G, Roger V, Savitz L, Platt R. The common rule and continuous improvement in health care: a learning health system perspective [Internet]. Washington (DC): Institute of Medicine; 2011 Oct [cited 2014 Dec 17]. Available at: <http://www.iom.edu/-/media/Files/Perspectives-Files/2012/Discussion-Papers/VSRT-%20Common%20Rule.pdf>
4. Stoto MA, Oakes M, Stuart EA, Stuart L, Priest E, Zurovac J. Analytical methods for a learning health system: 1. Framing the research question. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):28.
5. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
6. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):30.
7. Langley GJ, Moen R, Nolan KM, Norman CL, Provost LP. *The improvement guide: a practical approach to enhancing organizational performance*. 2nd edition. San Francisco: Jossey-Bass Publishers; 2009. 512 p.
8. Psek WA, Stametz RA, Bailey-Davis LD, Davis D, Jonathan D, Faucett WA, Henninger DL, Sellers DC, Gerrity G. Operationalizing the Learning Health Care System in an Integrated Delivery System. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* 2015, 3(1), Article 6. Available at: <https://egems.academyhealth.org/articles/abstract/10.13063/2327-9214.1122/>
9. Schilling L, Dearing JW, Staley P, Harvey P, Fahey L, Kuruppu F. Kaiser Permanente's performance improvement system, part 4: creating a learning organization. *The Joint Commission Journal on Quality and Patient Safety* 2011, 36(12): 532-543.
10. Rogers EM. *Diffusion of innovations*. 5th edition. New York: Free Press; 2003. 576 p.
11. Pawson R, Tilley N. *Realistic evaluation*. 1st edition. London: SAGE Publications; 1997. 256 p.
12. Reiss-Brennan B, Brunisholz KD, Dredge C, Briot P, Grazier K, Wilcox A, Savitz L, James B. Association of integrated team-based care with healthcare quality, utilization, and cost, *JAMA* 2016, 316(8):826-834.
13. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. 2008 May-Jun; 27(3): 759-769.
14. Shrank W. The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Aff (Millwood)*. 2013 Apr; 32(4): 807-812.



15. Shrank W. The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Aff (Millwood)*. 2013 Apr; 32(4): 807-812.
16. Pronovost P, Jha AK. Did hospital engagement networks actually improve care? *N Engl J Med* 2014;371(8):691-693.
17. Rajkumar R, Press MJ, Conway PH. The CMS Innovation Center — a five-year self-assessment. *N Engl J Med*. 2015, 372(21): 1981-1983.
18. Casalino LP, Bishop TF. Symbol of health system transformation? assessing the CMS Innovation Center. *N Engl J Med*. 2015, 372(21): 1984-1985.
19. Shrank W. The Center for Medicare and Medicaid Innovation's blueprint for rapid-cycle evaluation of new care and payment models. *Health Aff (Millwood)*. 2013 Apr; 32(4): 807-812.
20. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
21. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):30.
22. Stoto MA, Oakes M, Stuart EA, Stuart L, Priest E, Zurovac J. Analytical methods for a learning health system: 1. Framing the research question. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):28.
23. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.
24. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, Herbosa T, Joseph S, Kibatala PL, Lapitan MC, Merry AF, Moorthy K, Reznick RK, Taylor B, Gawande AA, Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009 Jan 29; 360(5): 491-499.
25. Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med*. 2014 Mar 13; 370(11): 1029-1038.
26. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.
27. Overtveit J, Leviton L, Parry G. Increasing the generalizability of improvement research with an improvement replication programme. *BMJ Qual Saf*. 2011 Apr; 20(Suppl 1): i87-i91.
28. Leape LL. The checklist conundrum. *N Engl J Med*. 2014 Mar 13; 370(11): 1063-1064.
29. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.
30. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.
31. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
32. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):30.
33. Yin RK. *Case study research: design and methods*. 5th edition. Washington: SAGE Publications. 312 p.
34. Gilson L, Hanson K, Sheikh K, Agyepong IA, Ssengooba F, Bennett S. Building the field of health policy and systems research: social science matters. *PLoS Med*. 2011 Aug; 8(8): e1001079.
35. Yin RK. *Case study research: design and methods*. 5th edition. Washington: SAGE Publications. 312 p.
36. Gilson L, Hanson K, Sheikh K, Agyepong IA, Ssengooba F, Bennett S. Building the field of health policy and systems research: social science matters. *PLoS Med*. 2011 Aug; 8(8): e1001079.
37. Dixon-Woods M, Bosk CL, Aveling EL, et al. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q*. 2011;89:167-205.
38. Patton MQ. *Qualitative research & evaluation methods*. 3rd edition. Washington: SAGE Publications. 688 p.
39. Pawson R, Tilley N. *Realistic Evaluation*. London: Sage; 1997.
40. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. 2008 May-Jun; 27(3): 759-769.
41. Pawson R. Evidence-based policy: the promise of realist synthesis. *Evaluation July*; 2002 8(3):340-358.
42. Pawson R, Tilley N. *Realistic Evaluation*. London: Sage; 1997.
43. Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, Sexton B, Hyzy R, Welsh R, Roth G, Bander J, Kepros J, Goeschel C. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med*. 2006 Dec 28; 355(26): 2725-2732.
44. Dixon-Woods M, Bosk CL, Aveling EL, et al. Explaining Michigan: developing an ex post theory of a quality improvement program. *Milbank Q*. 2011;89:167-205.
45. Creswell JW, Klassen AC, Clark, VLP, Smith KC. Best practices of mixed methods research in the health sciences [Internet]. Bethesda (MD): National Institutes of Health; 2011 [cited 2014 Dec 17]. Available at: http://obssr.od.nih.gov/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf
46. Zhang W, Creswell J. The use of "mixing" procedure of mixed methods in health services research. *Med Care*. 2013 Aug; 51(8): e51-e57.
47. Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM. Research in action: using positive deviance to improve quality of health care. *Implement Sci*. 2009 May 8; 4: 25.
48. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.

49. Stoto MA, Oakes M, Stuart EA, Priest E, Savitz L. Analytical methods for a learning health system: 2. Design of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):29.
50. Stoto MA, Oakes M, Stuart EA, Brown R, Zurovac J, Priest E. Analytical methods for a learning health system: 3. Analysis of observational studies. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2016;5(1):30.
51. Parry GJ, Carson-Stevens A, Luff DF, McPherson ME, Goldmann DA. Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr*. 2013 Nov-Dec; 13(6 Suppl): S23-S30.