medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Evaluating quality improvement at scale: development of a reporting methodology for board-level insights in a UK mental health Trust

Kia-Chong Chua [1, 2 *], Barbara Grey [2], Michael Holland [2], Claire Henderson [1, 2], Nick Sevdalis [1, 2]

1. Centre for Implementation Science, Institute of Psychiatry, Psychology & Neuroscience, King's College London

2. Quality Improvement, South London & Maudsley NHS Foundation Trust

* corresponding author

Ethics statement:

According to the policy activities that constitute research at the South London & Maudsley NHS Foundation Trust, this work met criteria for operational improvement activities exempt from ethics review.

Data Availability Statement:

Data are available upon request from the corresponding author.

Funding:

NS' research is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) South London at King's College Hospital NHS Foundation Trust. NS and KCC are members of King's Improvement Science, which offers co-funding to the NIHR ARC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. Its work is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity and the Maudsley Charity. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, the ESRC or the Department of Health and Social Care.

COI statement:

NS is the director of the London Safety and Training Solutions Ltd, which offers training in patient safety, implementation solutions and human factors to healthcare organisations. The other authors have no conflicts of interest to declare.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background

The question of whether quality improvement (QI) actually improves quality in a large healthcare organisation requires overview of a diverse range of QI projects. To-date, an organisational overview remains elusive due to heterogeneity across project-specific outcomes. Where organisational-level assessments of QI success have been offered, the extent to which they could be attributed to QI activity remains unclear. Reporting on a retrospective evaluation of return on investment of QI in a UK National Health Service (NHS) Trust, we discuss the development of a reporting methodology for providing board-level insights on organisational needs and achievements in QI.

Methods

A researcher-in-residence worked with a resident QI team to develop and conduct a retrospective evaluation of QI projects in a large healthcare organisation in the UK that specialises in mental health. Using a survey, we assessed QI project outcomes and explored costs and benefits, as well as contextual, input and process factors that might be associated with whether QI projects led to a change in routine practice.

Findings

Out of 52 QI projects, 10 led to a change in routine practice. Across a diverse range of projects, time invested by staff is an opportunity cost that could serve as a common denominator for return on investment. Retrospective data accrual proved problematic for measuring benefits. Odds ratios from logistic regressions show that process factors had more apparent impact on project outcomes than contextual and input factors. Service user engagement and implementation of measurement plan showed major impact on project outcome. Of note, if QI projects had documentation for Plan-Do-Study-Act (PDSA) cycles, the odds of effecting a change in routine practice were much higher.

Conclusions

To inform organisational-level decision making and optimise return on investment from QI in large healthcare organisations, systematic monitoring of QI activity needs prospective data accrual, consistent measurement and coherent iterations. We recommend that this be undertaken by a resident QI unit with the data categories explored in this study. Developing a reporting methodology for routine organisational-level monitoring can be an asset for improving practice, one that is pivotal for a learning healthcare system.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Quality improvement (QI) in healthcare has seen a surge in investment. Significant financial and human resources have been developed and dedicated to QI – including the creation of specialist resident QI teams (Agency for Healthcare Research and Quality, 2013; Jones & Woodhead, 2015), national datasets that can help benchmark QI needs and achievements (Maggard-Gibbons, 2014; Raval & Pawlik, 2018), and the creation of 'researcher-in-residence' models in which clinical and academic institutions make shared investments to embed improvement scientists within the former (Marshall et al., 2014).

Despite significant investment, much of it made in times of severe financial austerity following the 2009 global financial crisis, there is conspicuous lack of evidence regarding the return on investment of improvement work within healthcare institutions. This impedes communication between QI leaders and executive boards regarding the extent to which QI is achieving its full potential within an organisation. It also hampers the ability to make a financial business case for the need to continue funding QI activities, hence putting such activities at significant risks when budget savings are sought (often the case in many healthcare systems globally).

Two notable attempts have recently been made on generating evidence of return on investment in QI. Swensen, Dilling, Mc Carty, Bolton, and Harper (2013) focused on four domains: patient needs, reputation, esprit de corps, and financial return. Along similar lines, Shah and Course (2018) identified six domains: patient, carer and family experience; staff experience; productivity and efficiency; cost avoidance; cost reduction and revenue. While both attempted to address the task of defining impact at scale, return on investment in one form or another was demonstrated using case studies (rather than organisation-wide QI implementations). Further, where organisational-level estimates were presented in these studies (e.g. revenue or staff engagement), the extent they could be attributed to QI at scale was unclear.

This conundrum is reflected in the broader literature on QI evaluation. Existing guidance is fundamentally centred on project-specific study design and outcomes to make a case study for adoption, spread and scaling up of a project (e.g., Ovretveit & Gustafson, 2002; Parry et al., 2018). While project-specific evaluation is necessary, it does not offer sufficient overview for healthcare leaders facing the full-scale of projects that employ disparate amount of resources with varying levels of impact and required to make prioritisation decisions. In fact, while QI has been advocated in healthcare for over 30 years, whether QI actually improves quality remains surprisingly difficult to answer (Dixon-Woods, 2019).

The study that we report here aims to begin addressing this conundrum. We report the development of a reporting methodology for providing board-level insights on organisational needs and achievements in QI. Specifically, we describe our experience in developing a retrospective evaluation of return on investment of a QI programme at scale, practical challenges that emerged, preliminary insights, and the lessons learned as we construct a reporting methodology for prospective monitoring of QI activities.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Method

Setting

The National Health Service (NHS) in the UK comprises organisational units, known as NHS Trusts, serving either a geographical area and/or a specialised function (e.g., mental health). A mental health NHS Trust in South London established a resident QI team in 2016 with mandate to foster a continuous improvement culture. The resident team supports QI projects led by frontline staff through training and coaching on QI methodology (e.g., Model for Improvement, driver diagrams, Plan-Do-Study-Act (PDSA) cycles). In partnership with an academic institution in 2016, the Trust also adopted a researcher-in-residence model that embeds an academic faculty in the resident QI team to support data and evaluation needs. With a steady rise of QI activities across the Trust between 2016 and 2018, the Trust executive board requested an evaluation of the Trust's QI programme to gauge the return on investment from these nascent developments.

Evaluation design process

A scoping exercise was first conducted by the researcher-in-residence (lead author) to develop a proposal of evaluation content. This involved individual consultations with three improvement specialists in the resident QI team, coupled with a rapid evidence scan, to gauge the breadth and depth of evaluation. Salient concerns that surfaced at this initial stage of consult included factors like staff turnover, service user involvement, and return on investment indications like whether project ideas were adopted and spread. The literature also flagged factors like fidelity of PDSA applications and implementation of data or measurement plan. A list of items was generated and circulated among the entire team of six improvement specialists in the resident QI unit for formative feedback on content relevance, acceptability and feasibility of the reporting methodology. This feedback was combined with further consultations with two scientists with academic expertise in implementation and improvement science. A revised version was then piloted with two colleagues back in the resident QI team before it was finalised as an online form for broader Trust-wide evaluation (see online supplement).

Project outcomes

To offer an overview of a diverse range of QI projects, we chose to look at whether projects achieved their aims, led to a change in routine practice (adoption), and triggered similar projects within the Trust beyond the site of their original conception (spread). These outcomes were considered core themes that are relevant for all projects despite heterogeneous and highly localised aims. Such a focus enables regular reporting at scale and acts as an early signal for areas where impact is lacking.

Costs and benefits

To attribute costs and benefits of QI projects within the Trust, the evaluation inquired on resource use (e.g. number of meetings with resident QI team). Considerable variation in documentation precluded project-specific economic analysis for return on

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

investment in the current retrospective evaluation, but a question was added on whether it was possible to quantify improvement in terms of cost savings and whether this had been attempted. QI is underpinned by an organisational management philosophy that recognises the critical need to empower frontline staff to learn and participate in continuous improvement in face of escalating complexity and change (Blumenthal & Kilo, 1998). On this basis, we also considered aspects of skill development and capacity building which may be attributed to each quality improvement project. Such a focus aligns with the philosophy of QI to move beyond performance management and cultivate an organisational culture of learning regardless of project outcomes. As a measure of skill development, the evaluation inquired about the extent and form of dissemination efforts (e.g., at / beyond local site or service and publication / conference dissemination). We also tracked capacity building by inquiring on whether project team members went on to develop more QI projects. Taken together, this information offers a broader picture of resource use to inform on the extent to which the resource commitment 'paid off'.

Contextual, input and process factors

To explore what may facilitate or impede QI project success, we inquired on contextual, input and process factors. Contextual factors refer to organisational conditions that are not within the influence of project teams. Besides setting (in-patient / community care), the resident QI team drew particular attention to contextual aspects like whether it was the team's first ever QI project, and whether protected time for QI activities was officially sanctioned. Input factors mainly included an inquiry on team characteristics (e.g. team profile and whether the project was disrupted by staff turnover). Process factors refer to actions or decisions of the project teams. They mainly include an inquiry on stakeholder engagement, PDSA cycles and measurement plans implemented.

Sampling

With approximately 200 QI projects initiated between 2016-18 within the study Trust, the task of retrospective information retrieval had to use a sample for feasibility reasons. To select a sample, each improvement specialist of the Trust's resident QI team was requested to identify up to five QI projects that they considered as 'successful' and up to another five that they considered as 'unsuccessful'. At this stage, we did not have an established definition of 'success' of a QI project (Morganti, Lovejoy, Haviland, Haas, & Farley, 2012). Instead, the improvement specialists were asked to rely on their own assessment of what did and did not work to maximise the gradient of contrast across the selection of QI projects. This in turn would also help surface insights on "work-as-done" rather than imposing a "work-as-imagined" criterion in our evaluation (Hollnagel, 2017).

Analysis

We first enumerated project outcomes in terms of whether they achieved their aims, introduced change ideas that were adopted in routine practice (adoption), and triggered similar projects at other sites (spread). We then compared projects that did and did not lead to a change in routine practice to see if they differ in terms of

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

contextual, input and process factors. To compare the associations between these factors and adoption, we compared effect sizes based on odds ratios (with confidence intervals from logistic regression). An odds ratio (OR) smaller than 1.5 was considered to be a small effect size, whereas OR > 5.0 was considered to be a large effect size (Chen, Cohen, & Chen, 2010).

Results

The study sample included 52 QI projects across five boroughs of London, UK served by the Trust (Table 1). Thirty projects were conceived by community mental health teams and the remaining (n=22) by inpatient care teams. Of the three themes of Trust Quality Priorities (South London and Maudsley NHS Foundation Trust, 2017: patient safety, clinical effectiveness, and patient experience), improving clinical effectiveness was the most common focus in project aims (29 / 52). A small handful focused on multiple priorities at once (6 community mental health and 7 inpatient care projects).

Project outcomes

In terms of project outcomes (Table 2), 18 of 52 projects (35%) reported a change in routine practice (adoption). However, only 10 of them reported formal project closure with aims achieved. Out of 7 of 52 projects (13%) that triggered similar projects in other sites (spread), only 3 reported formal project closure with aims achieved. A plausible explanation for these findings could be that some projects were Trust-wide initiatives that were adopted / spread across service sites regardless of the project outcome at specific sites. In light of this divergence between “work-as-imagined” and “work-as-done” (Hollnagel, 2017), we decided to retain the former as a more stringent definition of successful projects. We thus focused on the 10 projects which led to adoption after achieving their aims at formal closure. This offered a more interpretable benchmark for making comparisons with the remaining 42 projects.

Costs and benefits

Among the 10 successful projects, half required six or more months (median = 6.0) for completion. Those that were not successful after formal closure (n = 13) showed large variation. Half were completed in under three months (median = 2.8) but some took up to 12 months (Figure 1). Among projects that did not reach formal closure, those that terminated at the Planning stage of PDSA (n = 18, median = 1.8 months) showed a shorter life span than those that terminated in more advanced stages (n = 11, median = 6.0 months). Meetings with the Trust’s resident QI unit took place typically on a monthly basis for successful projects (median = 1.1 meetings monthly), a slightly higher rate than for all others (median = 0.6 – 0.9 meetings monthly). Monthly communication between project and Trust’s resident QI unit (email/phone) shows a similar picture, with slightly higher activity levels in successful projects (median = 7.5 vs 3.8 – 4.8 monthly communications).

Retrospective estimates were requested for the number of service users and staff who directly benefitted from the undertaken QI projects. This proved problematic as indications were available for only a handful of projects (8 / 10 successful projects vs 11 / 42 for all others). Similarly, when inquired about whether it was possible to

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

quantify improvement in terms of cost savings, more than half of the projects reported “not known”.

Among the 10 successful projects, seven disseminated publications of their findings (2 locally, 4 beyond local site/service, 1 not known). Among the remaining 42, five did so (2 locally, 2 beyond local site/service, 1 not known). For most projects, survey responses indicated “not known” in both respects. Two of the successful project teams went on to develop two new projects, whereas six in the latter group developed 10 new projects.

Contextual factors

We examined a range of factors that might be associated with QI project outcomes (Table 3). Seven (out of 30) QI projects in community mental health settings led to a change in routine practice. These odds were halved (Odds ratio, OR = 0.5, 95% CI: 0.1 – 2.3) for projects in inpatient care settings where three (out of 22) led to adoption. The odds of adoption were twice higher for QI projects that required funding (OR = 2.1, 95% CI: 0.4 – 10.4). However, among the five projects for which funding was available, none achieved adoption (consequently, we could not calculate an OR for comparing odds). The differences in odds hitherto did not attain statistical significance. The odds of adoption were five times lower (OR = 0.2, 95% CI: >0.1 – 0.9) if it was the team’s first attempt at carrying out a QI project. These odds were five times higher if protected time for QI activities was officially sanctioned for the project (OR = 5.2, 95% CI: 1.2 – 23.4).

Input factors

The odds of adoption increased slightly (OR = 1.4, 95% CI: 0.7 – 2.7) with team size but decreased if the project was led by non-clinical staff (OR = 0.6, 95% CI: 0.1 – 5.1) or non-managerial staff (OR = 0.8, 95% CI: 0.2 – 3.4). They were similar between QI projects that did or did not make a budget plan (OR = 1.1, 95% CI: 0.1 – 10.6). Staff turnover appeared to be slightly detrimental to the odds of adoption (OR = 0.9, 95% CI: 0.2 – 3.6). However, the effect sizes of these input factors generally indicated a small impact, and none attained statistical significance.

Process factors

The odds of adoption were higher if the project team engaged their team leader, stakeholders (e.g., staff members not in project team), and service users. Only service user engagement showed a statistically reliable impact, with a large effect size (OR = 7.4, 95% CI: 1.6 – 34.9).

The odds of adoption increased moderately with the number of outcome measures attached to the aim statement of the driver diagram produced by the project team (OR = 3.7, 95% CI: 1.1 – 9.8). This was also the case for the number of primary (OR = 2.7, 95% CI: 1.3 – 5.9) and secondary drivers (OR = 1.5, 95% CI: 1.1 – 1.9) in these diagrams.

Among 35 projects that quantified their target outcomes in the aims statement, 10 led to adoption. In the remaining 17 that did not quantify their target outcomes in the aims statement, none achieved adoption (consequently, we could not calculate an

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

OR for comparing odds). The odds of adoption were much higher if measures were tagged to the primary and secondary drivers (OR = 7.5, 95% CI: 1.7 – 33.7 and OR = 6.0, 95% CI: 1.3 – 27.2 respectively). Projects that included balancing measures also had much higher odds of adoption (OR = 7.4, 95% CI: 1.6 – 34.9).

The odds of adoption increased slightly with the number of PDSAs carried out (OR = 1.5, 95% CI: 1.1 – 2.2). Projects with PDSA that completed more than one cycle showed much higher odds of adoption (OR = 7.5, 95% CI: 1.7 – 33.7). The odds were even higher for projects that had PDSA documentation (OR = 85.5, 95% CI: 8.5 – 860.2). Despite the wide confidence interval, the lower bound interval estimate suggests that this latter aspect of PDSA had a major impact on project outcome.

Projects that collected data before implementing change ideas showed much higher odds of adoption (OR = 9.5, 95% CI: 1.9 – 47.6). This also applies to projects that established the median value of random variation in outcome measures (OR = 5.0, 95% CI: 1.1 – 22.0). We could not calculate OR for comparing odds for projects that established the median value of random variation in process and balancing measures because all that did so achieved adoption. Projects that collected data after implementing a change idea also show much higher odds of adoption (OR = 7.5, 95% CI: 1.7 – 33.7).

Discussion

This is the first study to our knowledge that aims to develop a reporting methodology for generating board-level insights on programmatic QI activity within a large healthcare delivery organisation, such as the mental health NHS Trust that we studied. The study yielded exploratory insights and a number of lessons were learnt from the practical challenges. We reflect on these with the view of informing what is needed for a fuller development of a return on investment evaluation.

We found that time invested in meetings, communications (virtual and face-to-face), and overall project duration are data with minimal information retrieval burden. Across a diverse range of projects, time invested by staff is an opportunity cost that could serve as a common denominator for return on investment.

In contrast, we encountered major challenges in measuring benefits. Cost savings is often not an immediate focus in QI work. Consequently, it is usually not part of the measurement plan. Improvement in patient safety and experience needs a reporting methodology that translates gains into cost savings.

The onus of developing structured reports is likely to fall on the resident QI unit due to their expertise (this was the case with the QI team of the study Trust). This will enable consistent measurement and coherent iterations. The level of skill development and capacity building in QI was not known for most projects in our retrospective evaluation. As with economic benefits, monitoring project team outcomes (as opposed to project outcomes) is best carried out by the resident QI unit.

A fundamental goal of QI is to bring about a change in routine practice (Batalden & Davidoff, 2007). Of the 52 QI projects led by frontline staff in a mental health NHS

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Trust, 10 achieved this goal. Among factors that reliably increased the odds of effecting a change in routine practice, service user involvement showed at least a moderate to large impact. This finding is consistent with the wider literature on QI in mental health (Robert, Hardacre, Locock, Bate, & Glasby, 2003). It also highlights the need for more granular data on the level of involvement of this stakeholder group, to guide efforts in optimising their potential contributions (Armstrong, Herbert, Aveling, Dixon-Woods, & Martin, 2013).

With the Model for Improvement as the dominant QI activity paradigm, the driver diagram is a common tool for problem recognition and analysis. If measures were tagged to primary drivers and balancing measures in the driver diagram, the odds of effecting a change in routine practice increased by at least a moderate to large effect size. This was also the case if data were collected before and after implementing a change. Our findings concur with existing improvement science literature. Measurement and the use of data is at the heart of the Model for Improvement (Boland, 2019). A transparent, data-driven approach is paramount (Birdas et al., 2019), otherwise project teams may get stuck in the 'Do' phase of a PDSA, or reach no actionable insight in the 'Study' phase.

As structured interventional experiments for testing changes, iterative cycles in PDSA are key for learning (Reed & Card, 2016). Improvement work is less likely to succeed if iterative cycles are too few (Ogrinc & Shojania, 2014). In our evaluation, we found that the odds of effecting a change in routine practice increases by at least a moderate to large effect size if projects reported PDSA that completed more than one cycle. The task of PDSA documentation by far show the strongest impact even by conservative estimates. Documentation of each stage of the PDSA cycle supports scientific quality, learning and reflection; even if PDSA cycles are well-executed, poor documentation would impede organisational memory and transferability of learning (Ogrinc & Shojania, 2014; Taylor et al., 2014). Documentation is a critical part of fidelity in PDSA. It is not a simple task. Training should not overemphasise the conceptual simplicity of PDSA (Reed & Card, 2016). Achieving high fidelity in PDSA application as a QI method will require a gradual and negotiated process to explore different perspectives and encourage new ways of working (McNicholas, Lennox, Woodcock, Bell, & Reed, 2019). Collecting routine data on PDSA fidelity can aid conversations in this effort. When PDSA cycles are performed and reported appropriately, knowledge is accumulated, which can be shared readily (Ogrinc & Shojania, 2014). Of note, QI is not synonymous with improving quality (Ross & Naylor, 2017). Even if there was no improvement, PDSA cycles with rigorous measurement plans would still generate learning. This would also be considered a QI success (Reed & Card, 2016).

Study limitations

To minimise burden on top of routine workload of the Trust's resident QI team, we collected retrospective data from a convenience sample of 52 QI projects. Consequently, our findings must be considered exploratory. The small sample size also meant that while some differences in odds are large and attained statistical significance, the confidence intervals are wide. The impact of the factors we

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

identified on QI project outcomes will require more precise estimates from larger studies, even though they showed statistically reliable associations with project outcome (i.e., adoption) in the present study. Retrospective recall difficulties in the absence of routine prospective monitoring also meant that we could not exclude missing data bias.

While several process factors showed an apparent impact on project outcomes, contextual and input factors generally showed little impact. But this does not mean that these latter aspects are not important. It is well-established in the improvement science literature that contextual factors play a prominent role (Kaplan, Provost, Froehle, & Margolis, 2012). However, the sheer number of variables and the unpredictability of their interactions make it hard to predict the distal impact of contextual factors on project outcomes (Braithwaite, 2018). Consequently, the need to minimise survey burden led to a narrow scope of contextual and input factors. Instead, we prioritise the scope of process factors (e.g. fidelity of data and measurement plans), to illuminate proximal influences that are potentially amenable to staff training and interventions.

Future operational and research development

Routine Trust-wide evaluation of a QI programme entails a reporting methodology that reaches beyond evaluation of specific QI projects. A key priority and challenge in such Trust-wide evaluations is to offer an accurate overview of a diverse range of projects. To this end, iterative development and learning is needed to identify challenges in defining return on investment. Operationally, it is clear that prospective monitoring is necessary to ensure systematic data accrual. To enable routine reporting, information retrieval burden may be alleviated by distributing data accrual in small bundles over time. A well-timed bundle optimises relevance and acceptability of data collection. A small bundle also minimises respondent burden (e.g., 5 - 10 minutes). Data can also be collected from different types of respondents (e.g., project team lead, sponsor, resident QI team). Well-targeted respondents will provide the most valid data. This will also ensure that only the most relevant questions are asked, thereby also minimising respondent burden.

An evaluation at scale that feeds regular reports for monitoring an organisation-wide ‘big picture’ of QI programme success (or otherwise) can generate strategic insights and inform decisions on resource commitments. Such a big picture also enables return on investment comparisons not only within but also between organisations, where organisational strategies and quality priorities differ. In a system as complex as healthcare, changing routine practice is hard (Braithwaite, 2018). Developing a reporting methodology for routine organisational-level monitoring can be an asset for improving practice (Dixon-Woods, 2019), one that is pivotal for a learning healthcare system.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

Agency for Healthcare Research and Quality. (2013). Creating Quality Improvement Teams and QI Plans. Retrieved from https://www.ahrq.gov/ncepcr/tools/pf-handbook/mod14.html

Armstrong, N., Herbert, G., Aveling, E. L., Dixon-Woods, M., & Martin, G. (2013). Optimizing patient involvement in quality improvement. *Health Expect, 16*(3), e36-47. doi:10.1111/hex.12039

Batalden, P. B., & Davidoff, F. (2007). What is "quality improvement" and how can it transform healthcare? *Qual Saf Health Care, 16*(1), 2-3. doi:10.1136/qshc.2006.022046

Birdas, T. J., Rozycki, G. F., Dunnington, G. L., Stevens, L., Liali, V., & Schmidt, C. M. (2019). "Show Me the Data": A Recipe for Quality Improvement Success in an Academic Surgical Department. *J Am Coll Surg, 228*(4), 368-373. doi:10.1016/j.jamcollsurg.2018.12.013

Blumenthal, D., & Kilo, C. M. (1998). A report card on continuous quality improvement. *Milbank Q, 76*(4), 625-648, 511. doi:10.1111/1468-0009.00108

Boland, B. (2019). Quality improvement in mental health services. *BJPsych Bulletin*, 1-5. doi:10.1192/bjb.2019.65

Braithwaite, J. (2018). Changing how we think about healthcare improvement. *BMJ, 361*, k2014. doi:10.1136/bmj.k2014

Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860-864. doi:10.1080/03610911003650383

Dixon-Woods, M. (2019). How to improve healthcare improvement-an essay by Mary Dixon-Woods. *BMJ, 367*, l5514. doi:10.1136/bmj.l5514

Hollnagel, E. (2017). Why is work-as-imagined different from work-as-done? In R. L. Wears & E. Hollnagel (Eds.), *Resilient Health Care* (Vol. 2). London: CRC Press.

Jones, B., & Woodhead, T. (2015). Building the foundations for improvement. Retrieved from https://www.health.org.uk/publications/building-the-foundations-for-improvement

Kaplan, H. C., Provost, L. P., Froehle, C. M., & Margolis, P. A. (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. doi:10.1136/bmjqs-2011-000010

Maggard-Gibbons, M. (2014). The use of report cards and outcome measurements to improve the safety of surgical care: the American College of Surgeons National Surgical Quality Improvement Program. *BMJ Qual Saf, 23*(7), 589-599. doi:10.1136/bmjqs-2013-002223

Marshall, M., Pagel, C., French, C., Utley, M., Allwood, D., Fulop, N., . . . Goldmann, A. (2014). Moving improvement research closer to practice: the Researcher-in-Residence model. *BMJ Qual Saf, 23*(10), 801-805. doi:10.1136/bmjqs-2013-002779

McNicholas, C., Lennox, L., Woodcock, T., Bell, D., & Reed, J. E. (2019). Evolving quality improvement support strategies to improve Plan-Do-Study-Act cycle fidelity: a retrospective mixed-methods study. *BMJ Qual Saf, 28*(5), 356-365. doi:10.1136/bmjqs-2017-007605

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Morganti, K. G., Lovejoy, S., Haviland, A. M., Haas, A. C., & Farley, D. O. (2012). Measuring success for health care quality improvement interventions. *Med Care, 50*(12), 1086-1092. doi:10.1097/MLR.0b013e318268e964

Ogrinc, G., & Shojania, K. G. (2014). Building knowledge, asking questions. *BMJ Qual Saf, 23*(4), 265-267. doi:10.1136/bmjqs-2013-002703

Ovretveit, J., & Gustafson, D. (2002). Evaluation of quality improvement programmes. *Qual Saf Health Care, 11*(3), 270-275. doi:10.1136/qhc.11.3.270

Parry, G., Coly, A., Goldmann, D., Rowe, A. K., Chattu, V., Logiudice, D., . . . Nambiar, B. (2018). Practical recommendations for the evaluation of improvement initiatives. *Int J Qual Health Care, 30*(suppl_1), 29-36. doi:10.1093/intqhc/mzy021

Raval, M. V., & Pawlik, T. M. (2018). Practical Guide to Surgical Data Sets: National Surgical Quality Improvement Program (NSQIP) and Pediatric NSQIP. *JAMA Surg, 153*(8), 764-765. doi:10.1001/jamasurg.2018.0486

Reed, J. E., & Card, A. J. (2016). The problem with Plan-Do-Study-Act cycles. *BMJ Qual Saf, 25*(3), 147-152. doi:10.1136/bmjqs-2015-005076

Robert, G., Hardacre, J., Locock, L., Bate, P., & Glasby, J. (2003). Redesigning mental health services: lessons on user involvement from the Mental Health Collaborative. *Health Expect, 6*(1), 60-71. doi:10.1046/j.1369-6513.2003.00209.x

Ross, S., & Naylor, C. (2017). Quality improvement in mental health. *The King's Fund.* Retrieved from https://www.kingsfund.org.uk/publications/quality-improvement-mental-health

Shah, A., & Course, S. (2018). Building the business case for quality improvement: a framework for evaluating return on investment. *Future Healthc J, 5*(2), 132-137. doi:10.7861/futurehosp.5-2-132

South London and Maudsley NHS Foundation Trust. (2017). *Quality Account 2016/2017*. Retrieved from https://m.slam.nhs.uk/media/481802/slam_quality_report_201617.pdf

Swensen, S. J., Dilling, J. A., Mc Carty, P. M., Bolton, J. W., & Harper, C. M., Jr. (2013). The business case for health-care quality improvement. *J Patient Saf, 9*(1), 44-52. doi:10.1097/PTS.0b013e3182753e33

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ Qual Saf, 23*(4), 290-298. doi:10.1136/bmjqs-2013-001862

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Study sample of 52 quality improvement projects in South London and Maudsley NHS Foundation Trust (2016 – 2018)

	Community Mental Health n = 24 (6)*	Inpatient n = 15 (7)*
Borough		
Croydon	8	10
Lambeth	7	4
Lewisham	7	4
Southwark	7	3
Wandsworth	1	1
Patient Safety		
Reduce use of restrictive interventions on service users	1 (2)	0 (3)
Safer staffing	0	1 (1)
Risk assessments	1 (1)	1 (1)
Clinical Effectiveness		
Physical healthcare screening	4 (1)	4 (3)
Care planning	10 (5)	6 (3)
Developing electronic systems to improve care delivery	3 (1)	2 (1)
Patient Experience		
Reducing number of acute out-of-area treatments	3 (2)	0 (2)
Carer's assessment and associated care plan	1 (3)	0 (1)
Quality of environment and food	1 (1)	1 (2)

* numbers in parentheses refers to number of projects that included multiple Quality Priorities in project aims

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Project outcomes

		Adopted (n = 18)	Spread (n = 7)
Formal closure (n = 23)	Aims achieved (n = 12)	10	3
	Aims not achieved (n = 11)	3	2
Terminated (n = 29)	Aims achieved (n = 4)	4	1
	Aims not achieved (n = 25)	1	1

[+] Project Status: Closed, Terminated at xP (Plan), xD (Do), xS (Study), xA (Act)
Adopted = Change idea adopted, Spread = Triggered similar projects

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

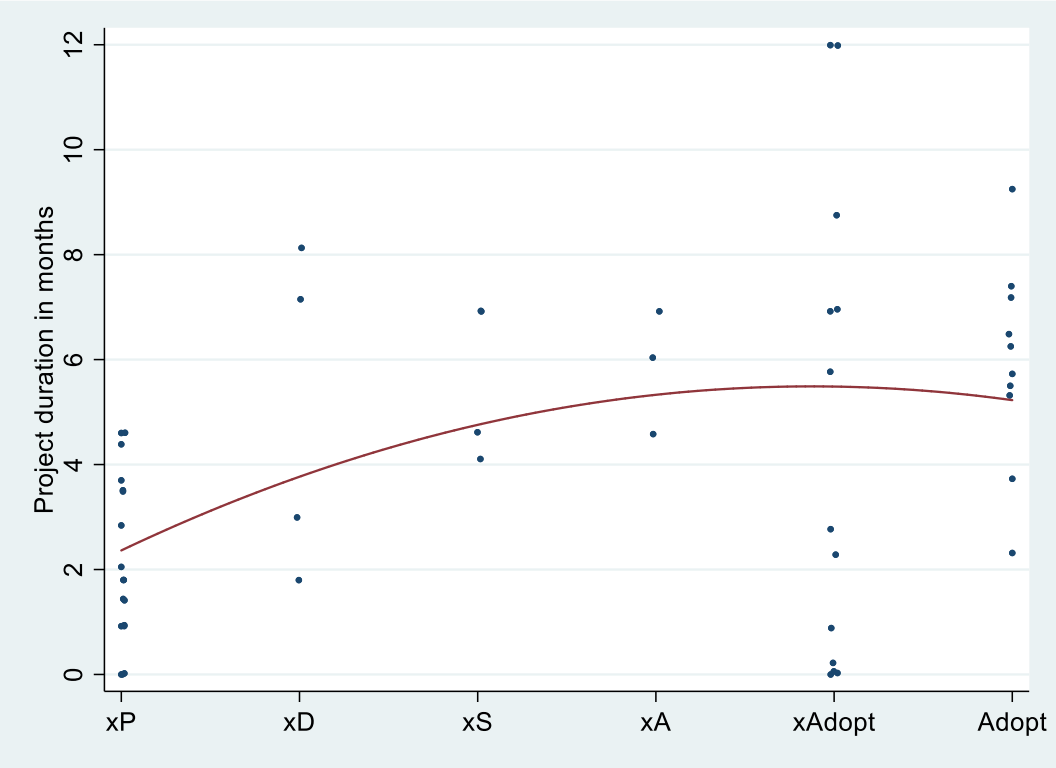

Figure 1. Duration (months) of projects that terminated in planning (xP), doing (xD), studying (xS) and Acting (xA) stage of PDSA cycle, as well as projects that had formal closure with and without a change in routine practice (xAdopt / Adopt).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.13.20022475; this version posted March 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3. Logistic regression models for association between project outcome of adoption (dependent variable) and contextual, input, and process factors (independent variables).

	n_1	n_2	Odds Ratio	95% CI
Contextual factors				
Inpatient	22	3	0.5	0.1 – 2.3
First QI project +	27	2	0.2	>0.1 – 0.9
Time officially sanctioned +	20	7	5.2	1.2 – 23.4
Funding available	5	0	#	#
Funding required	10	3	2.1	0.4 – 10.4
Input factors				
Made budget plan	5	1	1.1	0.1 – 10.6
Project lead: non-clinical staff	8	1	0.6	0.1 – 5.1
Project lead: non-managerial staff	18	3	0.8	0.2 – 3.4
Team size	-	-	1.4	0.7 – 2.7
Staff turnover	22	4	0.9	0.2 – 3.6
Process factors				
Engaged team leader	28	6	1.4	0.3 – 5.5
Engaged stakeholders	14	5	3.7	0.9 – 15.5
Engaged service users ++	10	5	7.4	1.6 – 34.9
No. of outcome measures +	-	-	3.2	1.1 – 9.8
No. of primary drivers +	-	-	2.7	1.3 – 5.9
No. of secondary drivers +	-	-	1.5	1.1 – 1.9
Aims quantified	35	10	#	#
Primary drivers tagged with measures ++	13	6	7.5	1.7 – 33.7
Secondary drivers tagged with measures +	11	5	6.0	1.3 – 27.2
Balancing measures ++	10	5	7.4	1.6 – 34.9
No. of PDSA completed +	-	-	1.5	1.1 – 2.2
PDSA that completed >1 cycle ++	13	6	7.5	1.7 – 33.7
PDSA with documentation ++	13	9	85.5	8.5 – 860.2
Data collected before implementing change idea ++	9	5	9.5	1.9 – 47.6
Median value of random variation in outcome measures +	12	5	5.0	1.1 – 22.0
Median value of random variation in process measures	4	4	#	#
Median value of random variation in balancing measures	2	2	#	#
Data collected after implementing change idea ++	13	6	7.5	1.7 – 33.7

n_1: total number of projects that satisfy the condition described by the independent variable
n_2: total number of projects (in n_1) that led to a change in routine practice (adoption).
independent variables for which odds of project outcome could not be calculated
+ independent variables that show statistically significant odds ratio
++ independent variables for which conservative estimates (lower bound of 95%CI) show at least a moderate effect size (OR > 1.5, or in opposite direction: OR < 0.7)