QualDash: Designing and evaluating an interactive dashboard to improve quality of care

Protocol version number and date: 2.0, 11/04/2018

IRAS number: 227139
Funder's number: 16/04/06
Project duration: 33 months (October 1st 2017 to June 30th 2020)

Principal Investigator: Dr Rebecca Randell, School of Healthcare, University of Leeds

Version control

Version	Date	Author	Edits
1.0	03/4/2017		Approved by University of Leeds School of Healthcare Research Ethics Committee (ref: HREC16-044)
2.0	11/04/2018		Addition of observation of meetings in Phase 2

Acknowledgement:
This project is funded by the NIHR Health Services & Delivery Research Programme (project number: 16/04/06). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

SUMMARY OF RESEARCH

Background: Quality dashboards provide visualisations of quality data which can lead to the identification of previously unnoticed patterns in that data,[1] informing quality improvement (QI) initiatives, and more efficient and effective decision making.[2]

Aim: To develop and evaluate QualDash, a web-based quality dashboard for exploring National Clinical Audit (NCA) data. To ensure users can interact with and view audit data in ways that are meaningful to them,[3] QualDash will be interactive and tailored to meet the needs of different users. Through the process of individualisation and using novel visualisation techniques, QualDash will support clinical teams, quality sub-committees, NHS Trust boards, and commissioners to better understand and make use of NCA data to improve quality of care and clinical outcomes.

Methods: The project is based on MRC guidance for design and evaluation of complex interventions.[4] To ensure that QualDash has a robust theoretical basis and to enhance the probability of its widespread implementation, we will combine the principles of realist evaluation (which involves building, testing and refining the theories of how and in what contexts an intervention works)[5] and principles of co-design.[6,7] Our project is comprised of five phases:

Phase 1: Interviews with members of clinical teams, quality sub-committees, and boards across five NHS acute Trusts and relevant commissioners will be used to articulate how NCA data are currently used (or not) in practice, identifying blockages to effective use and how these might be overcome. Interviews will consider a range of NCAs but will focus on the Myocardial Ischaemia National Audit Project (MINAP) and the Paediatric Intensive Care Audit Network (PICANet). Initial requirements for the design of QualDash derived from the interview data will be discussed at a workshop with suppliers of other NCAs to determine which requirements are generalisable to all NCAs.

Phase 2: QualDash will be developed through an iterative process, involving focus groups with clinical teams, quality sub-committee members, and board members from one Trust, relevant commissioners, and patients and carers. A controlled user experiment will assess comprehension, usability, and acceptability of QualDash prototypes, in comparison with existing formats for feedback of NCA data.

Phase 3: An implementation strategy for QualDash, tailored to the five Trusts and relevant Clinical Commissioning Groups (CCGs), will be developed through focus groups with clinical teams, quality sub-committees, and boards from the five Trusts and relevant commissioners. Planned implementation activities will be delivered across the five Trusts and with relevant commissioners.

Phase 4: QualDash will be made available in the five Trusts and relevant CCGs. A controlled interrupted time series (CITS) analysis will investigate impacts of QualDash, using process measures for MINAP and PICANet. Ethnographic observations and interviews over 12 months will provide insight into contexts and mechanisms that lead to those impacts. A questionnaire will be used to gather data on perceived usefulness of QualDash.

Phase 5: The feasibility of conducting a cluster randomised controlled trial (CRT) of QualDash will be assessed. If progression criteria are met, a CRT will be designed, using the CITS results to decide what effect size the trial should be powered to detect. Two focus groups will explore the suitability of QualDash for a range of other NCAs.

Phase 1 will provide understanding of how NCA data are currently used (or not) by various health service audiences (topic i of the commissioning brief), while Phases 2-5 will test the principles of effective communication of data to key audiences and provide the basis for a future trial (topic ii). We have a multidisciplinary team, including expertise in the visualisation and analysis of large data sets, addressing the requirement for research that provides an understanding of how academics from such fields and health service researchers can work with NCA suppliers to improve NCA outputs (topic a). By working with our Lay Advisory Group (LAG) in Phase 2, we begin to explore the most effective ways to present NCA data to the public (topic c).

BACKGROUND AND RATIONALE

In the UK, the Healthcare Quality Improvement Partnership (HQIP) centrally develops and manages a programme of over 60 NCA projects each year through the National Clinical Audit and Patient Outcomes Programme (NCAPOP). In addition, there are over 50 independent NCAs. The aim of the NCAs is not just to systematically measure the quality of care delivered by clinical teams and provider organisations but to use those data to stimulate QI.[8] While there is evidence of positive impacts of NCAs[9-11] and clinicians consider the annual reports that NCAs produce as important for identifying QI opportunities,[12] there is variation both within and between NHS Trusts in the extent to which they engage with NCA data.[12,13] A number of NCAs provide Trusts with online access to more recent data and allow users to download data for local analysis. However, with a recognised shortage of data analysis skills within the NHS,[14,15] use of such data poses challenges for some Trusts.[13] Consequently, NCA data are substantially underutilised and the potential for NCA data to inform QI is not being realised. It is also unclear how data from the NCAs are being used by different users in Trusts, such as clinical teams, quality sub-committees, or boards.

Below, we consider what insight theories of audit and feedback (A&F) can provide into contextual factors that affect engagement with NCA data, before turning to the literature on dashboards to identify how and in what contexts a quality dashboard may support more effective use of NCA data.

Audit and feedback

NCAs represent a form of A&F. A&F is a QI strategy that involves generating a summary of clinical performance over a specified period of time and presenting that information to clinicians to enable them to assess and adjust their performance.[16] While the most recent Cochrane Review of 140 randomised trials of A&F interventions found a median

4.3% absolute increase in clinicians' compliance with recommended practice, there was significant heterogeneity in the results.[17] This has led to a number of calls for A&F research to focus on identifying the contextual factors that impact the success of A&F interventions.[18,19]

Theories of A&F suggest a number of such contextual factors. With their origins in psychology, such theories tend to focus on how A&F interventions are intended to change the reasoning and behaviour of individuals and thus operate at the individual, or 'micro', level. They are premised on the idea that clinicians have an intrinsic motivation to improve and that if clinicians are presented with evidence that there is a discrepancy between their own performance and that of a target or standard, this will lead to efforts to improve performance. These theories therefore focus on how characteristics of the A&F intervention itself might trigger an individual's intrinsic desire to improve. For example, Contextual Feedback Intervention Theory (CFIT) suggests the standard or target has to be perceived as desirable and achievable and feedback about the discrepancy between their performance and the standard has to be perceived as accurate.[20] CFIT contends that such feedback is more likely to change behaviour if it is timely, frequent, cognitively simple, e.g. presented graphically, unambiguous, and provides concrete suggestions of how to improve performance. Hysong et al.'s model of actionable feedback suggests clinicians are more likely to respond to audit data if it is perceived to be timely and non-punitive and if they receive feedback about their own individual performance rather than aggregated data about overall performance of the organisation.[3] If they are able to customise how they view the data, this leads to active engagement in sensemaking, further increasing the likelihood the data will be acted on.

While the A&F literature is largely concerned with feedback to individual clinicians,[21] the majority of NCAs provide feedback at the clinical team or provider level. Studies of national audits suggest A&F will only have an impact if clinical teams are given adequate time to engage with and use the data[13] and if they perceive they have the power to make changes,[22] emphasising the importance of contextual factors at the organisational, or 'meso', level. At this level, theories of benchmarking can provide useful insight. Benchmarking refers to the identification, sharing, and implementation of practices at the team or organisational level which lead to excellent performance.[23] Van Helden and Tilemma's model of benchmarking draws on institutional reasoning and resource dependency theories and hypothesises that an organisation's response to benchmarking is determined by different types of institutional pressures exerted on them by government, professional groups, interest groups, and the general public.[24] An organisation's response pattern, which can vary from passive compliance to pro-active manipulation of pressures, depends on its willingness and ability to conform to institutional pressures, which in turn is shaped by which stakeholders exert pressure and how, stakeholders' reasons for exerting pressure, what the pressures are, and the context in which the pressure is exerted.

That theories of benchmarking point to the role of government and other stakeholders is a reminder of the need to consider 'macro' level theories regarding wider social, cultural, and political factors.[25] A broad class of theories at this level focuses on governance arrangements, and in particular relationships between regulators (such as the Care Quality Commission (CQC), NHS Improvement, and CCGs), Trust managers, and clinicians. It includes theories of 'new professionalism' that draw on Foucauldian concepts of disciplinary power and governmentality.[26,27] Governmentality, which refers to forms of self-surveillance that ensure performance meets expectations, negating the need for managerial surveillance,[28] has been found to be relevant to understanding QI in the NHS. For example, Martin et al. revealed the importance of clinical leaders in championing quality and safety, and regular meetings which bring clinicians together to discuss quality and safety constructively.[27] This theoretical lens would suggest that professional community is important to the success of NCAs, with social obligations amongst members of the community offering opportunities for social control.[27,29] This class of theories also includes new institutional theories, which focus on the problems associated with traditional top-down regulation, and the emergence of regulation through networks of actors - what some have referred to as network governance.[30] More generally, engagement with NCAs is likely to be impacted by professional jurisdictions and status hierarchies, leading to different levels of engagement based on professional group, while political priorities will determine both the goals providers experience pressure to meet and the funds available for investing in QI.[31]

While there are theories regarding how A&F may be impacted by contextual factors at micro, meso, and macro levels, how these interact to produce improvements in quality is unclear. **There is a gap in the literature concerning how these theories translate to A&F interventions, such as NCAs, that are aimed at clinical teams and managers.**

Dashboards

Dashboards are a specific form of A&F. Originally developed for the business sector, in healthcare dashboards have been developed to either a) provide feedback to managers on performance on standardised quality/organisation metrics at ward or organisational level to inform operational decision making and QI efforts (quality dashboards)[32] or b) to provide feedback to individual clinicians on their performance compared to quality metrics to inform decisions about patient care (clinical dashboards).[33] Dashboards may be either paper-based or computer-based but the key characteristic of dashboards is their use of data visualisation techniques to provide feedback.[34] It is thought that use of such techniques improves data comprehension[35] and reduces cognitive load.[36]

In a recent systematic review of the evidence base for the use of computer-based quality and clinical dashboards undertaken by two of us (Dawn Dowding (DD), Rebecca Randell (RR)), 11 research studies that had evaluated their impact on either quality or clinical outcomes were identified.[34] Of the 11 studies, only one met the definition of a quality dashboard, which was designed to provide data for unit managers on quality measures to improve service delivery.[33] The other 10 dashboards that were evaluated were all clinical dashboards implemented across a variety of specialties including intensive care, acute and primary care settings. The studies had mixed results. Zaydfudim et al. found a

dashboard improved adherence to the ventilator bundle and was associated with a reduction in the rate of ventilator associated pneumonia[37] and Pablate found the introduction of a dashboard increased the timeliness of anaesthetists' preoperative antibiotic administration.[38] Both Linder et al. and Batley et al. found no significant difference in antibiotic prescribing rates after the introduction of a dashboard but provide data to suggest antibiotic prescribing rates improved amongst those clinicians who accessed the dashboard.[39,40] The review also highlighted heterogeneity in dashboard design (e.g. use of different graphs, user interfaces) and targeted users, meaning at present **it is unclear how dashboards can be designed effectively and integrated into the workflow of clinicians and managers to inform clinical and organisational decision making.**

Healthcare providers, both in the UK and internationally, are increasingly using dashboards as a way of measuring the quality of care they provide and as the basis for QI initiatives. The introduction of dashboards into the NHS has been encouraged by a number of policy documents. Lord Darzi's *Next Stage Review*[41] and the *Health Informatics Review*[42], both published in 2008, recommended greater use of dashboards by NHS organisations. Two distinct strands of thinking permeated these reviews and subsequent policies and guidance on dashboards. One emphasised the need for summary real-time information, for use by clinicians in their clinical work. The former NHS Connecting for Health agency encouraged these developments through its Clinical Dashboard Project.[43-45] More recently, the need for integrated real-time information on care quality, which could be interrogated by clinical teams for the purpose of learning, was emphasised in the Berwick report in 2013.[15] The other strand focused on the need for Trust boards and regulators to have summary performance information. NHS Foundation Trusts already had to publish performance dashboards, required by Monitor, but otherwise Trusts did not use dashboards for their own internal reporting, or to report to the Care Quality Commission or (as they were then) primary care Trusts. Developments in this area were prompted by two major reports in 2013, the second Francis report on substandard care provided at Mid Staffordshire NHS Foundation Trust[46] and the post-Francis Keogh review of 14 NHS Trusts with persistently high mortality rates.[14]

Two of us (Justin Keen (JK), RR) are undertaking the only empirical study of quality dashboard use in NHS acute Trusts in England. A survey of 15 Trusts found most had some form of quality dashboard in place for reporting measures such as those included in the NHS Safety Thermometer.[47] However, there was significant variation in the sophistication of those dashboards, with the term being used to refer both to IT systems and the outputs from these systems, typically in the form of printed reports. The majority of dashboards still depend on resource-intensive manual collation of information from a number of systems by central performance management teams, with retrospective reports then being circulated to wards, directorates and Trust boards. While such quality dashboards allow comparison of wards within a Trust, they do not allow clinical teams to compare their performance with clinical teams at other Trusts. At board level, we have found Trusts currently undertake relatively limited comparisons of their performance with that of other Trusts, focusing on mortality rates, and engagement with NCA data is limited to reporting within Trust board reports whether they are participating in particular NCAs. However, our work has also revealed a clear 'direction of travel', with a desire for real-time quality dashboards.

While the focus in the literature on clinical dashboards means we know little about mechanisms through which computer-based quality dashboards drive improvement, as a form of A&F quality dashboards share many of the same theories as other types of A&F interventions. A number of the A&F theories described above suggest that, by changing the context at the individual (micro) level, an interactive quality dashboard would lead to improvements in care quality, through **graphical presentation** of audit data[20] and **the ability for clinicians to view audit data in ways that are meaningful to them**.[3] While paper-based dashboards are inevitably static, both the computer-based quality dashboards we have identified in our study and other such quality dashboards that have been introduced into the NHS (e.g. Stethoscope, developed by Methods Analytics) present information in a standard format. The choice of measure and perceptions of its accuracy impact clinicians' motivation to improve their performance,[20] which is another contextual factor that an interactive quality dashboard can influence, **enabling access to a range of measures and allowing the user to focus on those that matter to them**. An interactive quality dashboard would also allow the integration of much richer data than is used at present, such as patient feedback to support the interpretation of quantitative audit data.

While quality dashboards can change elements of the context at the individual clinician (micro) level, their use and subsequent impact will also be influenced by contextual factors at the micro, organisation (meso), and policy (macro) level. In addition to those factors described in the theories of A&F described above, at the micro level, it is necessary to consider the relationship between the technology and work practice.[48] For example, Fit between Individuals, Task and Technology (FITT) is based on the idea that adoption of health IT (HIT) depends on the fit between the attributes of the individual users (e.g. computer anxiety, motivation), attributes of the technology (e.g. usability, functionality, performance), and attributes of the work practice (e.g. organisation, task complexity). The Technology Acceptance Model (TAM), widely used in HIT research, draws particular attention to perceived usefulness and perceived ease of use, contextual factors that influence a user's attitude towards a technology and consequently their intention to use it.[49] In considering the attributes of individual users, attention should be given to their ability to interpret different forms of data visualisation, with recent studies suggesting an individual's ability to understand information presented in dashboards is often determined by their level of graph literacy.[50,51] For example, Gaissmaier et al. found individuals with low graph literacy understood numerical information better if it was presented as numbers, whereas individuals with high graph literacy had better comprehension when the same information was presented in a graphical format.[50]

At the organisation (meso) level, with any form of HIT it is also necessary to consider the process through which the technology was developed and introduced. For example, Normalisation Process Theory suggests if clinical teams,

quality sub-committees, NHS Trust boards, and commissioners are able to 'make sense' of a quality dashboard, have been engaged in the process of implementation, have been able to adapt their work processes to incorporate the technology, and are able to identify potential benefits to its introduction, it is more likely to become embedded into practice, being used routinely and successfully to inform QI initiatives.[52-55]

At the policy (macro) level, we can again turn to the class of theories discussed above that focus on governance arrangements. A quality dashboard could potentially increase access to NCA data, enabling greater scrutiny of clinicians' practices by managers and Trust performance by commissioners.[26,27] If perceived in this way, clinicians' views regarding the appropriateness and legitimacy of this could serve to justify non-participation.[28] Thus, power relations between clinical teams, quality sub-committees, Trust boards, and CCGs form an important macro context, shaping how quality dashboards are perceived and used and subsequent impacts. There is also a second class of theories, focusing on technology, which is relevant at this level. UK-based researchers have drawn on a range of theories over the years, including structuration theory and practice theory,[56,57] but historical institutional approaches have come to dominate the field over the last decade.[58,59] They emphasise the extent to which current HITs are rarely self-contained but rather are linked to one another in 'e-infrastructures'.[60] Quality dashboards, and NCAs more generally, depend on NHS e-infrastructures that have evolved over time, such as the IT infrastructure that allows Trusts to upload large datasets to regulators and other bodies, and the working practices that have evolved with them.

In summary, existing research focuses on clinical dashboards, so **there is a gap in the literature concerning how and in what contexts quality dashboards lead to improvements in quality of care and clinical outcomes**. Quality dashboards have the potential to impact the micro level context of the feedback of audit data through use of visualisation techniques and integration of a range of sources of information, but the details of any design need to consider the ability of individual users to interpret different forms of data visualisation, reflect the concerns of intended users, involve users in the design process, and take existing governance and IT-infrastructures into account.

In response to both the gap in the evidence and theoretical issues summarised above, we propose to develop and evaluate QualDash, an interactive web-based quality dashboard for exploring NCA data. The research will answer the following questions:

A. In what contexts and by what mechanisms are NCA data currently used (or not) to improve quality of care?
B. What other data need to be presented alongside audit data to build a more complete picture of care quality?
C. What are the most effective means for clinical teams, quality sub-committees, boards, and commissioners to view and interact with audit data?
D. In what contexts and by what mechanisms do quality dashboards lead to increased engagement with audit data and subsequent improvements in care quality?

EVIDENCE EXPLAINING WHY THIS RESEARCH IS NEEDED NOW

Significant NHS resources, including clinicians' time, are invested in the delivery and running of NCAs. NCAs have the ability to make a major contribution to monitoring care quality in the NHS. However, differences both within and between Trusts in the extent to which they engage with NCA data mean NCA data are substantially underutilised and the potential for NCA data to inform QI is not being realised.[12,13] With a shortage of data analysis skills within Trusts,[14,15] there is recognition of the need for user-friendly means of supporting clinical teams and providers to analyse, interpret, and present findings, with indicated requirements including enabling clinicians to make comparisons with providers located nearby and those similar in terms of size of population, being able to 'drill down' from summary data to more detailed data, and having online access.[13] Discussions with HQIP have revealed they are currently working with CQC to develop paper-based quality dashboards on key measures for each NCA, to be used by CQC and Trust boards, and have future plans to develop quality dashboards that can be used by commissioners, although they are aware of the gap in the evidence relating to the impact of quality dashboards. Thus, the research proposed here, with its focus on a broader range of users and making data available in an interactive web-based format, complements work being undertaken by HQIP and provides evidence, on how and in what contexts quality dashboards work, that can inform the activities HQIP plan to undertake in future.

In addition to the immediate benefits to the NHS through the development of QualDash, evaluation of QualDash will provide an empirical test of existing A&F theories, generating knowledge of broader relevance concerning the design and implementation of quality dashboards. This is timely, given the current concern with reducing variations in care quality and clinical outcomes within the NHS,[61] with both the Francis Inquiry[46] and the Keogh review[14] highlighting the need for better monitoring of care quality at ward and board level and interest in technologies such as dashboards for this purpose.

This project addresses both key topics in the commissioning brief, providing understanding of how NCA data are used by various health service audiences (topic i) and testing the principles of effective communication of data to key audiences (topic ii). What is unique about the project is that the findings will be used to develop a novel web-based intervention, tailored to the needs of different user groups, which can be adapted for all NCAs.

AIMS AND OBJECTIVES

Aim: To develop and evaluate QualDash, an interactive web-based quality dashboard that supports clinical teams, quality sub-committees, NHS Trust boards, and commissioners to better understand and make use of NCA data, thereby leading to improved quality of care and clinical outcomes.

Objectives:

1. To develop a program theory that explains how and in what contexts use of QualDash will lead to improvements in care quality;
2. To use the program theory to co-design QualDash;
3. To use the program theory to co-design an implementation strategy for QualDash;
4. To understand how and in what contexts QualDash leads to improvements in care quality; and
5. To assess the feasibility of conducting a CRT of QualDash.

RESEARCH PLAN/METHODS

Design and theoretical/conceptual framework

This project uses the MRC guidance for design and evaluation of complex interventions, which emphasises that new interventions should have a coherent theoretical basis, with theory being used systematically in the process of development.[4] The need for use of theory in design and evaluation of A&F interventions,[18,19,62] and QI initiatives more generally,[29,63,64] is well recognised. This project draws on realist evaluation,[5] which involves building, testing, and refining the underlying assumptions or theories of how an intervention works, as well as wider literature on use of program theory for designing and evaluating interventions.[64,65] Realist evaluation has been recommended for studying QI,[66] offering a framework for understanding for whom and in what contexts complex interventions work.

While realist evaluation has been used for studying the implementation and impact of a number of complex interventions in healthcare,[67,68] including large-scale QI programmes,[69] there is growing acknowledgement of the value of realist approaches for design.[70-72] A realist approach enables the development of interventions that are responsive to specific aspects of context that impact their effectiveness and, through understanding how and why the intervention works, the adaptation of an intervention to other contexts is supported.[70]

In this project, we will combine a realist approach with co-design. The principles of realist evaluation and co-design have been demonstrated to be complementary.[70] In co-design, rather than passive recipients, the intended users are viewed as co-designers.[6] We will work with clinical teams, quality sub-committee members, Trust board members, commissioners, and patients and carers throughout the project, incorporating participants' knowledge from past experiences of working with NCA data, following Bate and Robert's four stage model for experience-based co-design: reflection, analysis, diagnosis, and description; imagination and visualisation; modelling, planning, and prototyping; and action and implementation.[7]

In **Phase 1**, interviews with members of clinical teams, quality sub-committees, and boards across five Trusts and relevant commissioners will be used to develop a program theory for QualDash, through exploring participants' theories about how and in what contexts NCA data are currently used (or not) to improve care quality and how QualDash could increase engagement with NCA data. This represents the reflection, analysis, diagnosis, and description co-design stage. In **Phases 2 and 3**, the QualDash program theory will provide the basis for the design of QualDash and an associated implementation strategy. This represents the imagination and visualisation and the modelling, planning, and prototyping co-design stages. The detail of the technology will be developed iteratively through focus groups with clinical teams, quality sub-committee members, and board members from one Trust, relevant commissioners, and patients and carers. Through this process, our understanding of the theories will be refined, adding detail regarding, for example, what data need to be presented, how, for whom, and in what contexts. Focus groups at each Trust will be used to develop an implementation strategy that is tailored to the local context of each Trust. In **Phase 4**, the QualDash program theory will be tested as QualDash is evaluated through a rigorous mixed methods evaluation. This represents the action and implementation co-design stage. In **Phase 5**, the feasibility of conducting a CRT of QualDash will be assessed, with the program theory, revised in light of the findings of Phase 4, being used to determine how and in what contexts QualDash should be evaluated in a CRT.

Setting/context

To ensure the study generates findings that are generalisable beyond a single NCA, we have chosen to work with two NCAs, MINAP and PICANet. MINAP and PICANet are delivered by different suppliers and involve different clinical specialties working with different patient groups, multiple professional groups (medical and nursing), and include multiple types of measures (structure, process, and outcome). They also differ in data quality and completeness. MINAP has been running continuously since 2000, delivered by the National Institute for Cardiovascular Outcomes Research, and data are contributed by all hospitals in England, Wales and Northern Ireland that admit patients with acute coronary syndromes. PICANet was set up in 2002 by the Universities of Leeds, Leicester and Sheffield and contains data from all NHS Paediatric Intensive Care Units (PICUs) in England and Wales.

There are three main funding arrangements for independent NCAs: subscription by NHS Trusts; funded by a charity or professional body; or funded by NHS England. To explore the impact of funding arrangements, in Phase 1 we will also gather data on the National Cardiac Arrest Audit (NCAA; funded by Trust subscription), the National Audit of Cardiac Rehabilitation (NACR; funded by the British Heart Foundation), and the Elective Surgery National PROMs Programme

(funded by NHS England). There are a number of independent surgery NCAs where participation is at the individual clinician, rather than Trust, level; to understand the impact of this difference, in Phase 1 we will gather data about the British Association of Urological Surgeons (BAUS) audits. In Phase 5, we will explore the applicability of QualDash to those independent NCAs where data is collected intermittently rather than continuously,

We will work with three NHS acute Trusts (sites A-C) that participate in both MINAP and PICANet. These Trusts have been selected to ensure variation in key outcome measures (MINAP: 30 day mortality for ST-elevation myocardial infarction (STEMI) patients; PICANet: risk adjusted standardised mortality ratio), as shown in Table 1.

Because Trusts that participate in PICANet tend to be larger and to be teaching hospitals, they are not representative of the range of Trusts that participate in MINAP. Therefore, MINAP use will also be studied in two district general hospitals (DGHs; sites D-E) that do not have a PICU. These have been selected to ensure variation in the same key MINAP measure.

To further ensure generalisability of our findings, we have selected Trusts with and without foundation status. Those Trusts with PICUs also vary in number of PICU patients treated per year. We have selected one Trust included in the Keogh review of 14 Trusts with persistently high mortality.[14]

Table 1: Case site characteristics

Site	Hospital type	MINAP: 30 day mortality for STEMI patients (%)	PICANet: risk adjusted standardised mortality ratio	Number of PICU patients per year
A	Teaching hospital	7.5	1.01	799
B	Teaching hospital	6.6	0.66	318
C	Teaching hospital, Foundation status	4.5	1.36	769
D	DGH	20.6		
E	DGH	8.3		

Phase 1: Situation analysis

Objectives

- To develop a program theory that explains how and in what contexts use of QualDash will lead to improvements in care quality (Objective 1); and
- To identify requirements for the design and implementation of QualDash.

Summary of method

Phase 1 addresses the question 'in what contexts and by what mechanisms are NCA data currently used or not to improve quality of care?' (Question A). A situation analysis identifies the nature and extent of the opportunities and problems to be addressed by an intervention and the context within which the intervention will operate,[65] referred to in co-design as the reflection, analysis, diagnosis, and description stage.[7] The situation analysis will involve interviews with members of clinical teams, quality sub-committees, and boards and relevant commissioners. By using the interviews to identify critical elements of context and mechanism that create and sustain the underutilisation of NCA data, we will be able to establish requirements for the design and implementation of QualDash, while also developing an understanding of the contextual factors that may support or hinder use of QualDash.[70] Effective program theories typically combine stakeholders' theories that are derived from experience with substantive theory.[5,64] Therefore, we will draw on theories from a range of disciplines across the social sciences, discussed earlier, that are concerned with the micro, meso and macro level in order to develop the findings from the interviews into a program theory for QualDash.

To assess the generalisability of the requirements and the QualDash program theory, a workshop will be held with suppliers of both NCAPOP and independent NCAs and representatives of HQIP. To identify which requirements are generalisable across NCAs, a variation of the nominal group technique will be used, based on methods used by one of us (RR) for a similar purpose[73] and involving card sort activities that have been recommended for the purpose of co-design.[7] The nominal group technique is a highly structured group process that can be used for identifying requirements and establishing priorities, that enables a substantial amount of work to be achieved in a relatively short space of time, providing immediate results with no requirement for further work.[74] The technique has previously been used successfully in the development of other complex interventions.[75]

Team

The majority of interviews will be undertaken by a Research Fellow (RF) and Research Assistant (RA) employed within the School of Healthcare (SoH). RR, Roy Ruddle (RAR), and the RF employed within the School of Computing (SoC) will also take part in a number of interviews, to hear directly the views of participants and for the purpose of relationship building.[7] The SoH RF and RA will be responsible for analysing the interview data. RR, Joanne Greenhalgh (JG), DD, and JK will provide methodological guidance for the design and analysis of the interview study. Chris Gale (CG), Mamas Mamas (MM), Roger Parslow (RP), and Julia Lake (JL), through their knowledge of how NCA data are used within the NHS, will contribute to the design of the interview topic guide and the interpretation of the findings. We will invite our LAG to review the interview topic guide, to ensure the topics explored include those that matter to patients/carers. The SoC RF will work with the SoH RF and RA to translate the findings into a requirements specification. The requirements

specification will be reviewed by the LAG for comprehensibility. During this time, the SoC RF will also be becoming familiar with the MINAP and PICANet data sets and exploring possible ways of presenting and interacting with these data.

The SoH RF and RA will be responsible for organising the workshop and will be assisted in running the workshop by RR, RAR, and the SoC RF. The SoC RF will be responsible for quantitative analysis of the data, while the SoH RF and RA will undertake qualitative analysis of workshop transcripts.

1a. Interviews across five Trusts

Sample: Given the intention of QualDash to support clinical teams, quality sub-committees, Trust boards, and commissioners, it is necessary to ensure all such groups are included. Our current dashboards study suggests recruitment of participants within these groups is feasible. A combination of purposive sampling and snowball sampling will be used. Experience from our current dashboards study has revealed variation across Trusts in who is responsible for and engages with audit data. Therefore, in each Trust, we will begin by interviewing the MINAP and PICANet leads, asking them to identify others it would be appropriate for us to interview. Using snowball sampling in this way will enable us to capture the implementation chain through which audit data are produced by certain stakeholders and accessed and analysed by others. Additionally, we will interview urological surgeons about their experience of the BAUS audits. At each of sites A-C 12 members of staff and two relevant commissioners will be interviewed, while at each of sites D-E seven members of staff and one relevant commissioner will be interviewed (total n=58).

Data collection: Interviews will be semi-structured and will follow guidance on the undertaking of realist 'theory gleaning' interviews.[76] Interviews will start with questions about the participant's role and their experience of, and involvement with, NCAs. To begin to elicit the participant's theories about how and in what contexts NCAs lead to improvements in care quality, we will ask about their perceptions of the impacts of the NCAs, before encouraging them to expand on the contexts in which they perceive those impacts to occur and their ideas about how those impacts are achieved. We will investigate what MINAP and PICANet data are collected (for example, MINAP contains over 130 fields, only 20 of which must be completed) and how, which of these measures are used in making assessments of care quality and informing QI initiatives, what other data participants would like to use for this purpose, and why. In interviews with members of quality sub-committees and Trust boards and commissioners, we will also explore their experience of NCAA, NACR, the Elective Surgery National PROMs Programme, and the BAUS audits. As recommended for co-design, interviews will be conducted at the participants' places of work, providing the opportunity to view and discuss workplace artefacts that participants perceive as relevant.[7] We will discuss the idea of QualDash, to gather participants' perceptions of what QualDash would need to provide, what they perceive the impact of that would be, and the contexts in which they perceive it would provide benefit. The research team will agree revisions to the interview topic guide in light of emerging theories. All interviews will be audio recorded and transcribed verbatim.

Analysis: An iterative approach to data collection and analysis will be taken, to support the refinement of emerging theories. Interview transcripts will be entered into a qualitative software program (NVivo 10) for indexing.

Framework analysis, an approach developed for analysing qualitative data for the purpose of applied research, will be used.[77] Framework analysis is an approach that has previously been used within realist evaluation studies,[78] including our own,[79] and we and have found it to be well suited to working with large data sets. Framework analysis allows for systematic and comprehensive analysis of the collected data and enables both between- and within-case analysis.

Framework analysis begins with *familiarisation*. This involves gaining an overview of the collected data through immersion in the data, while listing recurrent themes and issues which emerge as important to the participants. From this, a *thematic framework* will be identified for indexing the data, bringing together a priori issues (those introduced into the data via the interview topic guide) and emergent issues. Following the realist strategy, indexing of the data will focus on identifying participants' accounts of how outcome patterns are formed by mechanisms and contexts.[67] As multiple members of the research team will be undertaking the analysis, clear definitions of each code will be created and, after applying the codes to a sample of the data by different members of the team, we will measure the inter-rater reliability of the indexing[80] and refine the definitions. *Indexing* of all the collected data will then take place.

Once all the data has been indexed, *matrix displays*[81] will be created, to build up a picture of the data as a whole. This involves abstraction and synthesis of the data but referencing the original text.[77]

The final stage of analysis is that of *mapping and interpretation*. The matrices will be used to support both within-case comparisons (similarities and differences according to e.g. role) and between-case comparisons, returning to the original data where necessary. Similarities and differences in participants' theories will be identified. Because of the range of participants involved in the interviews, we anticipate that we might encounter conflicting theories. However, realist evaluation encourages the testing of multiple, contradictory theories, so the intention is not to remove or ignore such conflicting theories but to refine them.[69] Finally, we will compare the participants' theories with relevant substantive theories, such as those described above, using those theories that relate to the participants' theories to develop an integrated program theory. Given that QualDash use is likely to be supported or constrained by contextual features at the micro, meso, and macro level, we will adopt Westhorp's approach of 'layering' theories, in order to fully understand how contextual features shape the mechanisms through which QualDash will work.[82,83]

The resulting QualDash program theory will be summarised in both diagrammatic and narrative form.[64,65] It will be translated into a requirements specification, detailing both functional requirements (what QualDash should do) and non-functional requirements (including look and feel, usability, performance, and maintainability and support requirements, and requirements relating to implementation).[84] The requirements will be presented using the Volere format, where each requirement includes a description, a rationale, a source, and a fit criterion.[85] Each requirement will also state which categories of user it is relevant to (e.g. functionality that is of interest to one category of user but not others), e.g. clinical team, quality and safety sub-committee, Trust board, commissioners, or more fine-grained categories of user that may become apparent through the analysis. RAR will lead a review of the requirements with representatives from at least two of the Trusts' IT departments, to identify practical implementation issues and workarounds. In this way, we will use our team's multidisciplinary experience to act as a bridge between QualDash end-users (clinical teams, etc.) and IT specialists who play a critical part in systems implementation.

1b. Workshop with NCA suppliers

Sample: For the purpose of prioritising the functional and implementation-related requirements for QualDash, we will use a purposive sample of participants, involving suppliers representing a range of NCAs (covering both NCAPOP and independent NCAs and the three main funding mechanisms for independent NCAs, as well as NCAs where participation is at the independent clinician level) and representatives of HQIP. Currently, NCAs (both NCAPOP and independent) are delivered by a total of 37 suppliers, with some suppliers responsible for the delivery of up to five NCAs. Where a supplier delivers more than one NCA, we will approach the national clinical lead for each NCA. On the basis of experience of running events using similar techniques, we will seek to have a maximum of 32 participants. Groups of between five to eight participants are recommended,[86] so participants will be organised into four groups of eight participants.

Data collection: The workshop will be organised into two two-hour sessions.[87] In preparation for the workshop, participants will be sent the draft requirements specification in a user-friendly format.

The morning session will focus on prioritisation of functional requirements and will follow standard guidance for the running of nominal groups.[74,86,87] It will begin with an opening statement that describes the overall task, the contribution of participants, the procedures, and how the results will be used. The first activity will be the 'silent generation of ideas'; each participant will be given a set of cards describing all of the functional requirements and will select the functional requirements that they consider essential, as well as writing down on blank cards any additional functional requirements they consider to be essential. In the second activity, the group will merge their requirements, creating an 'agreed' and 'not agreed' list on a flipchart, the 'agreed' list comprising functional requirements considered essential by all group members and the 'not agreed' list having a 'weighting' for each item, i.e. how many people considered this requirement essential. Participants will then be asked to individually select the 10 functional requirements that they consider most important, ranking them in order of priority. The results will be collated and fed back to the group as a whole for discussion, before a final round of ranking the priorities.

The afternoon session will follow the same format as above, this time focusing on requirements relating to the implementation strategy.

To capture the richness of participants' observations, the discussion of each of the groups and the discussion of the group as a whole will be audio recorded and transcribed verbatim, providing valuable context to the ranked priorities.

Analysis: Quantitative analysis will involve combining the lists created by the four groups to identify which functional and implementation requirements were considered essential by all groups and for which requirements, while not considered essential by all participants, there was substantial support. The final ranking of the priorities will be combined to produce a list of functional requirements and a list of implementation requirements ordered by priority.

The transcripts of the discussions will be entered into NVivo 10 and analysed using framework analysis, to identify participants' reasoning about why a requirement was or was not necessary, with an emphasis on seeking configurations of contexts, mechanisms, and outcomes. This analysis will be used to make further refinements to the QualDash program theory.

Outputs

- A program theory that explains how and in what contexts QualDash use leads to improvements in care quality; and
- A requirements specification to support design of QualDash, generalisable to all NCAs.

Phase 2: Design of QualDash

Objective

- To use the program theory to co-design QualDash (Objective 2).

Summary of method

Phase 2 represents the conceptual design, prototyping and software development stages. QualDash will be developed iteratively, with input from a series of interactive focus groups. Focus groups are group interviews, coordinated by a moderator, that concentrate on a particular topic and generate qualitative data through discussion between

participants.[88] Differences in participants' opinions and accounts can be explored, helping to draw out the needs of different categories of user and to refine our understanding of the contextual factors that are likely to affect QualDash use.[89] Focus groups enable group ownership of problems and solutions and so are appropriate for the purpose of co-design.[88] It is important to ensure that the focus group is organised in such a way that participants are actively involved in the process of design, rather than just consulted,[7] so the focus groups will incorporate co-design activities such as cognitive walkthroughs and creating storyboards and paper-based prototypes. This will be complemented by ethnographic observation of meetings where quality and safety are discussed, to understand the functionality that QualDash needs to provide to support discussion of NCA data in those meetings and how QualDash may best be incorporated into those meetings.

Prototypes of QualDash will be discussed in the focus groups to explore comprehension, usability, and acceptability. A prototype is one manifestation of a design that allows stakeholders to interact with it and explore its suitability, and thus prototypes play an important role when co-designing interventions.[84] While the design of the prototypes will be based on the requirements specification, the focus groups will also provide the opportunity to explore the benefits of presenting more complete NCA data, such as longitudinal data (not just the current snapshot), all of the measures for a particular NCA rather than just a selection, the actual data values rather than just traffic lights, portraying confidence limits, showing context (a given Trust's performance against others), and integrating other sources of data, such as patient stories. Through this process, we will address the question 'what other data need to be presented alongside audit data to build a more complete picture of care quality?' (Question B) and begin to address the question 'what are the most effective means for clinical teams, quality sub-committees, boards, and commissioners to view and interact with audit data?' (Question C).

Once two high fidelity QualDash prototypes have been developed that focus group participants consider to be comprehensible, usable, and acceptable, a controlled user experiment will be undertaken, providing further data to answer Question C. The results of the controlled user experiment will be used to determine which QualDash prototype will be evaluated in Phase 4 and to determine if any further revisions to the QualDash prototype are required before the evaluation. Once any revisions have been made, the selected QualDash prototype will be subjected to extensive testing to ensure system stability.

Team

The focus groups will be run by the SoH RF and RA, with RR, RAR, and the SoC RF also present to hear directly the views of participants.[7] The SoH RF and RA will be responsible for analysing the focus group data, under the guidance of RR and JG. The SoC RF will be responsible for ongoing development of the QualDash software, under the guidance of RAR, and for the production of a technical specification that describes how the requirements have been translated into a functioning system. CG, MM, Robert West (RW), RP, and JL, through their familiarity with MINAP and PICANet and knowledge of how NCA data are used within the NHS, will contribute to the design of the QualDash prototypes.

The controlled user experiment will be undertaken by the SoH RF and RA, with technical support provided by the SoC RF. The SoC RF will be responsible for the analysis of the quantitative data generated by the controlled user experiment, while the SoH RF and RA will be responsible for the analysis of the qualitative data. Methodological guidance for the design and analysis of the controlled user experiment will be provided by RAR, RR, and DD. CG, MM, RP, and JL, through their knowledge of how NCA data are used within the NHS, will contribute to the design of the tasks to be undertaken within the controlled user experiment. The SoC RF will make any further revisions to QualDash in light of the findings, under the guidance of RAR, and will be responsible for testing of the final system.

2a. Iterative development

Sample: While all five Trusts will be involved in the co-design process, through participating in the Phase 1 interviews, co-designing the implementation strategy in Phase 3, and providing feedback on QualDash in Phase 4, we are intentionally only involving one Trust in Phase 2 so that we can assess, in Phase 4, to what extent the success of QualDash is dependent on this level of staff involvement in the design process. We anticipate that a minimum of three focus groups will be held (more if separate focus groups are held for MINAP and PICANet). An additional focus group will be held with our LAG, to gain a patient and carer perspective on the proposed designs. Guidance on focus groups suggests 8-12 participants as a suitable number.[88] While generally homogenous groups are recommended, it can be beneficial to bring together a diverse group to maximise exploration of different perspectives;[89] to support the identification of similarities and differences in the needs and preferences of our different stakeholders, the first focus group will bring together a purposive sample of members of the cardiology and PICU clinical teams, the quality sub-committee, and the Trust board, and local commissioners (total n=12). On the basis of this focus group we will decide whether in the subsequent focus groups there would be benefit in holding separate focus groups relating to MINAP and PICANet.

Data collection: The SoH RF will take on the role of moderator, assisted by the SoH RA who will make notes on the order of speakers and non-verbal interaction (e.g. body language suggesting lack of engagement).[88,90] Each focus group will be audio recorded.

Co-design requires careful preparation but also a level of flexibility, to allow participants involvement in the design of the approach as well as the intervention.[7] While open to revision and further specification, we plan to move from

developing paper-based storyboards and prototypes with participants (focus group 1), through to rapid prototyping where prototypes are purposely ambiguous and incomplete to encourage design input from participants (focus group 2), through to high fidelity prototypes (focus group 3). Discussion of the QualDash prototypes will include discussion of how comprehensible they are (in terms of both the functionality and the presentation of the data) and alternative ways in which the data could be presented.

Throughout the focus groups the moderator will encourage participants to discuss differences in their opinions.[89] At the end of each focus group, the moderator will present any tentatively identified issues to participants for confirmation or clarification.[90]

Analysis: The audio recordings will be transcribed verbatim, using the notes to identify speakers, and then annotated with notes of the non-verbal interaction. Analysis of the data will be conducted after each focus group using framework analysis as described above. This analysis will be used to make further refinements to the QualDash program theory and determine priorities for the next iteration of the QualDash prototypes.

Software development: The SoC RF will develop the QualDash software in two stages. The first (months 1-12) will turn the focus group outputs into high fidelity prototypes that implement the most important functionality identified in Phase 1, and then a version of QualDash that is suitable for evaluation by Trusts and CCGs in Phase 4. The key novelty of QualDash's design will come from the way we present data in visualisations, tailored to the needs of different users, and allow users to interact with those visualisations. The practicality of QualDash will be ensured through discussion with representatives from the Trusts' IT departments. The second stage of development (months 16-33) will incorporate feedback from the Phase 4 evaluation and functionality identified as important during the requirements prioritisation but initially omitted due to time constraints.

2b. Observation of meetings

Sampling: In each site, we will aim to observe at least one meeting at ward, quality sub-committee, Trust board, and CCG level.

Data collection: Researchers will observe the meetings, recording their observations in fieldnotes, to be written up in detail as soon after data collection as possible.

Analysis: An iterative approach to data collection and analysis will be taken, to feed in to the ongoing development of QualDash and to enable the gathering of further data in light of emerging themes. Fieldnotes will be entered into NVivo 10. Analysis of the data will be conducted using framework analysis as described above. This analysis will be used to make further refinements to the QualDash program theory and the requirements specification.

2c. Controlled user experiment

Sample: The key outcome measure we will use is the overall system usability scale (SUS) score (described below). Given a standard deviation for SUS of 10 units (scale is 0-100 with SUS<50 being poor and SUS>80 being very good) we want the signal to be clear when the difference recorded is 20 units distinguishing between poor and good usability. When there are at least 10 participants per group such a difference can be achieved with 98.8% power for a single comparison. Multiple comparisons however will be made, and the power drops to 97.2% for two comparisons, but remains more than 90% for up to seven comparisons. This allows a sufficient number of comparisons to be meaningfully made in our context. In order to assess the comprehensibility, usability, and acceptability of the QualDash prototypes to all the intended categories of user and to gain a patient and carer perspective, a purposive sample of members of the cardiology and PICU clinical teams, the quality sub-committee, and the Trust board, local commissioners, and members of our LAG will be invited to participate, with a minimum of 10 participants per group.

Data collection: A controlled user experiment will be run, using a mixed factorial experimental design, with technology (current presentation of NCA data and two QualDash prototypes) and task as within-participant variables and role, NCA (MINAP or PICANet), and graph literacy (high or low) as between-participant variables.

Prior to the experiment, participants will be asked to complete an online demographic questionnaire, which will include Galesic and Garcio-Retamero's scale for measuring graph literacy.[91] This scale, developed specifically for the health domain, measures both basic and advanced graph reading skills and comprehension across different types of graphs. The scale consists of 13 items and measures three levels of graph comprehension: i) the ability to read the data; ii) the ability to read between the data (to find relationships in the data); and iii) the ability to read beyond the data (e.g. to be able to predict a future trend from a line chart). Participants will be categorised as high or low literacy, using the median split of a high literacy score of nine or more correct responses, and then stratified prior to randomisation into an experimental group (high vs low literacy).

Each participant will complete a series of tasks in each condition. Details of the QualDash prototypes to be evaluated and tasks to be undertaken will be dependent on the outputs of Phase 1 and the Phase 2 focus groups. The order of technology used and the task order will be determined by a computer randomisation program. On completing each task, participants will be asked to record their decision, certainty and, using the NASA task load index,[92] their mental workload.

Time to complete each task will be recorded. Morae software will be used for recording participants' interaction with the QualDash prototypes, enabling capture of audio, video, on-screen activity, keyboard and mouse input.

To assess usability and acceptability of the QualDash prototypes, following completion of tasks on each prototype participants will be asked to complete the system usability scale (SUS), a flexible questionnaire designed to assess any technology.[93] The SUS is quick and easy to complete, consisting of 10 statements that are scored on a 5-point scale of strength of agreement, with final scores ranging from 0-100. On completion of tasks in all conditions, we will carry out semi-structured interviews with participants to explore their views of the two QualDash prototypes in more detail.

Analysis: Quantitative data (time to complete task, accuracy, certainty, perceived workload, and SUS score) will be analysed using analyses of variance (ANOVA). Qualitative data (from the Morae recordings and interviews) will be analysed using the methods previously described.

The analysis will be used to: make further refinements to the QualDash program theory, adding details regarding the impact of an individual's graph literacy as a contextual factor; determine which QualDash prototype should be used in the evaluation; determine if further revisions to the QualDash prototype are required before the evaluation and what they should be.

Outputs

- An interactive web-based quality dashboard for clinical teams, quality sub-committees, boards, and commissioners, using novel visualisation techniques and tailored to meet the needs of different users, that can be adapted for all NCAs, the design based on existing theory and significant engagement with suppliers and users of NCA data; and
- A technical specification for QualDash.

Phase 3: QualDash implementation

Objective

- To use the program theory to co-design an implementation strategy for QualDash (Objective 3).

Summary of method

Implementation is a key element in the MRC guidance on the development and evaluation of complex interventions.[4] Implementation refers to putting an intervention into practice, using strategies to support and encourage the use of that intervention in ways that will lead to the desired impact.[94] While an understanding of implementation requirements will have been developed in the Phase 1 interviews and prioritised with NCA suppliers in the Phase 1 workshop, it is important that implementation strategies are tailored to the local context.[66] For example, discussions with HQIP suggest that training will need to cover both how to use QualDash and how to use it as part of a QI process, but this will depend on the extent to which the intended users at different sites are already engaging with NCA data for the purpose of QI and the processes and practices that they have already developed. Implementation strategies are most likely to be successful if stakeholders are involved in their design.[95] Therefore, focus groups will be used, again ensuring that they are organised in such a way that participants are actively involved, rather than just consulted. At the end of this phase, planned implementation activities will be undertaken across the five Trusts and with relevant commissioners.

Team

Activities in this phase will be undertaken by the SoH RF and RA and the SoC RF, under the guidance of RR, in parallel to the Phase 2 activities. Resources to support implementation, such as user guides, will be reviewed by the LAG for comprehensibility.

Sample: One focus group will be held at each Trust, involving members of the relevant clinical teams, the quality sub-committee, the IT department, and the Trust board, and local commissioners.

Data collection: As in the Phase 2 focus groups, the focus groups will be moderated by the SoH RF, who will encourage participants to discuss differences in their opinions. The SoH RA will make notes on the order of speakers and non-verbal interaction. Each focus group will be audio recorded. Participants will be asked to discuss what they think the likely problems might be with introducing QualDash into their practice and ways in which these problems could be overcome, using the prioritised implementation requirements from Phase 1 as a starting point for discussion. At the end of each focus group, the moderator will present identified implementation activities to participants for confirmation or clarification.

Analysis: The audio recordings will be transcribed verbatim, using the notes to identify speakers, and then annotated with notes of the non-verbal interaction. Analysis of the data will be conducted after each focus group using framework analysis as described above. This analysis will be used to make further refinements to the QualDash program theory and inform the design of the implementation strategy.

Outputs

- An implementation strategy and associated resources for QualDash, tailored to each of the five Trusts.

Phase 4: QualDash evaluation

Objective

- To understand how and in what contexts QualDash leads to improvements in care quality (Objective 4).

Summary of method

Phase 4 addresses the question 'in what contexts and by what mechanisms do quality dashboards lead to increased engagement with audit data and subsequent improvements in care quality?' (Question D). To evaluate QualDash, we will collect data that will enable testing of the program theory. Realist evaluation does not employ particular methods of data collection, but is explicitly a mixed method approach.[96] Outcome data, in the form of key MINAP and PICANet process measures, will be collected and analysed in a CITS study, while a multi-site case study[97] will provide insight into the contexts and mechanisms that lead to those outcomes. CITS studies provide a robust method of assessing the effect of an intervention and have been used to assess the effectiveness of a variety of complex interventions.[98]

In the multi-site case study, data will be collected through ethnographic observation and interviews. Ethnography is the study of people in their environments where the researcher participates in the setting in order to collect data.[99] Ethnographic methods, such as non-participant observation, have been used in previous realist evaluations as part of the process of theory testing and refinement,[69,100] and such methods have been argued as essential for studying both the implementation of QI interventions[29] and the introduction of technology in healthcare.[101] We will follow the Biography of Artefacts (BoA) approach,[59] which has been used successfully in a number of studies of HIT[58,102] and which we have found useful in our current study of dashboards. This approach involves 'strategic ethnography',[59] studying longitudinally, across different locations, and at different levels of the organisation how practices surrounding a new technology evolve and how previous practices enable and constrain this process.[58]

Interviews are an important complement to ethnographic observation, providing an opportunity for the researcher to ask questions about aspects of practice that might not be immediately intelligible to an observer, as well as for gaining interviewees' reasoning about QualDash. Logfiles will provide additional data on when and how QualDash is used.

A&F interventions, and QI interventions more generally, require longitudinal evaluation to allow sufficient time for staff to implement changes and incorporate them into their practice.[103-105] Similarly, the evaluation of HIT should allow time for staff to integrate the technology into their practices and evolve those practices to take advantage of the functionality offered by the technology.[58] Therefore, data will be collected over a 12 month period.

Team

The CITS will be designed and undertaken by RW. Data collection and analysis for the multi-site case study will be undertaken by the SoH RF and RA, under the leadership of RR, with JG, DD, and JK providing additional methodological guidance. CG, MM, RP, and JL, through their knowledge of how NCA data are used within the NHS, will contribute to the design of the multi-site case study and the interpretation of the findings of both the CITS and the multi-site case study. Amanda Farrin (AF) will advise on the data to be collected for informing the CRT feasibility assessment and design. The SoC RF will analyse the logfile data to identify patterns of QualDash use (at what sites, by whom, when, and using what functionality). Participants' feedback on QualDash will be used to make revisions to the software, which will be undertaken by the SoC RF, under the guidance of RAR.

4a. Controlled interrupted time series study

Sample: Data will be collected for sites A-E, with two control Trusts per intervention Trust, matched on comparable outcomes pre-intervention. Contextual factors in the QualDash program theory may be used as additional matching criteria. A CITS study requires data for a minimum of three time points pre-intervention and three time points post-intervention and must also allow for any seasonal effect on the outcomes.[106] Therefore, monthly data will be obtained for 24 months pre-intervention and 12 months post-intervention. For MINAP, Cumulative Missed Opportunities for Care (CMOC) will be used as the primary process measure (described below). The average CMOC for patients on a ward will be averaged for each month, so that there are 36 observations clustered within each of 15 hospitals. Taking the intra-class correlation to be 0.15, this yields a design effect of 6.25. Hence the effective number of observations is 15*36/6.25=86.4. Using Cohen's approach to sample size calculation this means that an effect size of 0.17 can be estimated with 80% given that there are six parameters in the model (including the coefficient for QualDash). Converting this to the percentage of variation that can be explained by the model, this yields 20.5%. Translating this back to CMOC, currently 49.6% are discharged from hospital without missing any of the nine opportunities for care, and we would be powered at the 80% level to detect an improvement from an average of 8.33 opportunities achieved to 8.46. Thus our study has good power to detect small but meaningful clinical improvements. For PICANet, the primary process measure will be the use of non-invasive ventilation first for patients requiring ventilation, 10% of the admitted population.[107] On average there are 5.25 ventilation cases per month per hospital. With a further design effect from patients clustered within hospitals, based on the reported ICC=0.065 giving DEff=1.276, the actual anticipated number of patients is 1701 giving an effective number of 213: 71 exposed to QualDash and 142 controls. This yields 80% power to detect a change from 32% to 53% which is a feasible target given that the use of non-invasive ventilation first will be highlighted by QualDash.

Data collection: For MINAP, the primary outcome will be the composite process measure CMOC. This has nine components (pre-hospital ECG, acute use of aspirin, timely perfusion, prescription at hospital discharge of aspirin, thienopyridine inhibitor, ACE-inhibitor, HMG-CoA reductase inhibitor, beta blocker, referral for cardiac rehabilitation) and, in previous work by members of the team, CMOC has been shown to be associated with mortality.[108] For PICANet, the primary outcome will be use of non-invasive ventilation first for patients requiring ventilation, which has been shown to be associated with reduced mortality.[107] Secondary outcomes will be determined on the basis of findings from Phase 1 regarding interviewees' perceptions of the value of particular measures and the extent to which they can be impacted by QI initiatives and through consultation with our LAG.

Analysis: Monthly MINAP and PICANet data will be extracted to spreadsheets for analysis with R software.[109] For each intervention Trust, there will be 72 data points prior to introduction (24 for the intervention Trust and 48 for the control Trusts) and 36 data points post intervention (12 for the intervention Trust and 24 for the control Trusts). For both NCAs, the outcome will be regressed upon time and the intervention (QualDash). The time component will include a seasonal effect (quarterly effect) and will allow for a (linear) time trend. To account for clustering of monthly observations within hospitals a random intercept will be fitted, although a fixed effect for hospital as a sensitivity analysis will be explored. Although the intervention is abrupt, its impact may well be 'phased in' over a few months, perhaps three. The timing of the bedding in of the intervention will be reported from the multi-site case study. Then a partial effect can be considered for this period with the interaction effect stepping up in a linear fashion.

4b. Multi-site case study

Sampling: In sites A-C we will undertake a minimum of 24 four hour periods of observation per Trust and in sites D-E we will undertake a minimum of 12 four hour periods of observation per Trust (total n=384 hours). While the researchers will return to each Trust monthly, to understand how use of QualDash changes over time, more time will be spent in the first few months following the introduction of QualDash, because this is when users are most likely to engage with and explore the affordances of QualDash and establish new practices around it, generating information with implications for both system enhancement and future implementation strategies.[60]

Data collection: At each case site, there will be an initial phase of general ethnographic observation, informed by the implementation chains identified in the Phase 1 interviews. Following the methods we have used in our current study of dashboards, this is anticipated to include: i) observation of formal meetings at ward, quality sub-committee, Trust board, and CCG level where quality and safety are discussed; and ii) observation in clinical areas to understand the working practices of the clinical teams and capture what Waring refers to as 'corridor committees' where issues of quality and safety are discussed more informally.[28] Following this, observation will be more focused, with details of what to pay attention to depending on the refined QualDash program theory. How and in what contexts QualDash is used (or not) at different levels will be observed, understood in the context of broader practices and use of other sources of information for monitoring care quality. Researchers will record observations in fieldnotes, to be written up in detail as soon after data collection as possible.

Observations will be complemented by semi-structured interviews with staff. These will gather data on: i) staff perceptions of mechanisms that result from QualDash use; and ii) outcomes that cannot be easily gathered by other means, particularly those relating to the perceptions of staff. Interviews will be audio recorded and transcribed verbatim.

Patterns of QualDash use identified through the analysis of the logfiles will be used to inform qualitative data collection (e.g. asking in interviews why participants use particular functionality), while the logfiles will provide detailed data on QualDash use observed during the course of data collection.

At the end of the data collection period, we will ask participants to complete a questionnaire based on the Technology Acceptance Model, using well validated items that have been used in numerous previous studies of health IT[49] and in previous studies of dashboards.[110] This will provide participants' perceptions of the usefulness of QualDash and data on whether they intend to continue using QualDash after the study period.

Depending on the refined QualDash program theory, it might be appropriate to add additional data collection methods. For example, if the safety culture of a ward or Trust is seen as an important contextual factor, a questionnaire to measure safety culture may be included.[111]

Analysis: An iterative approach to data collection and analysis will be taken, to enable: ongoing testing and refinement of the program theory; gathering of further data in light of such revisions; and refinement of QualDash in response to participants' feedback. Fieldnotes and interview transcripts will be entered into NVivo 10. Narrative analysis will be undertaken to develop a 'biography' of QualDash, which will describe the use of QualDash by a range of stakeholders at different levels and the interconnections between them (clinical team, quality sub-committee, Trust board, and CCG). Narrative analysis is consistent with a realist approach due to its emphasis on preserving connections within the data, thereby helping to understand causality.[112] Incorporating the results of the CITS analysis and analysis of the logfiles and questionnaire data, the biography will focus on how contextual factors shape the evolution of practices around QualDash and how this leads to the resulting outcome pattern, thereby providing evidence to support, refute, or revise the QualDash program theory.

Outputs

- Evidence regarding how and in what contexts use of QualDash leads to improvements in care quality.

Phase 5: Trial feasibility assessment and design

Objective

- To assess the feasibility of conducting a CRT of QualDash (Objective 5); and
- To assess the extent to which QualDash and the program theory are applicable to other NCAs.

Summary of method

The commissioning brief emphasises the need for trials within NCAs to test the principles of effective communication of NCA data to key audiences. We will use the findings of Phase 4 to determine for whom and in what contexts a trial of QualDash is appropriate. This avoids the cost of a trial if QualDash is found to be ineffective while, if the results of Phase 4 suggest QualDash does provide benefit, we will be able to design a trial that is fit for purpose, with high internal and external validity, for which subsequent funding would be sought.[113,114]

Progression criteria will be: (i) QualDash is used by 50% or more of the intended users; (ii) data completeness in the NCA improves or remains the same; and (iii) participants perceive QualDash to be useful and express the intention to continue using it after the study period. Criteria (i) and (iii) are concerned with acceptability and uptake of the intervention, and therefore have implications for recruitment to a trial. The second criterion is concerned with ensuring that the intervention does not have unintended negative consequences which would affect both the success of the intervention (as QualDash will be less useful if data completeness is reduced) and the feasibility of outcome assessment. The third criterion is also concerned with participants' perceptions of the impact of QualDash on care. While not formally assessed as part of the progression criteria, the impact of QualDash on processes of care as identified in the CITS will also be considered in determining whether a future CRT is justified. A traffic light system will be used to determine if a trial is feasible (green), feasible with modifications to QualDash and/or the implementation strategy (amber), or not feasible (red).[113,115]

If a trial is found to be feasible, a CRT will be designed. It is anticipated that the primary outcome will be a patient outcome. Data from the CITS will be used to inform the selection of NCAs to be included in the trial (MINAP and/or PICANet) and will provide information about variability of outcomes and about how long a trial intervention period would need to be. Findings from the multi-site case study will be used to inform the selection of categories of user to be included in the trial and, associated with this, the level of randomisation (Trust, hospital, or ward).

If the feasibility assessment indicates that QualDash and/or the implementation strategy need to be modified, Phase 4 findings will be used to determine what those modifications should be, while also identifying components of QualDash and the implementation strategy that, being associated with mechanisms that produce desired outcomes, should be preserved in the trial.

In addition, two focus groups will be undertaken in the Phase to assess the extent to which QualDash and the program theory are applicable to other NCAs.

Team

Trial feasibility assessment and design will be led by AF, working in collaboration with RR and the SoH RF, with support from a statistician and trial manager from the Leeds Institute for Clinical Trials Research (LICTR). If modifications to QualDash are required, these will be undertaken by the SoC RF, under the guidance of RAR. If modifications to the implementation strategy are required, these will be undertaken by the SoH RF, under the guidance of RR. RR and the SoH RF will undertake the focus groups.

Sample: Two focus groups will be held. In the first focus group, we will bring together quality sub-committee members from across the five sites. In the second focus group, we will bring together suppliers of intermittent audits.

Data collection: The focus groups will be moderated by RR, who will encourage participants to discuss differences in their opinions. The SoH RF will make notes on the order of speakers and non-verbal interaction. Each focus group will be audio recorded. In the first focus group, participants will be asked to discuss the extent to which QualDash would be suitable for NCAA, NACR, the Elective Surgery National PROMs Programme, and the BAUS audits. The discussion will not just consider the features of QualDash and how they may need to be adapted for the particular types of data held within these three audits, but also contextual factors that may impact use of QualDash for engaging with these audits. In the second focus group, participants will be asked to discuss whether and how QualDash might be adapted to provide value to intermittent audits.

Analysis: The audio recordings will be transcribed verbatim, using the notes to identify speakers, and then annotated with notes of the non-verbal interaction. Analysis of the data will be conducted using framework analysis as described above. This analysis will be used to make further refinements to the QualDash program theory.

Outputs

- Assessment of the feasibility of a CRT of QualDash; and
- If criteria for progression are met, design for a CRT of QualDash.

DISSEMINATION AND PROJECTED OUTPUTS

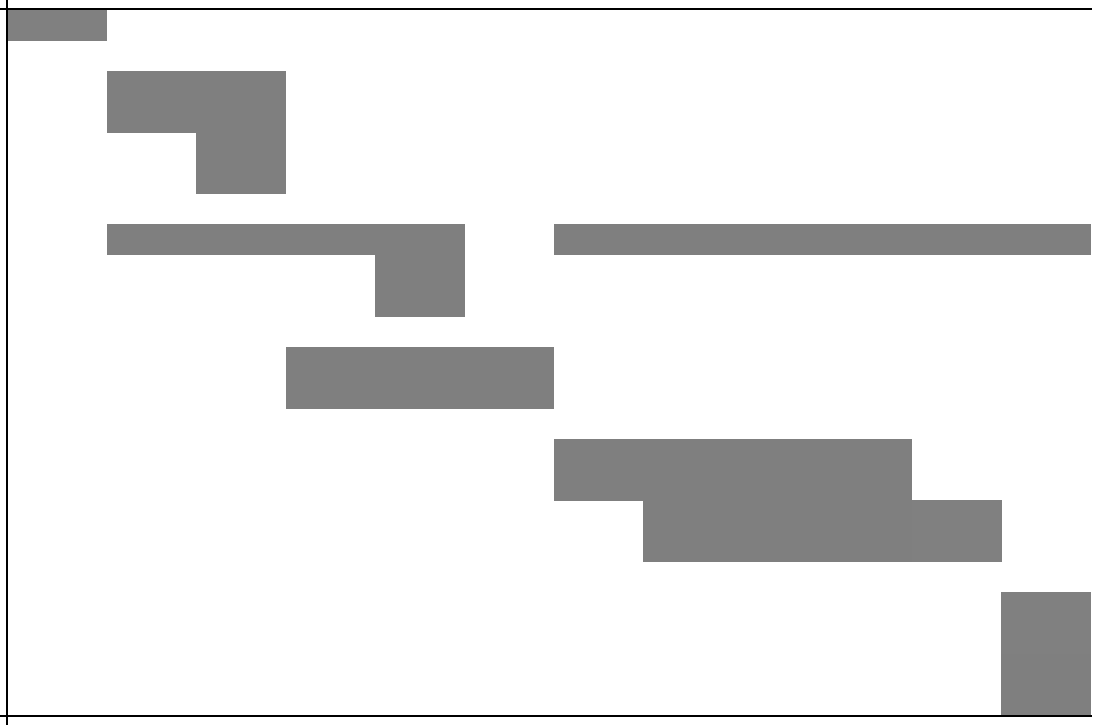
The main outputs will be: i) a program theory that explains how and in what contexts QualDash use leads to improvements in care quality; ii) based on the QualDash program theory, a requirements specification to support design of QualDash, generalisable to all NCAs, delivered by end of month 6 with revisions as the project progresses; iii) a technical specification for QualDash, delivered by end of month 12 with revisions as the project progresses; iv) an interactive web-based quality dashboard for clinical teams, quality sub-committees, boards, and commissioners, using novel visualisation techniques and tailored to meet the needs of different users, that can be adapted for all NCAs; v) a QualDash implementation strategy, the success of which is not dependent on involvement in design of QualDash; vi) evidence regarding how and in what contexts use of QualDash leads to improvements in care quality; vii) assessment of the feasibility of a CRT of QualDash; and, if progression criteria are met, viii) design for a CRT of QualDash.

Outputs i-iii will be sent to HQIP and NCA suppliers and will be made available on a dedicated project website. Outputs i and ii will enable suppliers of other NCAs to draw on our findings to inform the design of feedback they provide, and output ii can be used by suppliers if employing a commercial company to develop a quality dashboard. Output iii will support suppliers of other NCAs to develop their own quality dashboards based on the design of QualDash. We will hold an end of project dissemination event for NCA suppliers, with presentations video recorded and made available on the project website.

Other dissemination activities will include Open Access publications in a range of academic journals (e.g. BMJ Quality & Safety, Implementation Science, Journal of the American Medical Informatics Association), and presentation of findings at a national conference, such as the HSRUK Symposium or the HQIP-sponsored Clinical Audit for Improvement conference.

PLAN OF INVESTIGATION AND TIMETABLE

The project will be for 33 months. The timetable below describes the key tasks to be completed within the project and the timing of key milestones.



	Months											
	Pre-study	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24	25-27	28-30	31-33
Tasks												
Ethics & R&D approval												
PHASE 1												
Interviews across 5 Trusts												
Workshop with NCA suppliers												
PHASE 2												
Iterative development												
Controlled user experiment												
PHASE 3												
QualDash implementation												
PHASE 4												
QualDash evaluation – data collection												
QualDash evaluation – analysis												
PHASE 5												
Trial feasibility assessment & design												
Focus groups												
Final report												
Milestones												
Requirements specification delivered			*									
Technical specification delivered					*							
Implementation strategy delivered						*						
QualDash software release						*						*
QualDash evaluation data collection complete										*		
QualDash evaluation analysis complete											*	
Progress reports			*		*		*		*		*	

PROJECT MANAGEMENT

All applicants will be part of the Project Management Group (PMG), chaired by RR (PI). This group will initially meet monthly and then every two months, where the progress of the research will be reported and decisions regarding next steps will be made. Between meetings, email will be used to keep the PMG up to date. A meeting with RR, RAR, JK, and the research staff will be held fortnightly to monitor the progress of the research. In addition, RR will meet at least weekly with the SoH research staff and RAR will meet at least weekly with the SoC RF. Other members of the PMG will be involved in these meetings as and when their methodological and/or subject expertise is needed.

We will also convene a Study Steering Committee (SSC). The following people have agreed to be on the SSC: Professor Graham Martin (University of Leicester, expertise in relevant social science theories and the conduct of ethnographic studies of quality and safety); Dr Ana Manzano (University of Leeds, expertise in realist evaluation); Dr Clive Weston (Swansea University, MINAP clinical lead). We will have a representative of HQIP on the SSC (Yvonne Silove, Tasneem Hoosain, or Kirsten Windfuhr). We will also invite representatives of NCAA, NACR, and the Elective Surgery National PROMs Programme to join the SSC. We will appoint an NHS manager with responsibility for audit via the Yorkshire & Humber Effectiveness & Audit Regional Network, an NHS IT manager, and an expert in CITS. We will ask our LAG to nominate a member to sit on the SSC. SSC meetings will take place at key points in the project: shortly

after the beginning of the project; at the end of Phase 1, to discuss the requirements specification and its implications for the design of QualDash; at the end of Phase 3, to review the Phase 4 evaluation design in light of the revised QualDash program theory; and at the end of Phase 4, to discuss the findings and appropriate steps forward. The dissemination plan will be discussed at the first SSC meeting and will be a regular item on the agenda of the SSC meetings.

APPROVAL BY ETHICS COMMITTEES

We propose to carry out data collection across five NHS hospital Trusts, involving staff in the study. We will also be working with anonymised routinely collected data. The research team has considerable relevant experience, both in carrying out qualitative research in such settings and in the use of routinely collected data. NHS Research Ethics Committee approval will be sought. We have a draft ethics application which will be submitted by the end of May 2017. R&D governance approval will then be sought. Both MINAP and PICANet data are already held securely within the University of Leeds. We are aware of the third Caldicott report and will ensure that our data handling arrangements are compliant with any new guidance published in 2016/2017.

PATIENT AND PUBLIC INVOLVEMENT

A draft of the plain English summary was circulated to the Patient and Public Involvement (PPI) Group of our current dashboards study and discussed at the PPI meeting in April 2016. A LAG will be established, with participants recruited from the PPI Group of our current study, lay members of PICANet's Paediatric Intensive Care Families Group, and lay members of the MINAP Steering Group. LAG members will: Advise on data collection, e.g. review interview topic guides; Participate in a focus group to inform the design of QualDash; Participate in the controlled user experiment; Contribute to the design of the QualDash evaluation, e.g. selection of measures for the CITS; Provide their perspective on findings; Advise on dissemination of findings to relevant interest groups. They will also contribute to the creation of outputs, e.g. reviewing the requirements specification and user guide for comprehensibility. Five LAG meetings will be held over the course of the project, with the LAG also meeting to participate in the co-design activities and the controlled user experiment and with additional communication with members via telephone and email as the need arises. Four of the meetings will be scheduled to occur in advance of the SSC meetings, so that feedback from the LAG can be fed into the SSC meetings. An additional meeting will be held during Phase 1, prior to the workshop with NCA suppliers, for reviewing the requirements specification. A training session will be organised and de-briefing sessions will be held after activities to gather feedback on the process. This approach builds on successful PPI methods used in previous HS&DR projects. PPI will benefit the process and outputs of the project by: ensuring the topics explored in interviews include those that matter to patients/carers; making the design of QualDash more patient friendly; ensuring the measures used in the CITS are ones that matter to patients/carers; providing a more nuanced analysis (e.g. challenging clinicians' assumptions); ensuring comprehensibility of project materials; making communication of findings to the general public more effective.[116]

EXPERTISE AND JUSTIFICATION OF SUPPORT REQUIRED

Expertise

The project team is uniquely placed to carry out this work, with team members bringing significant relevant topic knowledge from current and previous projects in this area. The team will also have access to the knowledge and resources of the Leeds Institute for Data Analytics, which includes the £7m Leeds MRC Medical Bioinformatics Centre (ES/L011891/1), and the Applied Health Co-operative.

HIT design and evaluation: RR (study PI) has extensive experience of studying HIT use and contributing to the design of HIT. RAR is Professor of Computing and has expertise in the visualisation and analysis of large data sets. He is PI for QuantiCode (EP/N013980/1), a 3 year project funded by the EPSRC Making Sense of Data call, which is developing novel visualisation tools for analysis of electronic health records. DD brings unique expertise in development and evaluation of decision support tools. She is currently leading an NIH study on design and evaluation of dynamic dashboards for community nurses (AHRQ R21 HS 023855-01A1). JK is Professor of Health Politics and brings an organisational perspective on HIT use and knowledge of use of dashboards and related technologies within the NHS. He is PI for an HS&DR study (13/07/68) on the design and use of quality dashboards within NHS acute Trusts. We have expertise in human-computer interaction (RR, RAR) and experience of introducing new technologies into the NHS (RR, RAR).

Realist evaluation: JG, part of the RAMESES II team funded by HS&DR, provides expertise in realist evaluation. She led an HS&DR funded realist synthesis of the use of aggregate and individual PROMs data (12/136/31), which shares a number of program theories with the use of NCA data e.g. theories of A&F and benchmarking. RR was PI on a recently completed HS&DR funded realist evaluation of robotic surgery (12/5005/04), with JG and DD as co-investigators.

Qualitative methods: Team members also have methodological expertise in multi-site case studies (RR, DD, JK, JG) and qualitative research (RR, JG, DD, JK).

CITS and statistics: RW is Professor of Biostatistics and brings expertise in the design and conduct of CITS analyses. CG, MM, and RW all have experience of working with MINAP data. RP is the PICANet PI, bringing expertise in working with PICANet data.

Trial feasibility assessment and design: AF is Professor of Clinical Trials & Evaluation of Complex Interventions and brings expertise in trial feasibility assessment and design. She is co-investigator on the NIHR programme grant AFFINITIE which involves 2 CRTs of A&F interventions.
Clinical and professional expertise: CG and MM are both consultant cardiologists, bringing clinical expertise. An NHS perspective is provided by JL, Clinical Information & Outcomes Manager at Leeds Teaching Hospitals NHS Trust. CG, MM, RP, and JL provide knowledge of how NCA data are being used within the NHS and links to clinical teams who work with these data.

Justification of support required
Because A&F interventions require longitudinal evaluation to allow sufficient time for staff to implement changes and incorporate them into their practice, we are seeking funding for 33 months to allow us to follow use of QualDash over a 12 month period. Successful completion of this substantial and complex project requires significant input from a productive multidisciplinary team. Therefore, the majority of costs relate to staff costs and associated estates and indirect costs. A RF (1.0 FTE for 33 months) and RA (0.5 FTE for 30 months) will be employed in SoH and a RF (1.0 FTE for 33 months) will be employed in SoC. The PI has been costed at 0.3 FTE and will be responsible for project management and managing the SoH RF and RA. RAR has been costed at 0.15 FTE as he will be responsible for overseeing the work of the SoC RF and will have significant input into the design of QualDash. RW has been costed at 0.05 FTE in Years 1 and 2 then 0.1 FTE in Year 3 to undertake the CITS analysis. All other co-applicants have been costed at 0.05 FTE to support their contribution to the design, analysis and dissemination of the research. A statistician and trial manager from the LICTR have been costed to advise on Phase 5 and design of a definitive CRT (6 days each over 3 months).

REFERENCES

1. Tukey, J.W., *Exploratory Data Analysis.* 1977, Reading, Mass.: Addison-Wesley.
2. Vessey, I., *Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature.* Decision Sciences, 1991. **22**(2): p. 219-240.
3. Hysong, S.J., R.G. Best, and J.A. Pugh, *Audit and feedback and clinical practice guideline adherence: Making feedback actionable.* Implementation Science, 2006. **1**(1): p. 1-10.
4. Medical Research Council, *Developing and evaluating complex interventions: new guidance*, 2008.
5. Pawson, R. and N. Tilley, *Realistic Evaluation.* 1997, London: SAGE Publications.
6. Bate, P. and G. Robert, *Experience-based design: from redesigning the system around the patient to co-designing services with the patient.* Quality and Safety in Health Care, 2006. **15**(5): p. 307-310.
7. Bate, P. and G. Robert, *Bringing user experience to healthcare improvement: The concepts, methods and practices of experience-based design.* 2007: Radcliffe Publishing.
8. Phekoo, K.J., J. Clements, and D. Bell, *National Clinical Audit Quality Assessment - Overview of the self-assessment survey: "audit of audits"*, 2014, Healthcare Quality Improvement Partnership: London.
9. Birkhead, J.S., L. Walker, M. Pearson, C. Weston, A.D. Cunningham, and A.F. Rickards, *Improving care for patients with acute coronary syndromes: initial results from the National Audit of Myocardial Infarction Project (MINAP).* Heart, 2004. **90**(9): p. 1004-1009.
10. Bridgewater, B., A.D. Grayson, N. Brooks, G. Grotte, B.M. Fabri, J. Au, T. Hooper, M. Jones, and B. Keogh, *Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25 730 patients undergoing CABG surgery under 30 surgeons over eight years.* Heart, 2007. **93**(6): p. 744-748.
11. Neuburger, J., C. Currie, R. Wakeman, C. Tsang, F. Plant, B. De Stavola, D.A. Cromwell, and J. van der Meulen, *The Impact of a National Clinician-led Audit Initiative on Care and Mortality after Hip Fracture in England: An External Evaluation using Time Trends in Non-audit Data.* Medical Care, 2015. **53**(8): p. 686-691.
12. Taylor, A., J. Neuburger, K. Walker, D. Cromwell, and O. Groene, *How is feedback from national clinical audits used? Views from English National Health Service trust audit leads.* Journal of Health Services Research & Policy, 2016. **21**(2): p. 91-100.
13. Allwood, D., *Engaging Clinicians in Quality Improvement through National Clinical Audit*, 2014, Healthcare Quality Improvement Partnership: London.
14. Keogh, B., *Review into the quality of care and treatment provided by 14 hospital trusts in England*, 2013, NHS: London.
15. National Advisory Group on the Safety of Patients in England, *A Promise to Learn - A Commitment to Act: Improving the Safety of Patients in England*, 2013, Department of Health: London.
16. Jamtvedt, G., J. Young, D. Kristoffersen, M. O'Brien, and A. Oxman, *Audit and feedback: effects on professional practice and health care outcomes (Review).* The Cochrane Library, 2007. **2**.
17. Ivers, N., G. Jamtvedt, S. Flottorp, J.M. Young, J. Odgaard-Jensen, S.D. French, M.A. O'Brien, M. Johansen, J. Grimshaw, and A.D. Oxman, *Audit and feedback: effects on professional practice and healthcare outcomes.* Cochrane Database of Systematic Reviews, 2012(6).
18. Ivers, N.M., J.M. Grimshaw, G. Jamtvedt, S. Flottorp, M.A. O'Brien, S.D. French, J. Young, and J. Odgaard-Jensen, *Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis*

of Audit and Feedback Interventions in Health Care. Journal of General Internal Medicine, 2014. **29**(11): p. 1534-1541.
19. Ivers, N.M., A. Sales, H. Colquhoun, S. Michie, R. Foy, J.J. Francis, and J.M. Grimshaw, *No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention.* Implementation Science, 2014. **9**(1): p. 1-8.
20. Bickman, L., M. Riemer, C. Breda, and S. Kelley, *CFIT: A system to provide a continuous quality improvement infrastructure through organizational responsiveness, measurement, training, and feedback.* Report on Emotional and Behavioral Disorders in Youth, 2006. **6**(4): p. 86-87.
21. Hysong, S.J., M.K. Knox, and P. Haidet, *Examining Clinical Performance Feedback in Patient-Aligned Care Teams.* Journal of General Internal Medicine, 2014. **29**(2): p. 667-674.
22. Eldh, A.C., M. Fredriksson, S. Vengberg, C. Halford, L. Wallin, T. Dahlström, and U. Winblad, *Depicting the interplay between organisational tiers in the use of a national quality registry to develop quality of care in Sweden.* BMC Health Services Research, 2015. **15**(1): p. 1-9.
23. Meade, P., *A guide to benchmarking.* 2007: University of Otago.
24. Van Helden, G.J. and S. Tillema, *In Search of a Benchmarking Theory for the Public Sector.* Financial Accountability & Management, 2005. **21**(3): p. 337-361.
25. Waring, J., D. Allen, J. Braithwaite, and J. Sandall, *Healthcare quality and safety: a review of policy, practice and research.* Sociology of Health & Illness, 2016. **38**(2): p. 198-215.
26. Martin, G.P., N. Armstrong, E.-L. Aveling, G. Herbert, and M. Dixon-Woods, *Professionalism Redundant, Reshaped, or Reinvigorated? Realizing the "Third Logic" in Contemporary Health Care.* Journal of Health and Social Behavior, 2015. **56**(3): p. 378-397.
27. Martin, G.P., M. Leslie, J. Minion, J. Willars, and M. Dixon-Woods, *Between surveillance and subjectification: Professionals and the governance of quality and patient safety in English hospitals.* Social Science & Medicine, 2013. **99**(0): p. 80-88.
28. Waring, J., *Adaptive regulation or governmentality: patient safety and the changing regulation of medicine.* Sociology of Health & Illness, 2007. **29**(2): p. 163-179.
29. Dixon-Woods, M., C.L. Bosk, E.L. Aveling, C.A. Goeschel, and P.J. Pronovost, *Explaining Michigan: Developing an Ex Post Theory of a Quality Improvement Program.* Milbank Quarterly, 2011. **89**(2): p. 167-205.
30. Sørensen, E. and J. Torfing, *Theories of democratic network governance.* 2016: Springer.
31. West, E., *Organisational sources of safety and danger: sociological contributions to the study of adverse events.* Quality in Health Care, 2000. **9**(2): p. 120-126.
32. Pauwels, K., T. Ambler, B.H. Clark, P. LaPointe, D. Reibstein, B. Skiera, B. Wierenga, and T. Wiesel, *Dashboards as a Service Why, What, How, and What Research Is Needed?* Journal of Service Research, 2009. **12**(2): p. 175-189.
33. Daley, K., J. Richardson, I. James, A. Chambers, and D. Corbett, *Clinical dashboard: use in older adult mental health wards.* The Psychiatrist, 2013. **37**: p. 85-88.
34. Dowding, D., R. Randell, P. Gardner, G. Fitzpatrick, P. Dykes, J. Favela, S. Hamer, Z. Whitewood-Moores, N. Hardiker, E. Borycki, and L. Currie, *Dashboards for improving patient care: Review of the literature.* International Journal of Medical Informatics, 2015. **84**(2): p. 87-100.
35. Van Der Meulen, M., R.H. Logie, Y. Freer, C. Sykes, N. McIntosh, and J. Hunter, *When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care.* Applied Cognitive Psychology, 2010. **24**(1): p. 77-89.
36. Hutchinson, J., J.W. Alba, and E.M. Eisenstein, *Heuristics and biases in data-based decision making: Effects of experience, training, and graphical data displays.* Journal of Marketing Research, 2010. **47**(4): p. 627-642.
37. Zaydfudim, V., L.A. Dossett, J.M. Starmer, and et al., *Implementation of a real-time compliance dashboard to help reduce sicu ventilator-associated pneumonia with the ventilator bundle.* Archives of Surgery, 2009. **144**(7): p. 656-662.
38. Pablate, J., *The effect of electronic feedback on anesthesia providers' timely preoperative antibiotic adminstration*, 2009, University of North Florida. p. 90 p.
39. Batley, N.J., H.O. Osman, A.A. Kazzi, and K.M. Musallam, *Implementation of an emergency department computer system: design features that users value.* Journal of Emergency Medicine, 2011. **41**(6): p. 693-700.
40. Linder, J.A., J.L. Schnipper, R. Tsurikova, D.T. Yu, L.A. Volk, A.J. Melnikas, M.B. Palchuk, M. Olsha-Yehiav, and B. Middleton, *Electronic health record feedback to improve antibiotic prescribing for acute respiratory infections.* American Journal of Managed Care, 2010. **16**(12 Suppl HIT): p. e311-9.
41. Department of Health, *High quality care for all: NHS next stage review final report*, 2008, DH: London.
42. Department of Health, *Health Informatics Review report*, 2008, DH: London.
43. NHS Connecting for Health, *Clinicians driving Clinical Dashboards forward: A case study with Dr Anne Talbot*, NHS CFH: Leeds.
44. NHS Connecting for Health, *Clinical Dashboards case study: NHS Bolton staff dash to clinical improvement*, NHS CFH: Leeds.
45. NHS Connecting for Health, *Implementing Clinical Dashboards: Older Person Mental Health Services*, NHS CFH: Leeds.

46. The Mid Staffordshire NHS Foundation Trust Public Inquiry, *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*, 2013, The Stationery Office: London.
47. Randell, R., J. Keen, C. Gates, E. Ferguson, A. Long, C. Ginn, E. McGinnis, S. Willis, and J. Whittle, *Managing Quality and Safety In Real Time? Evidence From an Interview Study.* Studies in Health Technology & Informatics, 2016. **228**: p. 23-7.
48. Allen, D., *The importance, challenges and prospects of taking work practices into account for healthcare quality improvement: Nursing work and patient status at a glance white boards.* Journal of Health Organization and Management, 2016. **30**(4): p. 672-689.
49. Holden, R.J. and B.-T. Karsh, *The Technology Acceptance Model: Its past and its future in health care.* Journal of Biomedical Informatics, 2010. **43**(1): p. 159-172.
50. Gaissmaier, W., O. Wegwarth, D. Skopec, A.-S. Müller, S. Broschinski, and M.C. Politi, *Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information.* Health Psychology, 2012. **31**(3): p. 286-296.
51. Okan, Y., R. Garcia- Retamero, E.T. Cokely, and A. Maldonado, *Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension.* Journal of Behavioral Decision Making, 2012. **25**(4): p. 390-401.
52. Finch, T., F. Mair, C. O'Donnell, E. Murray, and C. May, *From theory to 'measurement' in complex interventions: Methodological lessons from the development of an e-health normalisation instrument.* BMC Medical Research Methodology, 2012. **12**(1): p. 69.
53. May, C. and T. Finch, *Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory.* Sociology, 2009. **43**(3): p. 535-554.
54. Murray, E., S. Treweek, C. Pope, A. MacFarlane, L. Ballini, C. Dowrick, T. Finch, A. Kennedy, F. Mair, C. O'Donnell, B. Ong, T. Rapley, A. Rogers, and C. May, *Normalisation process theory: a framework for developing, evaluating and implementing complex interventions.* BMC Medicine, 2010. **8**(1): p. 63.
55. Murray, E., J. Burns, C. May, T. Finch, C. O'Donnell, P. Wallace, and F. Mair, *Why is it difficult to implement e-health initiatives? A qualitative study.* Implementation Science, 2011. **6**(1): p. 6.
56. Feldman, M.S. and W.J. Orlikowski, *Theorizing practice and practicing theory.* Organization science, 2011. **22**(5): p. 1240-1253.
57. Walsham, G., *Cross-cultural software production and use: a structurational analysis.* MIS quarterly, 2002. **26**(4): p. 359-380.
58. Hyysalo, S., *Health technology development and use: from practice-bound imagination to evolving impacts.* 2010, New York: Routledge.
59. Pollock, N. and R. Williams, *Software and organisations: The biography of the enterprise-wide system or how SAP conquered the world.* 2008, New York: Routledge.
60. Pollock, N. and R. Williams, *e-Infrastructures: How Do We Know and Understand Them? Strategic Ethnography and the Biography of Artefacts.* Computer Supported Cooperative Work (CSCW), 2010. **19**(6): p. 521-556.
61. NHS England, Public Health England, Health Education England, Monitor, Care Quality Commission, and NHS Trust Development Authority, *Five year forward view*, 2014.
62. Foy, R., M.P. Eccles, G. Jamtvedt, J. Young, J.M. Grimshaw, and R. Baker, *What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review.* BMC Health Services Research, 2005. **5**: p. 50.
63. Grol, R.P.T.M., M.C. Bosch, M.E.J.L. Hulscher, M.P. Eccles, and M. Wensing, *Planning and Studying Improvement in Patient Care: The Use of Theoretical Perspectives.* Milbank Quarterly, 2007. **85**(1): p. 93-138.
64. Davidoff, F., M. Dixon-Woods, L. Leviton, and S. Michie, *Demystifying theory and its use in improvement.* BMJ Quality & Safety, 2015.
65. Funnell, S.C. and P.J. Rogers, *Purposeful program theory: effective use of theories of change and logic models.* 2011, San Francisco: John Wiley & Sons.
66. Robert, G. and N. Fulop, *The role of context in successful improvement.* Perspectives on context. A selection of essays considering the role of context in successful quality improvement. London: Health Foundation, 2014. **31**.
67. Byng, R., I. Norman, S. Redfern, and R. Jones, *Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation.* BMC Health Services Research, 2008. **8**(1): p. 274.
68. Ranmuthugala, G., F.C. Cunningham, J.J. Plumb, J. Long, A. Georgiou, J.I. Westbrook, and J. Braithwaite, *A realist evaluation of the role of communities of practice in changing healthcare practice.* Implement Sci, 2011. **6**: p. 49.
69. Greenhalgh, T., C. Humphrey, J. Hughes, F. Macfarlane, C. Butler, and R. Pawson, *How Do You Modernize a Health Service? A Realist Evaluation of Whole-Scale Transformation in London.* Milbank Quarterly, 2009. **87**(2): p. 391-416.
70. Westhorp, G., K. Stevens, and P.J. Rogers, *Using realist action research for service redesign.* Evaluation, 2016. **22**(3): p. 361-379.

71. Fletcher, A., F. Jamal, G. Moore, R.E. Evans, S. Murphy, and C. Bonell, *Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions.* Evaluation, 2016. **22**(3): p. 286-303.
72. Pearson, M., S.L. Brand, C. Quinn, J. Shaw, M. Maguire, S. Michie, S. Briscoe, C. Lennox, A. Stirzaker, T. Kirkpatrick, and R. Byng, *Using realist review to inform intervention development: methodological illustration and conceptual platform for collaborative care in offender mental health.* Implementation Science, 2015. **10**(1): p. 1-12.
73. Galliers, J., S. Wilson, R. Randell, and P. Woodward, *Final Report on Safer Handover*, 2009, City University London: London.
74. Carney, O., J. McIntosh, and A. Worth, *The use of the Nominal Group Technique in research with community nurses.* Journal of Advanced Nursing, 1996. **23**(5): p. 1024-1029.
75. Waugh, A., A. Austin, J. Manthorpe, C. Fox, B. Stephens, L. Robinson, and S. Iliffe, *Designing a complex intervention for dementia case management in primary care.* BMC Family Practice, 2013. **14**(1): p. 1-11.
76. Manzano, A., *The craft of interviewing in realist evaluation.* Evaluation, 2016. **22**(3): p. 342-360.
77. Ritchie, J. and L. Spencer, *Qualitative data analysis for applied policy research*, in *Analyzing qualitative data*, A. Bryman and R.G. Burgess, Editors. 1994, Routledge: London.
78. Cheyne, H., P. Abhyankar, and C. McCourt, *Empowering change: Realist evaluation of a Scottish Government programme to support normal birth.* Midwifery, 2013. **29**(10): p. 1110-1121.
79. Randell, R., N. Alvarado, S. Honey, J. Greenhalgh, P. Gardner, A. Gill, D. Jayne, A. Kotze, A. Pearman, and D. Dowding. *Impact of Robotic Surgery on Decision Making: Perspectives of Surgical Teams*. in *AMIA Annual Symposium Proceedings*. 2015. American Medical Informatics Association.
80. Thompson, C., D. McCaughan, N. Cullum, T.A. Sheldon, and P. Raynor, *Increasing the visibility of coding decisions in team-based qualitative research in nursing.* International Journal of Nursing Studies, 2004. **41**(1): p. 15-20.
81. Miles, M.B. and A.M. Huberman, *Qualitative data analysis: an expanded sourcebook.* 2nd ed. 1994, Thousand Oaks, California: SAGE.
82. Westhorp, G., *Using complexity-consistent theory for evaluating complex systems.* Evaluation, 2012. **18**(4): p. 405-420.
83. Westhorp, G., *Developing complexity-consistent theory in a realist investigation.* Evaluation, 2013. **19**(4): p. 364-382.
84. Rogers, Y., H. Sharp, and J. Preece, *Interaction design: beyond human-computer interaction.* 3rd ed. 2011, Chichester: John Wiley & Sons.
85. Robertson, S. and J. Robertson, *Mastering the requirements process: Getting requirements right.* 2012, Harlow: Addison-wesley.
86. Van de Ven, A.H. and A.L. Delbecq, *The nominal group as a research instrument for exploratory health studies.* American Journal of Public Health, 1972. **62**(3): p. 337-342.
87. Cantrill, J.A., B. Sibbald, and S. Buetow, *The Delphi and nominal group techniques in health services research.* International Journal of Pharmacy Practice, 1996. **4**(2): p. 67-74.
88. Sim, J., *Collecting and analysing qualitative data: issues raised by the focus group.* Journal of Advanced Nursing, 1998. **28**(2): p. 345-352.
89. Kitzinger, J., *Focus groups*, in *Qualitative research in health care*, C. Pope and N. Mays, Editors. 2006, Blackwell Publishing: Malden, Mass. p. 21-31.
90. Kidd, P.S. and M.B. Parshall, *Getting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research.* Qual Health Res, 2000. **10**(3): p. 293-308.
91. Galesic, M. and R. Garcia-Retamero, *Graph Literacy: A Cross-Cultural Comparison.* Medical Decision Making, 2011. **31**(3): p. 444-457.
92. Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* Advances in psychology, 1988. **52**: p. 139-183.
93. Burns Wendell, J., K. Holtzblatt, and S. Wood, *Rapid Contextual Design: A How to Guide to Key Techniques for User-Centered Design.* 2004, Burlington, MA: Morgan Kaufman.
94. Skolarus, T.A. and A.E. Sales, *Implementation issues: towards a systematic and stepwise approach*, in *Complex interventions in health: an overview of research methods*, D.A. Richards and I.R. Hallberg, Editors. 2015, Routledge: Oxford. p. 265-272.
95. McCormack, B., *Action research for the implementation of complex interventions*, in *Complex interventions in health: an overview of research methods*, D.A. Richards and I.R. Hallberg, Editors. 2015, Routledge: Abingdon. p. 300-311.
96. Pawson, R. and A. Manzano-Santaella, *A realist diagnostic workshop.* Evaluation, 2012. **18**(2): p. 176-191.
97. Yin, R.K., *Case study research: design and methods.* 3rd ed. 2003, Thousand Oaks, California: SAGE.
98. Hu, Y., *Stepped wedge, natural experiments and interrupted time series analysis designs*, in *Complex interventions in health: An overview of research methods*, D.A. Richards and I.R. Hallberg, Editors. 2015, Routledge: Abingdon. p. 200-212.
99. Hammersley, M. and P. Atkinson, *Ethnography: principles in practice.* 1995, London: Routledge.

100. Rycroft-Malone, J., M. Fontenla, D. Bick, and K. Seers, *A realistic evaluation: the case of protocol-based care.* Implementation Science, 2010. **5**(1): p. 38.
101. Greenhalgh, T. and D. Swinglehurst, *Studying technology use as social practice: the untapped potential of ethnography.* BMC Medicine, 2011. **9**(1): p. 45.
102. Mozaffar, H., K.M. Cresswell, L. Lee, R. Williams, and A. Sheikh, *Taxonomy of delays in the implementation of hospital computerized physician order entry and clinical decision support systems for prescribing: a longitudinal qualitative study.* BMC Medical Informatics and Decision Making, 2016. **16**(1): p. 1-14.
103. Øvretveit, J. and D. Gustafson, *Evaluation of quality improvement programmes.* Quality and Safety in Health Care, 2002. **11**(3): p. 270-275.
104. Ramsay, A.I.G., S. Turner, G. Cavell, C.A. Oborne, R.E. Thomas, G. Cookson, and N.J. Fulop, *Governing patient safety: lessons learned from a mixed methods evaluation of implementing a ward-level medication safety scorecard in two English NHS hospitals.* BMJ Quality & Safety, 2013. **23**: p. 136-146.
105. Benn, J., S. Burnett, A. Parand, A. Pinto, S. Iskander, and C. Vincent, *Studying large-scale programmes to improve patient safety in whole care systems: Challenges for research.* Social Science & Medicine, 2009. **69**(12): p. 1767-1776.
106. Cochrane Effective Practice and Organisation of Care Group (EPOC), *What study designs should be inlcuded in an EPOC review and what should they be called?*, 2013, Norwegian Knowledge Centre for the Health Services: Oslo.
107. Morris, J.V., P. Ramnarayan, R.C. Parslow, and S.J. Fleming, *Outcomes for Children Receiving Noninvasive Ventilation as the First-Line Mode of Mechanical Ventilation at Intensive Care Admission: A Propensity Score-Matched Cohort Study.* Critical Care Medicine, In press.
108. Simms, A.D., P.D. Baxter, B.A. Cattle, P.D. Batin, J.I. Wilson, R.M. West, A.S. Hall, C.F. Weston, J.E. Deanfield, and K.A. Fox, *An assessment of composite measures of hospital performance and associated mortality for patients with acute myocardial infarction. Analysis of individual hospital performance and outcome for the National Institute for Cardiovascular Outcomes Research (NICOR).* European Heart Journal: Acute Cardiovascular Care, 2013. **2**(1): p. 9-18.
109. R Core Team, *R: A language and environment for statistical computing*, 2014, R Foundation for Statistical Computing: Vienna, Austria.
110. Lee, K., S.Y. Jung, H. Hwang, S. Yoo, H.Y. Baek, R.-M. Baek, and S. Kim, *A novel concept for integrating and delivering health information using a comprehensive digital dashboard: An analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage.* International Journal of Medical Informatics, 2017. **97**: p. 98-108.
111. Etchegaray, J.M. and E.J. Thomas, *Comparing two safety culture surveys: Safety Attitudes Questionnaire and Hospital Survey on Patient Safety.* BMJ Quality & Safety, 2012. **21**(6): p. 490-498.
112. Maxwell, J.A., *A realist approach for qualitative research.* 2012, London: SAGE Publications.
113. Bugge, C., B. Williams, S. Hagen, J. Logan, C. Glazener, S. Pringle, and L. Sinclair, *A process for Decision-making after Pilot and feasibility Trials (ADePT): development following a feasibility study of a complex intervention for pelvic organ prolapse.* Trials, 2013. **14**(1): p. 353.
114. Wilson, D.T., R.E. Walwyn, J. Brown, A.J. Farrin, and S.R. Brown, *Statistical challenges in assessing potential efficacy of complex interventions in pilot or feasibility studies.* Statistical Methods in Medical Research, 2015.
115. Thabane, L., J. Ma, R. Chu, J. Cheng, A. Ismaila, L.P. Rios, R. Robson, M. Thabane, L. Giangregorio, and C.H. Goldsmith, *A tutorial on pilot studies: the what, why and how.* BMC Medical Research Methodology, 2010. **10**(1): p. 1-10.
116. Entwistle, V.A., M.J. Renfrew, S. Yearley, J. Forrester, and T. Lamont, *Lay perspectives: advantages for health research.* BMJ, 1998. **316**(7129): p. 463-466.