Implementing and Sustaining Healthcare Quality Improvement: Case Study Examining Feedback Structure and Dynamics

Andrada Tomoaia-Cotisel[1], Samuel Allen[1,2], Rod MacDonald[3], Jason Etchegaray[1], Eric Thomas[4]

1. The RAND Corporation
2. Worcester Polytechnic Institute, The Business School
3. James Madison University, College of Integrated Science and Engineering
4. Texas Medical Center

Abstract

Thinking of quality improvement (QI) as a simple negative feedback is insufficient. Due to challenges in sustaining improvements and concerns about importing a potentially-inappropriate paradigm, healthcare institutions develop custom QI approaches. We report on a case study (2015-2019) in one Neonatal Intensive Care Unit (NICU). This NICU was focused on developing and implementing a front-line oriented approach to QI training and project development: the unit-based approach (UBA). The successful implementation of UBA not only resulted in enhanced quality and remedied previous shortcomings but also showcased the efficacy of a front-line oriented approach to QI training and project development. We use mixed methods simulation modeling to provide comprehensive documentation of the intricate feedback dynamics within UBA and its intervention context. Furthermore, we present simulation scenarios which enable an improved understanding of what it takes to sustain QI. Through a visual comparative analysis with widely-used theories of health services improvement, our study highlights the significance of a concise feedback structure in unraveling dynamic patterns in complex contexts. Moreover, our findings emphasize the necessity of adopting innovative management and systems design approaches to ensure the long-term sustainability of the UBA's success.

Keywords

Quality improvement; context; neonatal intensive care unit (NICU); system dynamics; causal loop diagram

Introduction

Patient harms are an important problem in the neonatal intensive care unit (NICU), with preventable harms being due to variances in care delivery processes. As healthcare delivery becomes more complex due to patient conditions, complicated workflows, and available treatments (Grossmann, Goolsby et al. 2011), this complexity also impacts efforts designed to improve the delivery of such care. This article delves into the complexities of implementing QI approaches in healthcare and emphasizes the need for robust methods. To improve practical and theoretical understanding of efforts to improve care outcomes in complex healthcare contexts, we employ a mixed methods system dynamics case study in a NICU. Through a

rigorous examination of the interplay between operations and improvement, valuable insights will be gained to address the challenges and improve patient safety in healthcare.

Quality improvement (QI) efforts, one such approach to improving care, are typically undertaken via teams or committees and rely on coordinating improvements across multiple systems and practitioners within a healthcare setting (AHRQ 2022). Nevertheless, it appears that adopting a formal approach to QI does not necessarily lead to lasting improvements. For example, reviews of Lean and Lean-based methods of performance improvement in healthcare, whether aimed at efficiency, patient safety or both, have consistently expressed concern that only a small portion of improvements are reported to have been sustained over time, leading to repeated calls for more rigorous and critical evaluations in case studies (Vest and Gamm 2009, DelliFraine, Langabeer et al. 2010, Glasgow, Scott-Caziewell et al. 2010, J. Liberatore 2013, Al-Zuheri, Vlachos et al. 2021). Beyond Lean-based methods, recent reviews have noted contradictory findings about the impact of QI efforts in general on outcomes (Hill, Stephani et al. 2020).

Awareness of QI's failings is not new, and these failings are not specific to healthcare or to a specific QI approach. They have long been observed across industries, and researchers have worked for decades to improve formal approaches to improvement (c.f., (Weick 1993, Sitkin, Sutcliffe et al. 1994, Pfeffer and Sutton 2000, Repenning and Sterman 2001, Chassin and Loeb 2013)). Observing complementary strengths in formal methods, scholars have called for development of a "robust process improvement" approach that integrates valuable and complementary aspects of formal methods to account for physical and behavioral aspects of organizations (Repenning and Sterman 1997)(p.50). One such method is the Robust Process Improvement® (RPI) method pioneered for healthcare by The Joint Commission which thoughtfully integrates Change Management, Lean and Six Sigma (Adrian 2009, Chassin and Loeb 2013).

To build understanding of complexity and systems research in health, scholars have called for more case studies to be done using methods and concepts which take complexity into account (Rusoja, Haynie et al. 2018). These authors specifically suggest that by documenting their case study methods, insights about complexity can better transfer from theory to practice. We provide thorough documentation of our case study process in the methods sections (see also Appendix 1). We conducted research on efforts in a 118-bed Level II/III/IV NICU at one health system adopting RPI, in concert with other foundational elements, in developing and implementing a unit-based approach to QI efforts (herein, UBA). Such innovative UBAs adopt a systems approach to driving improvement by (1) explicitly acknowledging that there are a myriad of factors at various levels – individual, team, organizational, healthcare system, etc. – that impact safety by altering key feedbacks and (2) empowering front-line staff to lead QI efforts through formal training and other resource investments (*cf*.,(Sedlock, Ottosen et al. 2018)).

UBA is a complex intervention with components that interact with each other and with the organizational context where it is implemented. Scholars interested in safety in healthcare have called for research on ICU settings in order to improve understanding of complex interdependencies as well as for methodological innovations, including text analysis and pre-post data collection designs, to understand the impact of important difficult-to-measure variables (Tolk, Cantu et al. 2015). System dynamics (SD) is a method that leverages and

synthesizes qualitative and quantitative data to understand complex issues (Forrester 1980). Two complementary aspects of SD that are used to create a robust understanding of the system, including so-called soft variables, are developing qualitative diagrams (e.g., Causal Loop Diagrams (CLDs)) and running quantitative simulation models. Diagramming allows for the visualization of ways that multiple system factors are interrelated, and simulating allows one to test and improve the models that incorporate system parameters(Sterman 2000, Sterman 2006). Our study design iteratively cycles between diagramming and simulation and integrates qualitative and quantitative information, such as numeric operations data to develop insight into the dynamic consequences of the many factors that impact QI efforts in hospital settings. Here, we primarily report the feedback structures observed in the form of a CLD and briefly demonstrate policy analysis using our simulation model.

The adoption of RPI in the ICU context through UBA is a logical strategy; however, the anticipated challenges and complexities associated with the context and approach suggest that its implementation would not be straightforward. Scholarly research in the interdisciplinary field of implementation science, very often conducted in healthcare settings, indicates that the uptake of innovations is influenced by various contextual factors, not solely their effectiveness(Tomoaia-Cotisel, Scammon et al. 2013, Bauer and Kirchner 2020). The existing literature reveals significant gaps that need to be addressed. Despite the contributions of the Model for Understanding Success in Quality (MUSIQ) and Normalization Process Theory (NPT) to understanding implementation's complexity (May, Johnson et al. 2016, Reed, Kaplan et al. 2018), critical accounts suggest that that they may not adequately address the role of causal feedback loops, provide detailed causal mechanisms, handle overlapping constructs and causal loops effectively, and capture the temporal dynamics of implementation (Atkins, Lewin et al. 2011, Alharbi, Carlström et al. 2014, Alverbratt, Carlström et al. 2014, McEvoy, Ballini et al. 2014, May, Cummings et al. 2018, Reed, Kaplan et al. 2018, May, Albers et al. 2021). Addressing these gaps is crucial for advancing our understanding of implementation processes and improving the effectiveness of improvement initiatives in healthcare.

Although SD methods and concepts are frequently utilized in works on complexity in health (Rusoja, Haynie et al. 2018), a 2015 review of modeling studies in ICUs found a scarcity of SD studies, with only one identified (Bai, Fügener et al. 2018). Papers using that SD model (Demir, Lebcir et al. 2014, Lebcir and Atun 2020, Lebcir and Atun 2021) focused on understanding patient flows and explored interesting areas of medical decision-making including counter-intuitive reactions to policies to reduce length of stay, but failed to account for additional important issues, including quality of care and staffing. An SD study on ICUs study published more recently examined quality in the ICU, but only in terms of patient mortality, as it pertained to emergency patient flows in an ICU, and also failed to account for staffing (Mahmoudian-Dehkordi and Sadat 2017). More SD studies of ICU settings are needed to improve understanding of the role of complex interdependencies and soft variables specific to quality challenges in this context.

When SD has been applied in safety research across industries, many studies have suffered from a narrow focus (i.e., emphasizing a single cause for safety problems, usually an organizational issue) and researchers (Shire, Jun et al. 2018) have called on future research to consider multiple causes – including, for example, supervision, preconditions for unsafe acts, and/or unsafe acts – to better understand complex operational issues. In implementing the

UBA, an innovative reporting system was used which recorded the cause of the variance. Using this system in QI indicates that UBA aims to capture causes of safety problems holistically and to intentionally alter key feedbacks. Finally, another recent review noted that, despite a broad range of SD studies in health services and quality, articles failed to build on the long tradition of SD research across industries on service operations and formal QI implementation (Darabi and Hosseinichimeh 2020).

In this study, we present the implementation of an innovative approach to RPI (UBA) that aims to address the challenges experienced with formal quality improvement (QI) in healthcare. Specifically, we focus on a care setting characterized by a high degree of dynamic complexity, where innovations were implemented to prevent and overcome previous obstacles experienced in that setting and documented in the literature. To thoroughly investigate the causes of previous challenges and understand how the UBA and the NICU's management address them, we employ a rigorous pre and post evaluation design. This design allows us to uncover valuable insights into common challenges experienced in formal efforts to improve care delivery processes to make them safer for patients. Furthermore, we conduct a comparative analysis of our findings with related theories. This comparison provides a broader context and enhances the understanding of our study's contributions to the existing body of knowledge.

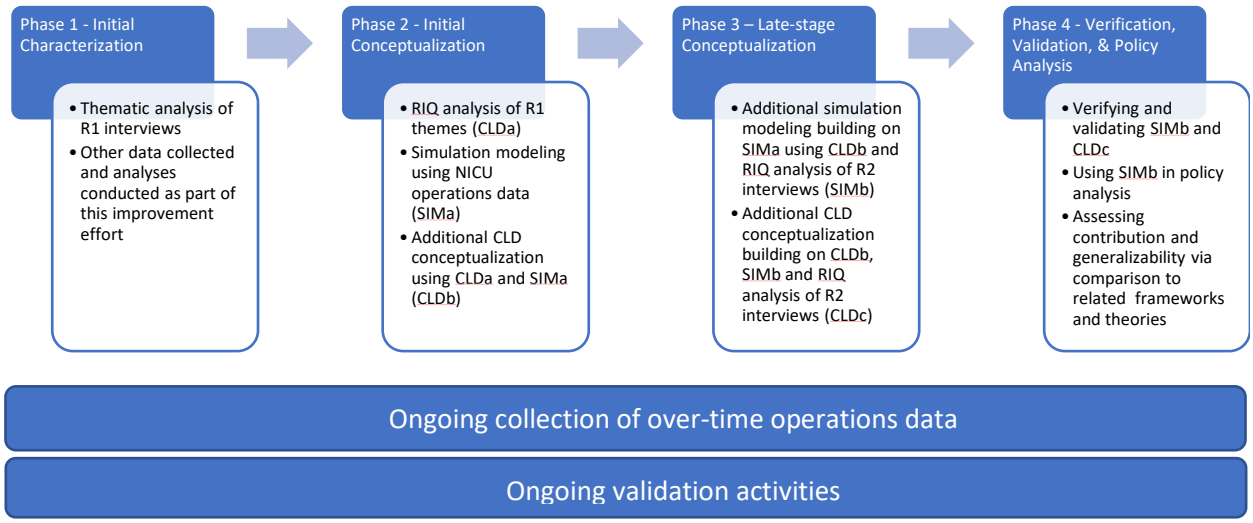
Methods

This study used a mixed methods approach which combines a case study framework(Yin 1994) with SD simulation modeling(Kopainsky and Luna-Reyes 2008) to analyze qualitative and quantitative data (Luna-Reyes and Andersen 2003, Miller, Crabtree et al. 2013, Yearworth and White 2013). More information about SD can be found in Appendix 1. It consists of four phases, where the first and second phases took place at the beginning of implementing UBA, the third phase took place close to the end of external funding for UBA implementation, and the fourth phase took place thereafter.

Qualitative and quantitative data were iteratively synthesized using CLDs as a comparison proxy for interviews and simulation models. Using the terminology introduced in (Fetters, Curry et al. 2013), we used a *convergent mixed methods* design where qualitative and quantitative data were collected over a similar timeframe. Taking an *interactive* (*ibid.*, p.5) approach to this design, initial diagramming and simulation modeling results informed subsequent qualitative data collection, and those qualitative data informed further collection of operations data. We *recurrently* integrated qualitative data collection and quantitative data collection (embedding) by: (1) *building* (*ibid.*, p.7) from the Causal Loop Diagram (CLD) to the simulation and then back again, and (2) *merging* (*ibid.*, p.7-8) the structure found in the CLD and that found in the simulation model to develop the diagram that visualizes a comprehensive understanding of what we have learned. Results presented in this paper are drawn from that diagram and from the final version of the simulation model.

Figure 1 below presents the research design.

Figure 1: Research Design



This research project was reviewed by the Committee for the Protection of Human Subjects (CPHS) of the University of Texas Health Science Center at Houston (HSC-MS-16-0542) and the Human Subjects Protection Committee at the RAND Corporation (2016-0447), and both have determined that this study is exempt from Institutional Review Board review. This research was funded by AHRQ (Grant # 1P30HS024459-01).

Phase 1

Phase 1 involved conducting a series of analyses to evaluate changes in harms (patient safety events) and variances (deviances from established practice) over time, as well as assessing the changes in QI capability ("UBA capabilities") among NICU staff. Additionally, contextual facilitators and barriers to implementation were captured(Sedlock, Ottosen et al. 2018). Purposive sampling guided selection of participants for Round 1 interviews (n = 20) conducted around the beginning of UBA implementation; considerations included breadth of experience (e.g., level of prior engagement with NICU QI, position) and contextual characteristics (e.g., shift, tenure in this NICU). Participants included NICU staff (nurses, physicians, and management) as well as relevant hospital and health system level leadership. These interviews primarily focused on (1) identifying the contextual facilitators and challenges previously at play and (2) understanding how these facilitators and challenges have impacted prior QI efforts at the hospital and health system in which the NICU we studied was embedded. Our interview protocol was developed using context and implementation frameworks from the health services research literature (see (Kaplan, Provost et al. 2012, Tomoaia-Cotisel, Scammon et al. 2013)).

These interviews were initially analyzed using thematic analysis (Braun and Clarke 2006) (themes are reported in: (Tomoaia-Cotisel, Etchegaray et al. 2022)). Because these results suggested causal mechanisms among interacting elements of the interventions developed and implemented during UBA implementation and the NICU context, Phases 2 – 4 of this study used SD methods(Sterman 2000) to document these causal mechanisms, to explore their over-time impacts, and to test potential policies individually and in combination.

Phase 2

Phase 2 focused on the *initial conceptualization*, with our goal being to gain an initial understanding of the structure of the system. We started by developing a CLD of the important variables and their causal relationships (CLDa). We did so by reanalyzing Round 1 themes using Rigorously Interpreted Quotation analysis (RIQ) (Tomoaia-Cotisel, Allen et al. 2022) to create a CLD to visualize the causal language expressed in interviews with NICU staff. Principles of CLD Combination (Tomoaia-Cotisel 2018) were applied in adjudicating how causal structure in each additional interview would contribute to the diagram developed to that point. CLDa focused on how context interacted with the NICU to impact QI prior to implementation of the improvement projects.

Separately, operations data were used to inform development of an SD computer simulation model to capture the flow of babies through the NICU and the impact of QI interventions on variances. At this point, the NICU's internal RPI training and improvement efforts had not yet been implemented, so the QI interventions were developed primarily by the hospital/health system and brought to the NICU for implementation. The simulation model (SIMa) captured these operations via a system of mathematical relationships between variables.

The diagram (CLDa) was then revised to reflect learning from the initial simulation model (SIMa) to produce an updated diagram (CLDb).

Phase 3

Phase 3 focused on *late-stage conceptualization*. Now that UBA had been implemented and utilized in several projects over the course of several years, a second round of semi-structured interviews (Round 2, n=10) was conducted with NICU staff including participants on 3 QI teams. Each team received the same training, but their experiences with QI implementation were not equally successful. These interviews aimed to capture the experiences of participants who were directly involved in the execution of QI projects in the context of the UBA intervention and to gather their perspectives on its effectiveness and impact. These interviews were used to add implementation of UBA into our modeling. In addition to the questions raised in Round 1, these interviews were guided by insights gained from the simulation model findings up to that point.

Using RIQ analysis, the data from the Round 2 interviews were analyzed to further refine the CLD (now CLDc) and the simulation model (SIMb). Quotations from the interviews were matched to the structure of CLDc and to the structure and behavior of SIMb. This process allowed for the identification of confirmations and discrepancies between stakeholder experiences and the diagram and/or simulation model, leading to revisions when necessary. If no diagram or model revisions were made, the reasoning was documented.

Available operations data, literature review and stakeholders' knowledge were used to identify parameter estimates for key variables and to ensure that the simulation model is a reasonable reflection of the real system. This included developing policy scenarios that reflect the variation in QI experiences identified in R2 interviews. The model is calibrated by comparing model output to empirical data, and if discrepancies exist, refining the model and parameter estimates.

Phase 4

Phase 4 focused on *validation and policy analysis* (scenario analysis of simulated decisions, program changes, and operating procedures). Ultimately, SIMb was used for identifying leverage points and policy analysis(Tomoaia-Cotisel, MacDonald et al. 2022).

The CLD and the simulation model were repeatedly tested during the study using informal discussions the vein of Disconfirmatory Interviews (Andersen, Luna-Reyes et al. 2012) with SD experts, NICU staff and leaders, and other key stakeholders. These discussions were conducted at various points throughout the study and in various ways.

Finally, we employed a visual structure-oriented CLD comparison method inspired by Schaffernicht and Groesser (2011) to visually overlay CLDc with related theories (Schaffernicht and Groesser 2011). This approach facilitated the observation of similarities and differences between CLDc and these theories. First, we did a comparative analysis of CLDC and the Model for Understanding Success in Quality (MUSIQ)(Kaplan, Provost et al. 2012) as it was used to inform our initial interview protocol development. Second, we compared CLDc to diagrams of a leading theory of implementation in complex situations, Normalization Process Theory (NPT)(May, Mair et al. 2009) and to the updated versions of MUSIQ(Reed, Kaplan et al. 2018) and NPT(May, Johnson et al. 2016) that consider dynamic complexity. To enable a meaningful comparison, we assumed that the diagram elements found in these frameworks are comparable to those used in a CLD. Given our interest in comparing our model to the feedback loops within these frameworks, we focus on the elements found in closed loops. For MUSIQ,

this meant collapsing lists of elements in a box into single variables. These two comparisons are reported in Appendix 3. Third, we compared CLDc with existing cross-industry SD work on service operations (Anderson Jr, Morrice et al. 2005, Rust 2013) and formal QI implementation (Repenning and Sterman 2002). Detailed insights from all three comparisons are elaborated on in the discussion section.

Results

Here we present the **UN**it base**D** **A**pproach in the NIC**U** co**NT**ext: f**E**edback **D**ynamics (UNDAUNTED) Model. The shaded areas in Figure 2 visualize the three major aspects of our study: (1) the general NICU operations that support patient care (NICU Operations & Patient Care), (2) NICU staff QI efforts, and (3) the UBA intervention (Unit Based Approach to QI). Outside this clinical context, there are various organizational policies impacting UBA implementation (see Figure 7).

The variables and links in Figure 2 visualize the causal interdependencies between these aspects using a CLD (Appendix 2 provides an orientation to how to read this CLD). Causal pathways form feedback loops that drive system performance over time, including both intended and unintended consequences.

The following subsections describe insights gained from this conceptual understanding. Unit operations and patient care are described in **Section A**. The role of unit staff in traditional unit QI is described in **Section B**. **Section C** describes the UBA QI approach. **Section D** describes unintended consequences arising from UBA's interactions with patient care. **Section E** describes supportive policies in the unit context.

Figure 2: UNDAUNTED—Core Feedback Structure of the Unit-Based Quality Improvement Approach

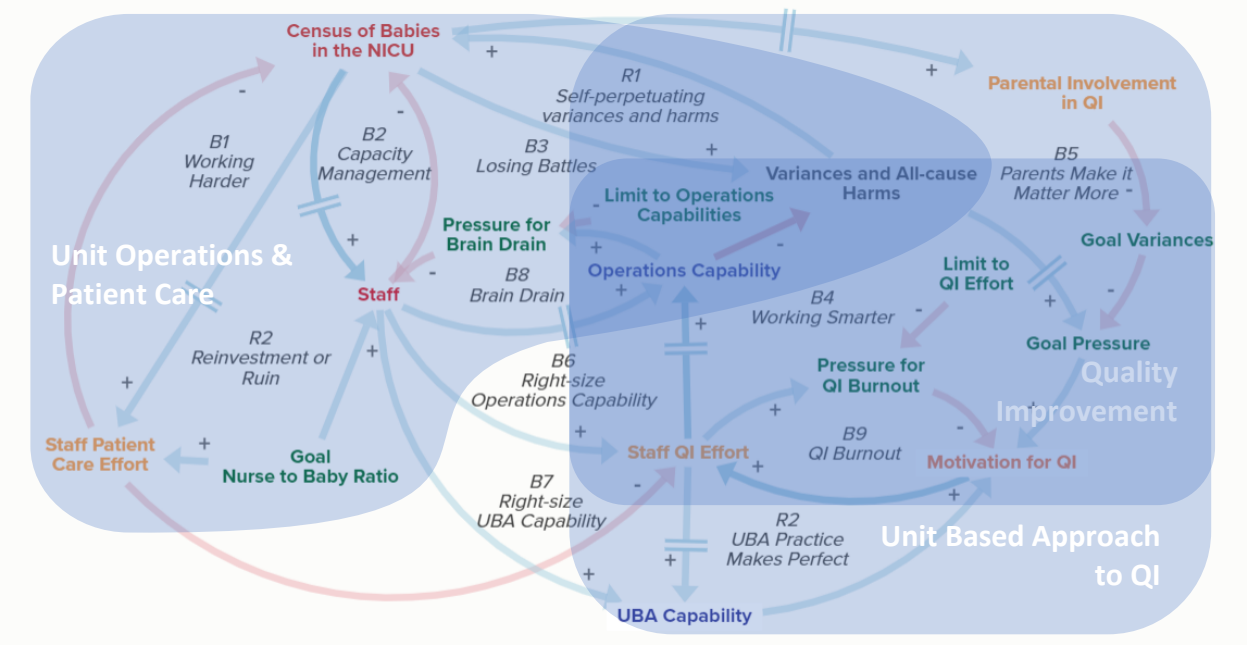


Figure Footnote: Variable Categories: **People**, **Goals, limits and their effect on pressure**, **Capabilities**, **Harms**, **Effort**, and **Motivation**. Link Polarity: Positive, Negative. Loops: *B*=balancing, *R*=reinforcing.
An interactive version of the CLD in this figure is available here.

(A) NICU operations and patient care context

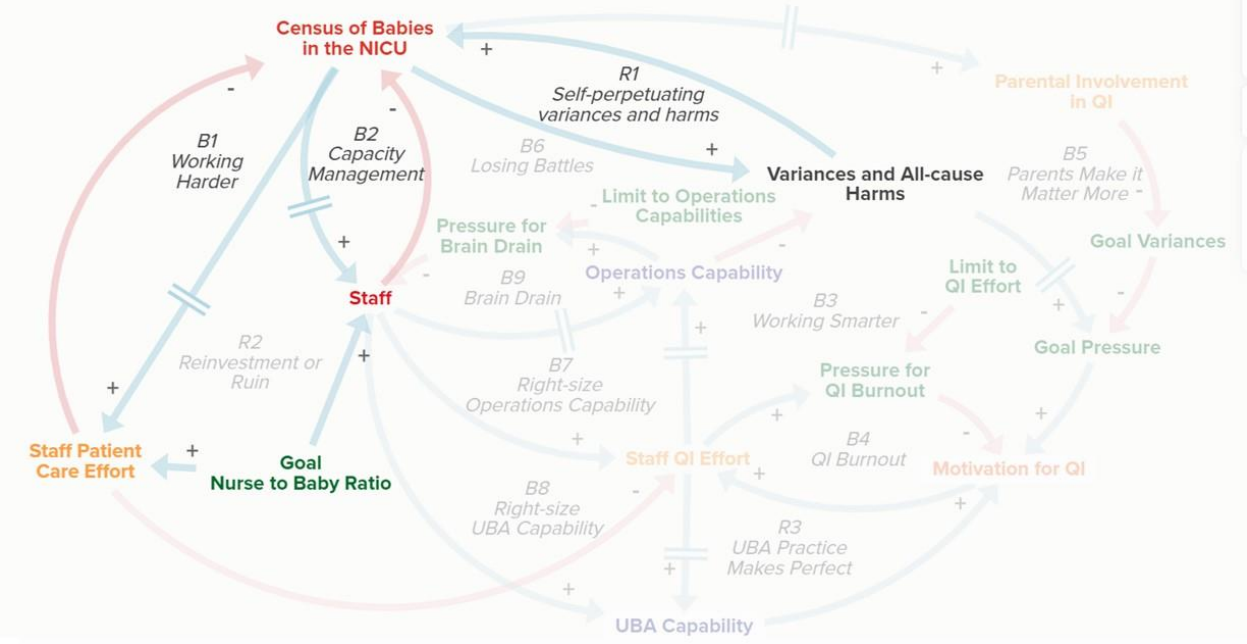
Figure 3 visualizes NICU operations in support of patient care. The NICU serves newborn babies needing intensive care. NICU managers and nursing staff respond to changes in the total number of babies, known as the 'census', in two ways. First, they respond by adjusting the

intensity of their patient care efforts. As more patient care is provided, patients recover more quickly, keeping the census under control (**B1 Working Harder**). Second, management adjusts the NICU staffing level. In the short term, charge nurses adjust scheduling from the larger pool of existing staff. In the longer term, unit and hospital management change the size of that staff pool (**B2 Capacity Management**). Whether by working harder or managing capacity, the NICU's ability to respond to changes in the census is limited by the desired nurse-to-baby ratio, which accounts for the acuity of current patients.

Beyond the daily variation in admissions to the NICU, the patient census changes over time depending on the quality of care provided. A portion of variances cause harms. When patients are harmed, they need to stay longer to finish their recovery. Variances can also cause inefficiencies that lead to longer length of stay, without involving harms. Either way, a longer length of stay increases the census. In the short term, an increasing census means more chances for variances and harms to occur (**R1 Self-perpetuating variances and harms**). Participants reported that passively sustaining a low level of all-cause harms is challenging because the sources of variance are part and parcel of healthcare. For example, one participant mentioned that the hospital size, patient acuity and changing management priorities create upward pressure on variances.

I think CLABSI will be [sustained] because that's something that comes from so high that they will be… and it's not going to go away. I think there's always going to be line stream infections that you can't maintain zero or one for very long periods of time. And an institution this big, the kids we have, I don't see that ever happening.

Figure 3: NICU operations and patient care context



(B) Engaging NICU Staff in *Externally* Led QI Improvement

Figure 4 below shows the traditional NICU QI approach. Whether at the hospital, health system, or both (as in the case of the healthcare system we examined) management outside of the unit tracks variances and harms. Before UBA, when managers noticed an uptick in these outcomes, they assigned outside QI experts to come to the NICU and work with staff to find solutions to the identified causes of harm. Unit staff who also noticed the uptick in outcomes became motivated to make changes, but the extent to which they were engaged was limited (minimally if at all for most people). A small number of staff participated in these processes, and they were allowed only a small amount of paid time to do so. They reported that the processes performed poorly at enhancing cohesion among NICU staff working in various roles and professions or at developing interventions that were based on a shared understanding across QI group participants, both being key elements in successfully implementing QI interventions. The outside solutions became interventions in the NICU with the potential to increase the operations capability of NICU staff, and thus decrease the associated variances and harms. Nevertheless, because of delays in implementation and in perceiving improvements, staff sometimes lost motivation for the changes being implemented. Thus, this feedback loop (**B3 Working Smarter**) was operational, and preferred to just B1 Working Harder, but its potential for QI was realized only to a limited extent.

Participants described NICU staff's perception of their pre-UBA QI efforts as largely perfunctory on the part of both staff and leadership.

[Several years ago], we did have QI projects. I honestly didn't really know of them as a staffer. They were done behind the scenes. It was something that was seen as the leadership job. We went to committee meetings and stuff, but we didn't really see any of that as QI. We just saw that as staff going to committee meetings. We'd try to make some decisions but then it never would go through leadership.

Participants reported that there is a limit to the amount of their day that they were allowed to spend doing QI. Pushing this limit led to reduced motivation for further participation in QI (**B4 QI Burnout**). NICU nurses are highly motivated to make Qis, but were continually disappointed by this QI approach, which contributed to cycles of joining and quitting. Continuing the previous quote,

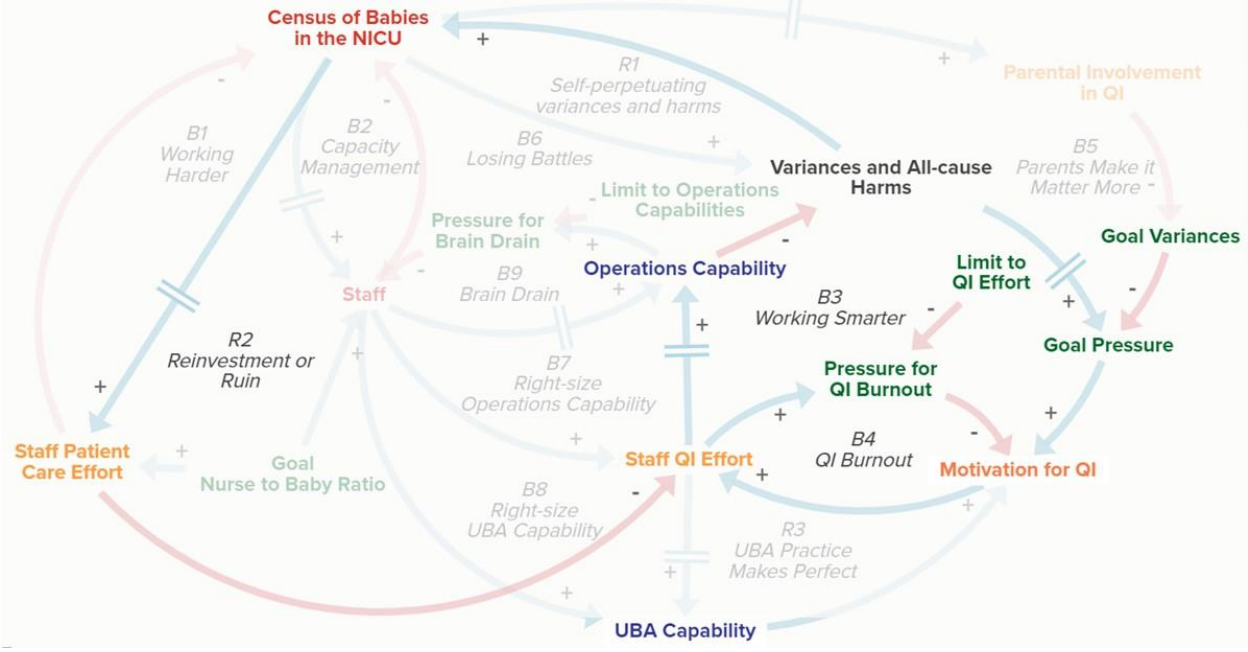
Nothing really ever changed and then people would just quit the committees. Truly. I think what we saw in our unit was this momentum of nurses wanting to make a difference. Joining a committee. Not really achieving the results they wanted. Quitting the committee. Then the committee starts back up again. Same cycle, over and over and over again.

As NICU staff engaged in QI effort, they improved their operations capabilities. With these new-found capabilities, variances and harms are decreased. With fewer harm-causing variances, patients leave the NICU more quickly. The census goes down and less patient care effort is needed. NICU staff can use the added time in their day to expend additional NICU QI effort. In doing so, NICU staff reinvest in their successful approach to QI (**R2 Reinvestment or Ruin**).

Alternatively, in cases where the census is increasing, staff must put in more of their effort for patient care, leaving less time for QI. Over time, this leads to a ruinous vicious cycle where neglecting QI causes reduced operations capability. The reduction in effort means the growth in operations capability from the limited QI effort now being expended cannot make up for capability losses due to new technologies disrupting processes, forgetting, staff turnover or other such causes. Falling capabilities means more variances and harms, and all else equal, a further increase in the census (**R2 Reinvestment or Ruin**).

Patient care takes precedence [over QI], and that is a time constraint for us. There's not a lot of extra time when we're at work.

Figure 4: Engaging NICU Staff in Externally Led QI Improvement



(C) Engaging NICU Staff in *Internally* Led QI Improvement – More QI Engagement, UBA Capabilities, and Parental Involvement

Figure 5 below shows how two key ingredients of UBA strengthen loop B3. UBA *relies* upon NICU staff to expend their own effort in internal QI – choosing projects, setting goals, implementing interventions. QI is now led and performed in the NICU by capable NICU staff. As NICU staff expend QI effort, they find solutions to variance problems and change the associated policies/process/technologies, thus increasing their capability to provide better care (i.e., the NICU's operations capability). The added operations capabilities decrease variances and harms. This success relieves the associated pressure for continued improvement – on to variances and harms that continue to be a challenge. NICU staff reported continuing to engage experts at the hospital or health system level, at international safety organizations like the VON, and outside consultants. This engagement involved mentoring on specific aspects of a method, sharing of ideas for interventions to address an identified problem, etc. Staff described a higher level of

motivation across the unit as a whole, and particularly for those staff directly involved in the UBA QI projects. Under UBA, more NICU staff were motivated to engage, and did so.

In addition to delegating QI tasks to NICU staff, a key ingredient of UBA is the introduction of UBA capabilities among a significant portion of NICU staff. Building UBA capabilities brought about increased motivation as staff felt capable of taking on harms in their NICU. In addition to technical skills, UBA training also included organizational factors like meeting facilitation and overall QI management, which further fostered NICU staff motivation for QI by increasing cohesion in interdisciplinary QI teams and building shared understanding around intervention designs. With increased motivation, a greater portion of staff were engaged in NICU QI, and they were engaged to a greater degree than before. As they participated in NICU QI projects, NICU staff strengthened their UBA capabilities through learning by doing (**R3 UBA Practice Makes Perfect**). One participant described how improving their knowledge of QI in general has through practice can provide a sense of success even when projects seem to have no impact on outcomes.

You can have success without actually moving a mark. … Our group hasn't really done anything in terms of moving a mark, but we've learned a whole lot along the way, which is helping us create new projects. I'm not sure that's not successful. Which is not the same type of success that we looked at in the past.

Another participant described the difference that UBA makes as a change in culture. This culture manifests in how nurses use their UBA training to lead the staff, and how all must be undaunted by the forces resisting sustained improvement.

[In] sustaining those [improved patient] outcomes over time … [in] the places where we've been challenged … much of that work was culture work. They did not do anything different. We didn't come up with some other way to keep things clean. We weren't missing a 'piece'. We just weren't doing what we said we were doing …

[When doing audits, our UBA-trained nurse staff leader] captured people's attention and made them think differently. [It wasn't] well received by everyone all the time. I would say that there was a lot of pushback on, obviously, "I know how to do this, I'm a nurse, I've been doing this a long time. I cannot believe that you're asking me to do these things." And she would just kindly say that "I am."

And she did. And then they did it consistently, and they did not give up when it got hard. And slowly but surely people figured out that, "maybe we did have a problem and maybe this is a better way to go about this". She just kind of kept going and she can deliver messages really well. … I think having a leader that was supportive but firm.

Implementing and sustaining QI involves persistently monitoring capabilities and working *undaunted* through learning delays (**B3 Working Smarter**). With UBA, this loop is now functioning more fully. Staff implement and are determined to learn. Leaders are not neglecting this important corrective function or taking the staff's resistance to change personally.

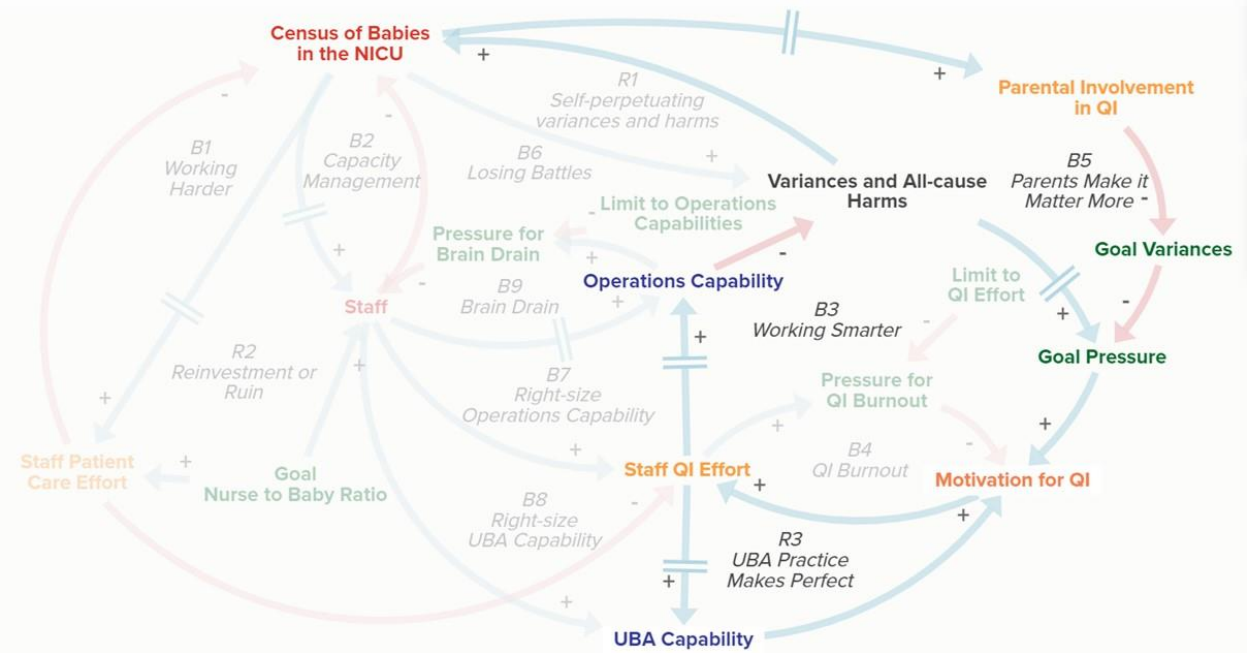
Another key ingredient of UBA is the initiation of efforts to formally engage parents of NICU babies in NICU QI. In UBA, parents support NICU staff patient care by participating in staff-

led UBA QI projects, usually by being invited after a significant amount time post discharge. NICU staff participants reported that the parent's voice is important because as parents share their experiences and perspectives, it helps QI team members to put variances in the proper context as causes of real harms to babies, leading them to set more ambitious QI goals (**B5 Parents Make it Matter More)**.

I have a lot of contact with [parents]. … I think their insight has changed our minds on a whole lot of things on how what we do impacts their lives. … They have seen things, … [for example,] on the IV infiltrate group. So, her baby has had IV infiltrates, … an IV burn … she had mentioned that it looked red to her, and the nurses were like, "it was fine", but then obviously something happened.

These two ingredients support NICU staff in their efforts to find ways of *working smarter* rather than only *working harder*.

Figure 5: Engaging NICU Staff in Internally Led QI Improvement – More QI Engagement, UBA Capabilities, and Parental Involvement



(D) Becoming a victim of one's own success – the unintended consequences arising from QI's interactions with patient care

Prior sections have described how UBA has improved the effectiveness of QI efforts in the NICU. However, signs show that they may have begun to be victims of their own success. QI success has led to improved operations capabilities, reductions in variances and harms. However, it has also led to a reduced census. Responding to long-term changes in the census with capacity management (Loop B2) is fraught with problems (see Figure 6 below).

Participants recognize the tension between successful QI, which is good for patients, and its effect on the census, which is not good for NICU operations.

We also have been looking at our length of stay, which is decreased on average by about three days, I think, this year, which was awesome, not awesome for our census, but really good for patients, so that's good. So, we're doing really good work.

One response to this is to increase the census by transferring patients from other NICUs. This impacts the overall patient acuity in the NICU, and the overall workload. In general, the cause of changes in census, variances and acuity are difficult for NICU staff to perceive.

I don't know if it's just our acuity is changing and it's, we're getting sicker kids, I'm not for sure. Like right now our census is kind of low, but our acuity is really, really high.

On the one hand, as a strategy for addressing a decreasing census, capacity management counteracts QI successes. With time, a falling census leads to a reduction in NICU staff. Reduced staff means a loss of operations capability. The staff members turning over have advanced capabilities compared to new hires, by virtue of their experience in this NICU. This kind of capability reduction means an increase in harms and variances over what they would otherwise have been (**B6 Losing Battles**).

Also, staffing reductions reduce the total amount of time that the NICU staff devotes to QI. As above, neglecting QI leads to a ceteris paribus decrease in capabilities. Thus, this is another way that capacity management policies counteract QI progress over time (**B7 Right-Size Operations Capability**).

Furthermore, staffing reductions lead to a decline in UBA capability because UBA capable staff are turning over. These staff had become motivated to participate in QI, meaning their loss leads to a lower degree of motivation in the unit. Less motivation, less QI effort, thus worsening the situation for sustaining prior QI progress **(B8 Right-Size UBA Capability)**.

One nurse describes how staff members with longer tenure have more experience with QI than newer ones:

I think a lot of people don't even know that [QI] happens [but] … nurses that have been here 15 or 20 years …. They know it's here, some of them do it because that's what they want to do. And then we have that 5 to 10 year group that they know about it, they do, some do it, some don't. And then we have that 1 to 5 year group who, "Oh, hmm. You are doing what?"

On the other hand, as a strategy for addressing an increasing census, capacity management takes too long. Not only does hiring staff take time, but there are also delays in the learning process for new hires in this NICU. This is because most of the learning is on the job – learning by doing (i.e., as they learn "the [health system name] way" for each patient care process). Delays in hiring and learning mean that the impact of hiring on reducing variances and harms is negligible in the short term. Compared to the nearly instant self-perpetuation in the R1 Loop described above, adding staff to improve quality means playing catch-up at best (**B6**

Losing Battles). One nurse describes how some capabilities can take a long time to master because they are only needed occasionally and may not be part of planned on-the-job training.

You only get what you need if it's available. ... You're only going to learn this skill if you have [a patient needing a] chest tube in a unit. So, we may go six or eight months a year and you not be [scheduled] when we had a chest tube [patient]. So [with a brief training period] I might not be here to have a kid on the ECMO pump because we have such long periods of time and that. [So, this will not be learned by new nurses on the unit who] get three months orientation and then they stay on day shift for a year.

Another nurse describes the shortcomings of training with ensuring that new staff absorb the best practices developed in prior QI projects.

So, it was a proven thing … and it's ingrained into this group of people, but it's ingrained, and you do it, but you might forget to teach it, because it's [taken for granted]. So, I think maybe that's it. So, then you have new people in here that weren't taught this specific thing and…, that [this is how we do it] … and so it maybe didn't come up during the orientation, so then there's a little fall-out that way.

An experienced nurse describes how, over the long term, new NICU staff members bring new ideas which helps the unit to adapt to the slow evolution in practice standards.

It's evolved over time, I've been here for a long time, and we've grown. So, I came from when we had maybe 50 beds, and now we have 120 beds. … with that comes good things and we get more perspectives come in, because there's newer people. … we have a lot of newer and younger attendings who come in with a lot of good ideas. … nurses as well, RTs. So, I think all in all … people are pretty accepting of new ideas in general as long as there's a basis behind them. So, things change all the time in the neo[natal] world. So, but, like I said, I think the people are adaptable pretty well here.

The good news is that although new staff do not usually participate in QI, they do increase the amount of time spent on QI overall in the unit by allowing existing staff to spend more time in QI efforts. Increased QI effort leads to increased operations capability beyond what otherwise would have been the case (**B7 Right-Size Operations Capability**).

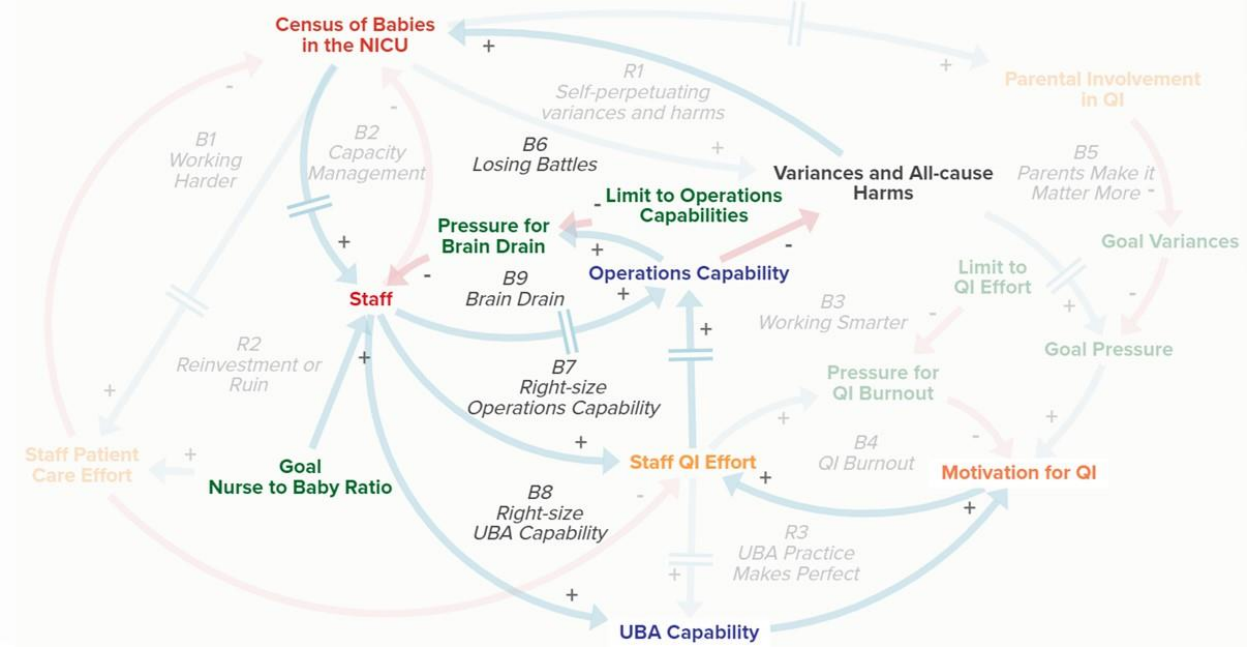
Finally, as new hires complete on-boarding they receive an initial training in UBA. This has a marginal impact on their motivation, and they have little time to participate in QI. Nevertheless, their new ideas are valued by existing staff, and they make a small contribution to overall improvements in operations capabilities **(B8 Right-Size UBA Capability)**.

Participants reported that the health system is in a very competitive market for skilled staff. Staff with increased operations capability are recruited to work in competitor health systems. Thus, as staff become more capable, turnover increases. In a scenario with low operations capability, staff are less attractive on the job market and feel less pressure to leave. This loop responds to the relative attractiveness of staying at this NICU versus going to another one – allowing capacity management policies to short-staff the unit after QI successes makes

the NICU a less attractive place to work. This communicates to staff that their motivation and efforts are less valued than they may have expected, unintentionally setting a low ceiling on the unit's operations capabilities (***B9 Brain Drain***).

I think we feel like we've implemented it, it's good, but I think everything needs some reinforcement at intervals, and I think that's where we fall short is the reinforcement part. … Maybe we (reach a target and) move on to another project or we feel like we've crossed that hurdle or fixed that problem, but then new people come in, things like that. And so, we fall off a little bit.

Figure 6: Becoming a victim of one's own success – the unintended consequences arising from QI's interactions with patient care



(E) Supportive Policies – what can be done?

The CLD enables creative thinking about policies that have the potential to impact the feedback process that is UBA implementation in a NICU context, (see Table 1). The left column suggests each variable as a leverage point. The right column identifies policies that participants noted were in place or that could be revised/added to support all cause harm reduction through UBA implementation are described in relation to the type of leverage they would exert. The last three rows in this table have greyed out cells indicating that these are not considered points of leverage at this time and an explanation is provided as to why.

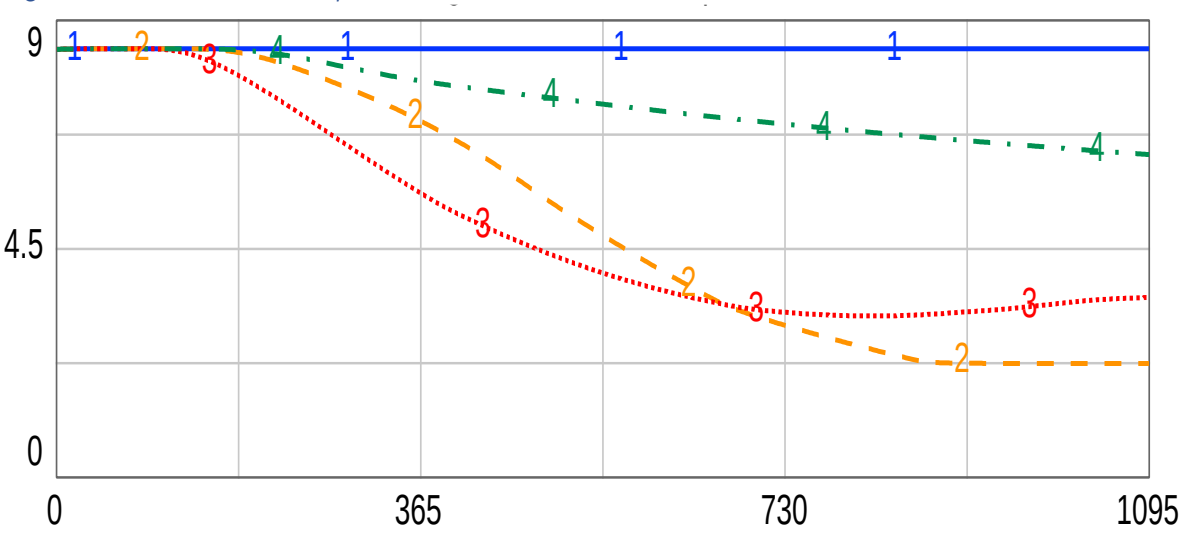
Table 1: Supportive Policies

Leverage Point	Policies identified for consideration
Census of Babies in the NICU	Policies to **change the inflow and outflow** of patients: - transfer policy (e.g., bring more high acuity patients from satellite hospitals, from lower-level units inside this hospital, also transfer low acuity patients from this NICU to other units inside this hospital or to satellite hospitals) - policy to reward referring internally inside the health system - marketing in the community
Goal Nurse to Baby Ratio	Policies that would **impact this goal**: - the current Desired Nurse to Baby ratio - Adjusting the ratio based on patient acuity - Adjusting the ratio based on QI effort anticipated
Goal Pressure	Policies that would **impact the perception** of variances and harms: - What data is collected and reported to staff (e.g., units of measure, what length of time does the report consider)? - How often is it reported to staff? - How is it reported to staff? (e.g., in QI meetings, mass emails, physical bulletin boards in the unit) - How much does management prioritize variances and harms?
Goal Variances	Policies that would **impact the variance goal**: - the current variance goal - quality mandates with financial repercussions - internal management continuous improvement goal
Limit to Operations Capabilities	Policies that would change **the value placed upon operations capabilities**: - incorporating operations capability into employee evaluation and promotion - Benefits offered that compensate capable staff for operating in harsher conditions (e.g., short staffing bonus, overtime) - accounting for operations capabilities (including newly-demonstrated ones) in nurse to baby ratio
Limit to QI Effort	Policies that would **recognize** that total **workload** includes non-patient care effort (e.g., QI) and show value for that part of the workload by: - allowing for space/time in the day when that work can be done (thus upping the limit of QI effort that can be expended) - managing the QI and patient care workload so that, in periods of short staffing, changes to the balance of effort take into account ongoing QI project needs
Motivation for QI	Policies that would **change staff members' motivation** for QI: - sources of external recognition (parties, conferences, etc.) - policy that QI involvement matters for employee evaluation and promotion

Leverage Point	Policies identified for consideration
Operations Capability	Policies that would **change staff operations capabilities**: - what is included in onboarding of new staff - what is included in on-the job training for new staff (and what is forgotten/taken for granted)
Parental Involvement in QI	Policies that would **change the amount of parental involvement** in QI: - when to engage (during, at some time point after having baby discharged from the NICU) - how much to engage (in PAC, in 1 QI project, in more) - for how long to engage an individual (they need to have a break) - orienting parent to PAC and to participation in QI projects (empowering them to provide their input)
Staff	Policies that would **change the number** of staff working in this NICU: - hiring policies (number of openings, delays in the hiring process) - benefits offered (short staffing bonus, overtime, etc.)
Staff Patient Care Effort	Policies that would **support** charge nurses in **scheduling** patient care so that there is dedicated staff time for QI
Staff QI Effort	Ways of **obtaining resources** to support staff QI effort: - outside networks (e.g., Vermont Oxford Network) that share QI experiences - outside grants and contracts to cover staff time doing QI - having hospital QI staff doing some of the tracking/monitoring, the analysis, and/or data access that might not be readily available Policies that would change the amount of Staff QI effort expended: - make QI meetings accessible to all staff (e.g., night shift should be able to participate too)
UBA Capability	Policies to **change staff UBA capabilities**: - one-off formal training for existing staff - formal training to new staff during on-boarding (or at some point after on-boarding) - who gets formal UBA training and when (e.g., new hires, seasoned employees at some point, those interested, everyone)
Pressure for Brain Drain	Staff decide how they feel the pressure
Pressure for QI Burnout	Staff decide how they feel the pressure
Variances and All-cause Harms	Accessible through the Census and Operations Capability only

Simulation was used to evaluate the impact of changes in selected policies, including but not limited to: Goal Variances, Parental involvement in QI, and UBA Capability. Figure 7 below presents simulation results for these policies, and for a run without any policy change (Status Quo). These runs project the endogenous impact of policy in the simulated NICU QI situation. As such, it does not represent the stochasticity endemic to the real context. Other policies followed trajectories similar to one of these and are not shown.

Figure 7: All Cause Harms per Month



Legend: Time units = Days. Scenarios: (1) Status Quo, (2) Parental Involvement, (3) Improvement Goal, (4) UBA Capabilities for New Staff

In each scenario, the policy achieves decreased harms compared to the Status Quo scenario (line #1, Figure 7). While the trajectory of harms for specific policies varied, four behaviorally different scenario trajectories are evident: gradual and shallow reduction, gradual and deep reduction, early reduction with some regression and no noticeable reduction. Table 2 below lists the policies which were evaluated in simulation, describes the mechanism and meaning of the policy, and summarizes the impact each has on all cause harms by the end of the simulation.

A policy of **Parental Involvement** (#2) accomplishes gradual harm reduction, ultimately leveling off at the greatest amount of reduction. Compared to this, the **Improvement Goal** (#3) policy makes for early harm reduction before bottoming out and ultimately starting to regress. The impact of increasing **UBA Capabilities for New Staff** (#4) is gradual and more shallow harm reduction. The remaining policies (External Motivation, UBA Capabilities for Existing Staff, UBA Capabilities for New Staff, Nurse Retention, and Nurse to Baby Staffing Ratio) all follow a similar trajectory. These policies have shallower trajectory because they face long delays in decreasing harms . In the timeframe considered, these policies have not yet attained their Improvement Goal – they continue to push the system to decrease variances (and thus harms) albeit very slowly.

Table 3: Trajectories Projected for Harms Under Different Policy Levers

Policy Lever	Model Structure	Scenario trajectory
Parental Involvement	Using a step to model a sustained increase to the level of parental involvement (more parents or more engagement per parent).	Slower initial harm reduction, ultimately leveling off at a substantial amount of harm reduction
Improvement Goal	Using a step to model a sustained increase NICU staff members' desired Improvement Goal (via intrinsic sources of motivation – because they are committed to a more ambitious reduction in variances and harms)	Quicker harm reduction, appearing to level off at a substantial amount of harm reduction, and ultimately starting to regress
External Motivation	Using a step or a pulse (as reported in interviews) to increase NICU staff motivation to participate in NICU UBA QI projects via external means (ones reported in interviews included recognition from leadership, parties for reaching desired outcomes, support for sharing stories of success at conferences)	Minimal harm reduction if any
UBA Capabilities for Existing Staff	Using a step to model a sustained increase to the capability of existing staff to implement a UBA approach to QI efforts (by increasing nurses' understanding of UBA on-the-job as they participate in NICU UBA QI projects)	Delayed harm reduction and to a lesser degree during the timeframe considered
UBA Capabilities for New Staff	Using a step to model a sustained increase to the capability of new staff to participate in UBA QI efforts (by increasing nurses' understanding of UBA during onboarding)	
Nurse Retention	Using a step to model a sustained increase to the tenure of nurses in this NICU (e.g., via improving benefits and satisfaction (e.g., via rewarding UBA QI project participation))	
Nurse to Baby Staffing Ratio	Using a step to model a sustained increase to the number of nurses assigned to one baby	

Scenario results for other variables are reported, along with the variable's location on the UNDAUNTED CLD, see Figure 8. In these charts, the variable name is the same as the units. The parenthesis gives the measurement scale. Normalized scaling helps to focus on behavior trends, rather than absolute values.

Staff capabilities to perform a UBA QI project increased in all policy scenarios implemented. This is what enables NICU staff to successfully execute UBA QI work. Staff operations capabilities also increased in all policy scenarios implemented. This is what enables variances (and thus harms) to decrease. These outcomes were expected; however unexpected outcomes were also observed when comparing interventions to the status quo and to each other:

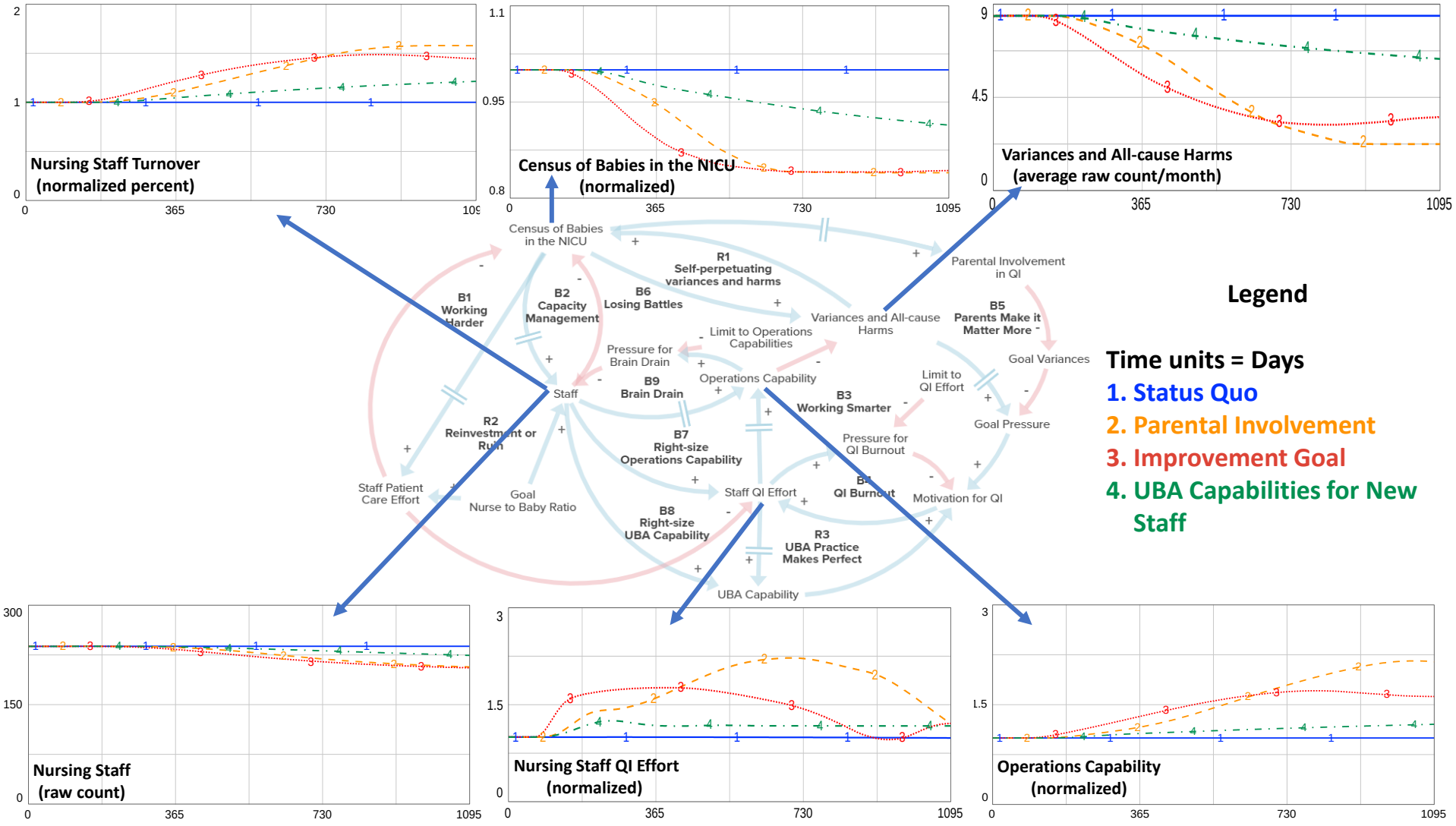
Comparing interventions to the status quo, we learned that in intervention scenarios:

- The census would decrease over time. With fewer harms, the same inflow of babies results in a lower census as babies are healthier and discharged more quickly.
- Staff turnover would increase over time. This is because, with a lower census, an externally set staffing ratio dictates that fewer nurses are needed on the floor.

Comparing interventions to each other:

- Slower paced harm reduction led to greater long-run reduction in harm because staff were able to sustain QI effort, and thus gain the operational capabilities they needed to sustain long term harm reductions (line 2).
- Faster paced harm reduction led to greater early-on harm reduction, but staff suffered increasing burnout, increasing staff turnover, and thus limiting and ultimately even reducing the improvements made (see rebounding all cause harm) (line 3).

Figure 8: UNDAUNTED—Core Feedback Dynamics and Examples from Simulation



Discussion

Insights

Over the course of four years (2015-2019), we collected and analyzed data with the aim of unraveling the interdependencies between the UBA intervention and its context to improve understanding of their dynamic implications. We used SD modeling to document the ways that these factors drove changes in outcomes. We moved fluidly between simulation modeling, numeric operations data and textual analysis mapping interview data onto CLDs. Both numeric data and the use of simulation helped us to expand beyond the perceptual limits of subjective mental models to include important operational causal structures. Furthermore, the qualitative data and diagrams also extended our understanding beyond the perceptual limits of operational modeling with numeric operations data. This mixed methods approach proved valuable in synthesizing our evolving knowledge of system structure with its dynamic behavior, contributing to a more comprehensive understanding.

Our application of this method to understand this NICU yielded several key findings, which hold practical implications for both organizational leaders and researchers. In this section, we describe insights for three key topics: motivation, parental involvement, and 'victim of success' dynamics.

One insight is that staff motivation for QI is vital to sustaining QI success. Organizations interested in realizing improvements from QI projects should measure, monitor and then act upon changes in staff motivation. Participants identified the following sources of motivation for QI: providing staff with accurate information feedback on the impact of their QI efforts on safety (variances and harms), attending to their QI and patient care workload (avoiding QI burnout), and supporting staff members' training on formal QI methods (UBA capabilities).

Study participants described that NICU staff members were much more motivated as a whole to do QI when QI teams were highly capable in UBA. Skilled UBA teams were more effective at developing shared mental models in their interprofessional QI teams. Shared mental models are important to establish because they have been associated with better team communication and coordination and allow stakeholders to understand the big picture as well as the influence of factors on key variables and outcomes. Skilled UBA leaders were *undaunted* by the staff's natural resistance to change, and they led the staff to implement and sustain QI changes.

CLDs are useful for developing shared understanding, more formally, for defining a system boundary and identifying the most likely loops to influence a particular problem over time(Martinez-Moyano and Richardson 2013). So, academic ((Tolk, Cantu et al. 2015, Shire, Jun et al. 2018) and gray literature on QI (c.f.,(Barry 2010, Kaufmann and Chieh 2012, Rai 2012, Rushing 2012, Hallowell 2016)) have often recommended that QI teams use CLDs and/or SD simulation models in QI efforts. Using CLDs in QI was a key aspect of early critical research on QI and on an early conceptualization(Repenning and Sterman 1997) of the RPI QI method. Nevertheless, CLDs and simulation are not part of the Joint Commission's RPI training, and they have not been used by QI teams at our case study institution. Our interviews focused on implementation experiences and as such did not focus on the use of specific tools in QI.

Further research is necessary to gain a better understanding of how the use of CLDs and simulation models in healthcare QI efforts affects shared understanding and motivation. Additionally, exploring optimal training techniques for front-line and formal QI staff, such as systems and industrial engineers, and assessing the impact of these methods on outcomes and the implementation of specific QI projects, will provide valuable insights. Our CLD could be useful in guiding these efforts. Our simulation could also be improved as a representation of healthcare QI dynamics by evaluating its ability to reproduce qualitatively described experiences, and key behavior patterns observable in numeric data which represent enduring changes in key variables such as patient flows, errors, and QI efforts, which have been documented in other contexts including NICUs and other situations where front-line QI and other policies are used. The CLD and simulation provide useful insights into data collection for this purpose.

Similarly, unit and hospital leadership could use our CLD in UBA-like program design, implementation and evaluation as well as in considering how to engage front-line staff in external QI efforts impacting the NICU (or other hospital units). Further studies can explore the value of this CLD for helping managers understand the dynamics of healthcare QI implementation and incorporate RIQ analysis to measure changes in mental models (Doyle, Radzicki et al. 2008). Research along these lines would also allow for extending the CLD.

Second, parental involvement in QI efforts in a NICU setting can help increase the staff's commitment to QI, which translates into more ambitious unit goals focused on reducing variances and harms. Having an active parental advisory council was crucial for staff to begin defining quality in terms more similar to the ones parents use, a critical element of any QI approach known colloquially as "voice of the customer". As of the close-out of our study, NICU staff were still working to find ways of sustaining parental involvement, (e.g., recruiting parents at least 2 years after discharge to allow them time to recover fully first). Few study participants had experienced this innovative practice and those who did described expectations more than experience, and none of our participants were QI-involved parents. Future research is needed to better understand the QI-involved parent's experience by focusing on this issue in participant selection, including of parents.

Third, one perplexing situation that arises is that having fewer neonates harmed means that their length of stay decreases initiating a cascade of situations that results in reduced operational and UBA capability to conduct QI projects. This confirms the findings of SD work on "the capability traps", which describes problems organizations across industries have with sustaining successes in improvement(Repenning and Sterman 2002, Sterman, Oliva et al. 2015). The insight that *improving capability now can harm capability improvement later* was a call for action two decades ago; that, in order to preserve hard-won capabilities, what is needed is to improve processes and organization simultaneously. Early research (Repenning and Sterman 1997) on the Capability Trap likely informed the RPI method (Adrian 2009, Chassin and Loeb 2013) at the core of UBA (or at least reached the same conclusions earlier). Indeed, RPI is designed to make up for deficiencies in traditional QI. We document how this succeeded in a challenging organizational context. Despite this, UBA faces the same contextual challenges that its formal QI predecessors faced. The need to coordinate process change with organizational change persists two decades later and across industries, including healthcare. Future research on implementing QI in healthcare could use our CLD to inform such coordination. This could

include studies of innovations that shift the capacity management paradigm from staff-mix toward a skill-mix perspective, whether they be scheduling grids, staffing models or regional health workforce policies (Dubois and Singh 2009).

Finally, we also documented specific policies that can be implemented alongside UBA to address different factors contained in our CLD and to improve understanding of potential unintended consequences. In future research, researchers can gather information on the current state of these policies and run simulation models to test their impact on outcomes of interest and then share these findings with organizational leaders to help inform decision making about which policy or policies to implement in particular contexts. For example, researchers might examine the impact of reducing staff turnover by improving their non-salary benefits on operational capability, harms, and other outcomes over time. Further, researchers can examine the impact of multiple policies simultaneously, which allows organizational leaders to factor in attributes of these policies such as the cost and complexity involved when enacting a policy.

Contributing to Dynamic Understanding in Existing Theory

Two commonly-used frameworks used in health services research are MUSIQ and NPT. These frameworks were designed to enhance understanding of improvement and implementation, respectively. They are built on accumulated findings from prior studies, and secondary analysis, and both have evolved over time to more explicitly represent complexity (May, Johnson et al. 2016, Reed, Kaplan et al. 2018). These changes were made by NPT authors partially in response to criticisms about the agent-structure dilemma (see also (Lane 2001b)). The improvements include focusing on *individuals'* capabilities (i.e., "participants' capacity and potential to respond"(May, Johnson et al. 2016)(p.3) and treating context "as a process rather than a place"(*ibid*., p.4). We take the same approach. Broadly speaking, our CLD provides more detailed insights into the causal mechanisms compared to the links presented in NPT or MUSIQ where common elements exist. Our changes include assigning link polarity (+/-), highlighting significant delays, removing duplicate causal chains and adding new links. With these improvements, we can consider the impact of important interdependencies (i.e., causal feedback loops) between an intervention and its context on the progression of a specific implementation over time (i.e., dynamics) in a manner not yet explored in these existing literature streams. Our CLD can serve as a valuable tool for leaders and their teams to identify key drivers of quality and anticipate consequences of changes in these drivers. To evaluate the external validity of our CLD, future research could employ secondary analysis of qualitative from studies data in NICUs, ICUs, and other relevant contexts, as was done in the development of NPT and MUSIQ. Such research can be informed by the analysis presented in Appendix 3. Furthermore, future research should explore expanding the boundary of our CLD to shed light on the intricate interdependencies among NICU processes and their interactions with other relevant hospital units.

In general, future implementation research in healthcare should explicitly consider the causal feedback loops in the theories and methods used to design interventions and account for context, by visualizing these loops transparently in CLD form. This will enable replication, application, comparison, and inspection, which is crucial considering the widely-accepted significance of feedback-oriented frameworks such as NPT, MUSIQ, high reliability, plan-do-study-act, and rapid learning in applied work to enhance healthcare organizations. By explicitly

considering feedback loops, presenting their results in CLD form, future research on implementing improvement initiatives situations studies can facilitate replication, application, and inspection. Even without further testing for external validity, our CLD offers several advantages over these frameworks. There are several issues experienced by NPT users, including handling overlapping constructs and causal loops (May, Albers et al. 2021), and of coding chunks of experience data where constructs are interdependent (Atkins, Lewin et al. 2011, McEvoy, Ballini et al. 2014)(p.10) which our CLD's greater detail could prevent. Additionally, this detail improves understanding of how implementation plays out over time, a persistent critique of NPT (Alharbi, Carlström et al. 2014, Alverbratt, Carlström et al. 2014, May, Cummings et al. 2018). Our CLD offers an attractive alternative to MUSIQ 2.0(Reed, Kaplan et al. 2018) for describing the dynamics of context, especially for situations involving team-based QI efforts.

Two established frameworks for service operations and implementation of QI have also been developed by scholars using SD, but have not yet been referenced in the peer-reviewed SD literature in healthcare (Darabi and Hosseinichimeh 2020). Here, we refer to these frameworks as Service Operations (Rust 2013) and Capability Traps (Repenning and Sterman 2002). While our CLD is inductively developed in one case study, these frameworks are based on accumulated findings from multiple prior studies across multiple industries, with Service Operations being generalized only within service industries (Anderson Jr, Morrice et al. 2005, Rust 2013). When compared to these models broadly, our CLD provides less detail into some links and includes a broader system boundary (i.e., more variables and feedbacks). All three share the core *Capacity Management* and *Working Harder* feedback loops. Two of the models, Capability Traps and our CLD, share the *Working Smarter* and *Reinvestment or Ruin* feedbacks. Our model includes additional feedbacks with QI capabilities, motivation for QI and parental involvement. Nevertheless, our CLD does contrast with these models in that it considers aspects of quality that are particular to healthcare (i.e., more patients with fewer harms rather than just more widgets produced, or services provided). While these additional loops in our CLD help us make sense of UBA in its context, future research is needed to explore the extent to which these structures exist in other healthcare services and QI situations, as well as service situations in other industries where front-line staff's QI capabilities matter.

Safety in healthcare is often tied to the concept of High Reliability Organizations. For example, UBA uses RPI, which has been tied very closely in the literature to this notion (c.f., (Chassin and Loeb 2013)). As with Lean and other concepts, healthcare scholars are interested in better understanding whether high reliability concepts apply to healthcare, because the theory was developed from cases in other industries. Specifically, Tolk *et al*.(Tolk, Cantu et al. 2015) propose that tight coupling may not exist in healthcare and suggest that ICUs are a healthcare situation which could clearly test this assumption. Future research testing this hypothesis would benefit from considering the feedbacks proposed in this paper, as well as incorporating insights from prior research on feedback-rich SD work in NICUs and ICUs (c.f., (Demir, Lebcir et al. 2014, Mahmoudian-Dehkordi and Sadat 2017) and in safety high reliability theory (c.f., (Cooke and Rohleder 2006)).

Conclusion

Organizational improvement is often based on the idea of a causal feedback loop as a solution to a complex problem (e.g., *plan-do-study-act* in QI, *double loop learning* in rapid learning). The problem is a black box with an unknown structure and these feedbacks slowly make sense of it, removing defects and solving difficulties one at a time while simultaneously adapting changes to an evolving context. Through inductive analysis of semi-structured interviews, ongoing stakeholder engagement, and iteration between rigorous diagramming and simulation with numeric operations data, we have discovered empirically valid feedbacks both in and between an intervention and its context. In this paper, we visualize these interdependencies, thus enabling a deeper understanding of what it takes to sustain QI. This understanding promotes the notion that, before blaming the person, consider how the system holds shared responsibility for important problems and the changes that it needs to make in the future to prevent them.

Our comparisons with two theories of complex interdependencies in healthcare demonstrate that complex interdependencies in organizational context need not be infinitely complex or entirely case dependent. Rather, a moderate amount of feedback structure can be sufficient to explain much of the specific dynamics occurring across various situations.

In its initial implementation, UBA succeeded on two important fronts: QI and meeting staff's high expectations for overcoming the failings of traditional QI. Nevertheless, our research suggests that sustaining UBA's success will require innovations in management and systems design.

Ethics approval and consent to participate

Ethical review was provided by University of Texas Health Science Center at Houston Internal Review Board. This study was approved (HSC-MS-16-0542). Consent was obtained prior to each interview.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the Agency for Healthcare Research and Quality (Grant # 1P30HS024459-01, PI: Eric Thomas). The views expressed in this paper are solely those of the authors and do not represent those of any U.S. government agency or any institution with which the authors are affiliated.

References

Adrian, N. (2009). "Don't just talk the talk." Quality Progress **42**(7): 30-33.
AHRQ (2022). Module 14. Creating Quality Improvement Teams and QI Plans. Practice Facilitation Handbook.
Al-Zuheri, A., I. Vlachos and Y. Amer (2021). "Application of Lean Six Sigma to Reduce Patient Waiting Time: Literature Review." International Journal for Quality Research **15**(1).

Alharbi, T. S., E. Carlström, I. Ekman and L.-E. Olsson (2014). "Implementation of person-centred care: management perspective." J Hosp Adm **3**(3): 107-120.
Alverbratt, C., E. Carlström, S. Åström, A. Kauffeldt and J. Berlin (2014). "The process of implementing a new working method-a project towards change in a Swedish psychiatric clinic." J Hosp Admin **3**(6): 174.
Andersen, D. L., L. F. Luna-Reyes, V. G. Diker, L. Black, E. Rich and D. F. Andersen (2012). "The disconfirmatory interview as a strategy for the assessment of system dynamics models." System Dynamics Review **28**(3): 255-275.
Anderson Jr, E. G., D. J. Morrice and G. Lundeen (2005). "The "physics" of capacity and backlog management in service and custom manufacturing supply chains." System Dynamics Review: The Journal of the System Dynamics Society **21**(3): 217-247.
Atkins, S., S. Lewin, K. C. Ringsberg and A. Thorson (2011). "Provider experiences of the implementation of a new tuberculosis treatment programme: a qualitative study using the normalisation process model." BMC Health Services Research **11**(1): 1-12.
Bai, J., A. Fügener, J. Schoenfelder and J. O. Brunner (2018). "Operations research in intensive care unit management: a literature review." Health Care Management Science **21**(1): 1-24.
Barry, K. (2010). "Six Sigma Tools & Templates: Causal Loop Diagram." https://www.isixsigma.com/wp-content/uploads/2010/02/Causal-Loop-Diagram.gif.
Bauer, M. S. and J. Kirchner (2020). "Implementation science: What is it and why should I care?" Psychiatry Research **283**: 112376.
Braun, V. and V. Clarke (2006). "Using thematic analysis in psychology." Qualitative research in psychology **3**(2): 77-101.
Chassin, M. R. and J. M. Loeb (2013). "High-reliability health care: getting there from here." The Milbank Quarterly **91**(3): 459-490.
Cooke, D. L. and T. R. Rohleder (2006). "Learning from incidents: from normal accidents to high reliability." System Dynamics Review **22**(3): 213-239.
Darabi, N. and N. Hosseinichimeh (2020). "System dynamics modeling in health and medicine: a systematic literature review." System Dynamics Review **36**(1): 29-73.
DelliFraine, J. L., J. R. Langabeer, 2nd and I. M. Nembhard (2010). "Assessing the evidence of Six Sigma and Lean in the health care industry." Qual Manag Health Care **19**(3): 211-225.
Demir, E., R. Lebcir and S. Adeyemi (2014). "Modelling length of stay and patient flows: methodological case studies from the UK neonatal care services." Journal of the Operational Research Society **65**(4): 532-545.
Doyle, J. K., M. Radzicki and W. Trees (2008). "Measuring change in mental models of complex systems." Complex decision making: Theory and practice: 269-294.
Dubois, C.-A. and D. Singh (2009). "From staff-mix to skill-mix and beyond: towards a systemic approach to health workforce management." Human Resources for Health **7**(1): 87.
Fetters, M. D., L. A. Curry and J. W. Creswell (2013). "Achieving Integration in Mixed Methods Designs—Principles and Practices." Health Services Research **48**(6pt2): 2134-2156.
Forrester, J. W. (1980). "Information Sources for Modeling the National Economy." Journal of the American Statistical Association **75**(371): 555-566.
Glasgow, J. M., J. R. Scott-Caziewell and P. J. Kaboli (2010). "Guiding inpatient quality improvement: a systematic review of Lean and Six Sigma." Jt Comm J Qual Patient Saf **36**(12): 533-540.

Grossmann, C., W. A. Goolsby, L. Olsen, J. M. McGinnis and Institute of Medicine (US) Roundtable on Value & Science-Driven Health (2011). Engineering a learning healthcare system: A look at the future: Workshop summary, National Academies Press (US).
Hallowell, D. L. (2016). "Causal Loop Diagrams: An Orientation in Software Context." https://www.isixsigma.com/tools-templates/causal-loop-diagrams-orientation-software-context/.
Hill, J. E., A.-M. Stephani, P. Sapple and A. J. Clegg (2020). "The effectiveness of continuous quality improvement for developing professional practice and improving health care outcomes: a systematic review." Implementation Science **15**(1): 23.
J. Liberatore, M. (2013). "Six Sigma in healthcare delivery." International Journal of Health Care Quality Assurance **26**(7): 601-626.
Kaplan, H. C., L. P. Provost, C. M. Froehle and P. A. Margolis (2012). "The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement." BMJ Qual Saf **21**(1): 13-20.
Kaufmann, U. H. and C. J. Chieh (2012). "Applying Systems Thinking to the Practice of Six Sigma." https://www.isixsigma.com/tools-templates/cause-effect/applying-systems-thinking-practice-six-sigma/.
Kopainsky, B. and L. F. Luna-Reyes (2008). "Closing the loop: promoting synergies with other theory building approaches to improve system dynamics practice." Systems Research and Behavioral Science **25**(4): 471-486.
Lane, D. C. (2001b). "Rerum cognoscere causas: Part II—Opportunities generated by the agency/structure debate and suggestions for clarifying the social theoretic position of system dynamics." System Dynamics Review **17**(4): 293-309.
Lebcir, R. and R. Atun (2020). "Should doctors use their judgment? How a system dynamics model elicited knowledge in neonatal care services." Journal of the Operational Research Society **71**(7): 1113-1123.
Lebcir, R. and R. Atun (2021). "Resources management impact on neonatal services performance in the United Kingdom: A system dynamics modelling approach." The International Journal of Health Planning and Management **36**(3): 793-812.
Luna-Reyes, L. F. and D. L. Andersen (2003). "Collecting and analyzing qualitative data for system dynamics: methods and models." System Dynamics Review **19**(4): 271-296.
Mahmoudian-Dehkordi, A. and S. Sadat (2017). "Sustaining critical care: using evidence-based simulation to evaluate ICU management policies." Health Care Management Science **20**(4): 532-547.
Martinez-Moyano, I. J. and G. P. Richardson (2013). "Best practices in system dynamics modeling." System Dynamics Review **29**(2): 102-123.
May, C., B. Albers, M. Bracher, T. L. Finch, A. Gilbert, M. Girling, K. Greenwood, A. MacFarlane, F. Mair and C. M. May (2021). "Evaluating Implementation Contexts, Mechanisms, and Outcomes: A Normalization Process Theory Coding Manual for Qualitative Research and Instrument Development."
May, C. R., A. Cummings, M. Girling, M. Bracher, F. S. Mair, C. M. May, E. Murray, M. Myall, T. Rapley and T. Finch (2018). "Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review." Implementation Science **13**(1): 80.

May, C. R., M. Johnson and T. Finch (2016). "Implementation, context and complexity." Implementation Science **11**(1): 141.
May, C. R., F. Mair, T. Finch, A. MacFarlane, C. Dowrick, S. Treweek, T. Rapley, L. Ballini, B. N. Ong and A. Rogers (2009). "Development of a theory of implementation and integration: Normalization Process Theory." Implementation Science **4**(1): 1-9.
McEvoy, R., L. Ballini, S. Maltoni, C. A. O'Donnell, F. S. Mair and A. MacFarlane (2014). "A qualitative systematic review of studies using the normalization process theory to research implementation processes." Implementation Science **9**(1): 1-13.
Miller, W. L., B. F. Crabtree, M. I. Harrison and M. L. Fennell (2013). "Integrating Mixed Methods in Health Services and Delivery System Research." Health Services Research **48**(6pt2): 2125-2133.
Pfeffer, J. and R. I. Sutton (2000). The knowing-doing gap: How smart companies turn knowledge into action, Harvard business press.
Rai, A. (2012). "Systems Thinking Results in Holistic Improvements." https://www.isixsigma.com/methodology/business-process-management-bpm/systems-thinking-results-holistic-improvements/.
Reed, J. E., H. C. Kaplan and S. A. Ismail (2018). "A new typology for understanding context: qualitative exploration of the model for understanding success in quality (MUSIQ)." BMC Health Services Research **18**(1): 584.
Repenning, N. P. and J. D. Sterman (1997). Working Paper: Getting quality the old-fashioned way: self confirming attributions in the dynamics of process improvement. Cambridge, MA, MIT.
Repenning, N. P. and J. D. Sterman (2001). "Nobody ever gets credit for fixing problems that never happened: creating and sustaining process improvement." California management review **43**(4): 64-88.
Repenning, N. P. and J. D. Sterman (2002). "Capability traps and self-confirming attribution errors in the dynamics of process improvement." Administrative Science Quarterly **47**(2): 265-295.
Rushing, W. (2012). "Causal Loop Diagrams: Little Known Analytical Tool." https://www.isixsigma.com/tools-templates/cause-effect/causal-loop-diagrams-little-known-analytical-tool/.
Rusoja, E., D. Haynie, J. Sievers, N. Mustafee, F. Nelson, M. Reynolds, E. Sarriot, R. C. Swanson and B. Williams (2018). "Thinking about complexity in health: A systematic review of the key systems thinking and complexity ideas in health." Journal of evaluation in clinical practice **24**(3): 600-606.
Rust, T. (2013). Dynamic Analysis of Healthcare Service Delivery: Application of Lean and Agile Concepts.
Schaffernicht, M. F. G. and S. N. Groesser (2011). "A comprehensive method for comparing mental models of dynamic systems." European Journal of Operational Research **210**(1): 57-67.
Sedlock, E. W., M. Ottosen, K. Nether, D. F. Sittig, J. M. Etchegaray, A. Tomoaia-Cotisel, N. Francis, L. Yager, L. Schafer and R. Wilkinson (2018). "Creating a comprehensive, unit-based approach to detecting and preventing harm in the neonatal intensive care unit." Journal of Patient Safety and Risk Management **23**(4): 167-175.

Shire, M. I., G. T. Jun and S. Robinson (2018). "The application of system dynamics modelling to system safety improvement: Present use and future potential." Safety Science **106**: 104-120.
Sitkin, S. B., K. M. Sutcliffe and R. G. Schroeder (1994). "Distinguishing control from learning in total quality management: a contingency perspective." Academy of management review **19**(3): 537-564.
Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World, Irwin/McGraw-Hill.
Sterman, J. D. (2006). "Learning from evidence in a complex world." American journal of public health **96**(3): 505-514.
Sterman, J. D., R. Oliva, K. Linderman and E. Bendoly (2015). "System dynamics perspectives and modeling opportunities for research in operations management." Journal of Operations Management **39–40**: 1-5.
Tolk, J. N., J. Cantu and M. Beruvides (2015). "High Reliability Organization Research: A Literature Review for Health Care." Engineering Management Journal **27**(4): 218-237.
Tomoaia-Cotisel, A. (2018). The Journey toward the Patient-Centered Medical Home: A Grounded, Dynamic Theory of Primary Care Transformation. PhD, London School of Hygiene & Tropical Medicine.
Tomoaia-Cotisel, A., S. D. Allen, H. Kim, D. F. Andersen and Z. Chalabi (2022). "Rigorously Interpreted Quotation (RIQ) Analysis for Evaluating Causal Loop Diagrams in Late-Stage Conceptualization: Method Presentation and Case Illustration." System Dynamics Review.
Tomoaia-Cotisel, A., J. M. Etchegaray, E. W. Sedlock, M. Ottosen, D. F. Sittig, S. D. Allen and E. J. Thomas (2022). "Motivation, Time in the Day & External Resources Strengthen NICU Quality Improvement: A Case Study."
Tomoaia-Cotisel, A., R. MacDonald, S. D. Allen, J. M. Etchegaray, L. Yager, A. Khan and E. J. Thomas (2022). "Curbing Long Term Oscillation in NICU All Cause Harm: An Integrative Simulation-based Policy Analysis."
Tomoaia-Cotisel, A., D. L. Scammon, N. J. Waitzman, P. F. Cronholm, J. R. Halladay, D. L. Driscoll, L. I. Solberg, C. Hsu, M. Tai-Seale and V. Hiratsuka (2013). "Context matters: the experience of 14 research teams in systematically reporting contextual factors important for practice change." The Annals of Family Medicine **11**(Suppl 1): S115-S123.
Vest, J. R. and L. D. Gamm (2009). "A critical review of the research literature on Six Sigma, Lean and StuderGroup's Hardwiring Excellence in the United States: the need to demonstrate and communicate the effectiveness of transformation strategies in healthcare." Implementation Science **4**(1): 35.
Weick, K. E. (1993). "Organizational redesign as improvisation." Organizational change and redesign: 346-379.
Yearworth, M. and L. White (2013). "The uses of qualitative data in multimethodology: Developing causal loop diagrams during the coding process." European Journal of Operational Research **231**(1): 151-161.
Yin, R. K. (1994). Case Study Research: Design and Methods. Thousand Oaks, CA, SAGE Publications.