Contents lists available at ScienceDirect

Social Science & Medicine

Taking the heat or taking the temperature? A qualitative study of a large-scale exercise in seeking to measure for improvement, not blame

Natalie Armstrong[a,*], Liz Brewster[b], Carolyn Tarrant[a], Ruth Dixon[c,d], Janet Willars[a], Maxine Power[e], Mary Dixon-Woods[f]

[a] Department of Health Sciences, University of Leicester, Leicester, UK
[b] Lancaster Medical School, Lancaster University, Lancaster, UK
[c] Blavatnik School of Government, University of Oxford, Oxford, UK
[d] Department of Politics and International Relations, University of Oxford, Oxford, UK
[e] Haelo, Salford, UK
[f] The Healthcare Improvement Studies (THIS) Institute, University of Cambridge, Cambridge, UK

ARTICLE INFO

ABSTRACT

Measurement of quality and safety has an important role in improving healthcare, but is susceptible to unintended consequences. One frequently made argument is that optimising the benefits from measurement requires controlling the risks of blame, but whether it is possible to do this remains unclear. We examined responses to a programme known as the NHS Safety Thermometer (NHS-ST). Measuring four common patient harms in diverse care settings with the goal of supporting local improvement, the programme explicitly eschews a role for blame.

The study design was ethnographic. We conducted 115 hours of observation across 19 care organisations and conducted 126 interviews with frontline staff, senior national leaders, experts in the four harms, and the NHS-ST programme leadership and development team. We also collected and analysed relevant documents.

The programme theory of the NHS-ST was based in a logic of measurement for improvement: the designers of the programme sought to avoid the appropriation of the data for any purpose other than supporting improvement. However, organisational participants - both at frontline and senior levels - were concerned that the NHS-ST functioned latently as a blame allocation device. These perceptions were influenced, first, by field-level logics of accountability and managerialism and, second, by specific features of the programme, including public reporting, financial incentives, and ambiguities about definitions that amplified the concerns. In consequence, organisational participants, while they identified some merits of the programme, tended to identify and categorise it as another example of performance management, rich in potential for blame.

These findings indicate that the search to optimise the benefits of measurement by controlling the risks of blame remains challenging. They further suggest that a well-intentioned programme theory, while necessary, may not be sufficient for achieving goals for improvement in healthcare systems dominated by institutional logics that run counter to the programme theory.

1. Introduction

Quality measurement that focuses on important processes and outcomes, including clinical care and patient experience, is often seen as an essential feature of well-functioning healthcare systems (Conway et al., 2013). Prominent uses of measurement include identifying variations in care between different organisations or practitioners, surfacing the factors associated with high performance, and supporting replication and scaling of apparently successful approaches (Bradley et al., 2012).

Measurement is a defining characteristic of many quality improvement efforts, where techniques such as statistical process control and audit and feedback are routinely deployed to help practitioners monitor their local system performance and the responses of that system to improvement interventions (Portela et al., 2015; Ivers et al., 2012). Increasingly, measurement is also deployed in the context of performance management regimes and as an element of pay-for-performance schemes to address the demand for accountability and transparency that has become one of the central tropes of current thinking on

governance and regulation in modern healthcare (Mukamel et al., 2014; Brewster et al., 2016). These differing goals of measurement embody a number of tensions, including those relating to the balance between stimulating improvement and provoking unintended consequences – such as gaming (manipulation of data to look good), effort substitution (focus on the things being measured to the exclusion of other important activities) (Kelman and Friedman, 2009), shrinkage of professional responsibility (narrow focus on the things being measured) and excessive bureaucratic burden associated with data collection and reporting (Chassin et al., 2010; Meyer et al., 2012). The question of whether it is possible to capture the potential benefits of measurement while minimising the risks is thus a critical one.

One frequently made argument is that optimising the yield from measurement (and, conversely, averting its unwanted effects) requires controlling the risks of *blame*. Solberg et al.'s much-cited article (Solberg et al., 1997) distinguishes *measurement for improvement* from *measurement for accountability,* proposing that the appropriation of locally collected data for external accountability purposes may thwart the goal of improvement. These authors argue that only when fear and blame are "out of the equation", can everyone "concentrate on improvement rather than defensiveness" (p.138). Yet cultures of blame are pervasive in healthcare (Dekker and Hugh, 2014), resulting, Don Berwick argues, in measurement fostering fear and defensiveness rather than improved quality and safety:

> *Any good foreman knows how clever a frightened work force can be. In fact, practically no system of measurement - at least none that measures people's performance - is robust enough to survive the fear of those who are measured […] The inspector says, "I will find you out if you are deficient." The subject replies, "I will therefore prove I am not deficient" - and seeks not understanding, but escape.* (Berwick, 1989:53)

It remains unclear, however, whether it is possible to design and operate measurement systems for improving healthcare quality that evade the apparently negative effects of blame: field studies of measurement of quality and safety have remained rare (Dixon-Woods et al., 2012).

We suggest one useful way of gaining clarity and analytic purchase is to understand "data for accountability" and "data for improvement" as two different logics. We propose, as a more general principle, that quality improvement (QI) efforts founded in a particular logic may be overwhelmed by institutional logics operating at the field level. We provide empirical support for this analysis using an ethnographic study of a large-scale data collection programme in England. Known as the NHS Safety Thermometer, the programme is of particular interest for our purposes because it explicitly embraced a principle of "data for improvement" that eschewed a role for blame.

1.1. Blame

It is useful to begin by acknowledging that, though blame is widely discussed in the healthcare literature and at policy level, for example in relation to patient safety (Wachter and Pronovost, 2009), its definition tends to be somewhat vernacular. Political science, by contrast, has developed an extensive literature on blame that offers some helpful pointers to a more formal approach. Christopher Hood, for example, defines blame as the act of attributing something bad or wrong to some person or entity (Hood, 2011:6); it involves some (actual or perceived) harm or loss, as well as, crucially, an attribution of agency. Though he emphasises that blame is not always bad, Hood explains that, faced with external demands for accountability, blame avoidance may become a dominant preoccupation for organisations and institutions.

Many political science analyses offer a fairly muscular view of blame avoidance and blame engineering, describing a range of techniques and strategies that are purposefully chosen and implemented with specific (albeit often undeclared) intentions of deflecting or evading blame. We propose that, though much of the scrutiny has

focused on the deliberate or purposeful creation of blame engineering schemes, it is possible for a system to function latently as a blame distribution and attribution system even when not designed with that goal in mind – or indeed, even, as we shall show using the example of the NHS Safety Thermometer, when it seeks explicitly to disavow a role in blaming.

1.2. The NHS Safety Thermometer

The declared aim of England's NHS Safety Thermometer (NHS-ST) programme is to "provide a quick and simple method for surveying patient harms and analysing results so that you can measure and monitor local improvement and harm-free care over time" (NHS Digital, 2017). The four harms measured by the NHS-ST tool – pressure ulcers, harm from falls, urinary infection in patients with catheters, and venous thromboembolism (VTE) – account for a large proportion of avoidable injury in healthcare settings, and incur high human and economic costs (Power et al., 2016). Patients who incur none of these harms are deemed "harm-free".

The programme requires that staff caring for NHS patients England-wide in hospitals or community nursing settings (e.g. patients' homes) record the presence and severity of the four harms on a pre-specified day each month. The NHS-ST thus creates a monthly census amounting to approximately 200,000 patients. Data collection is the responsibility of frontline teams, who are asked to record information according to the definitions in Table 1. The resulting data, which are entered into spreadsheets and aggregated at organisational, regional, and national levels, are publicly available online along with national benchmarking data. Following a 2011 pilot, the NHS-ST was introduced across England (Power et al., 2016). Since 2012/13 use of the NHS-ST has attracted financial incentives. A Commissioning for Quality and Innovation (CQUIN) payment was introduced in 2012/13 in which a financial reward was linked to data collection, with the aim of establishing a baseline. Other incentives were introduced over time; since 2015 the NHS Standard Contract has required the collection of data on a monthly basis using the NHS-ST or another local collection method.

1.3. Programme theories and institutional logics

The NHS-ST and its associated policy framework can be understood as a quality improvement (QI) programme (Portela et al., 2015). Recent years have seen growing recognition of the importance of explicating the theories or models that underlie such programmes (Davidoff et al., 2015), including elaboration of the causal assumptions - what is sometimes known as a programme's "logic". Even when QI programmes appear to have a sound underpinning theory and logic, success is often evasive (Dixon-Woods and Martin, 2016). Many reasons for such failures can be identified, but one that has remained little examined, despite its rich explanatory potential, lies in the relationship between programme theories and institutional logics.

First proposed as a feature of institutional theory more than 25 years ago (Friedland and Alford, 1991), the literature on institutional logics has expanded greatly over the last quarter century, and now accommodates several different definitions. Broadly, however, it proposes that large-scale supraorganisational social structures tend to be characterised by distinctive sets of assumptions, values, beliefs, practices, and symbolic constructions (Friedland and Alford, 1991), offering repertoires to social actors that constrain - though do not fully determine - their choices, behaviours, and understandings. Friedland and Alford's original work focused on the central institutions of the contemporary West, such as professions, markets, and bureaucracies, but later theorists identified a hierarchical character to logics, such that "organisational fields and industries are viewed as having their own logics nested within societal level institutional orders" (Goodrick and Reay, 2011:375).

Though institutional logics contribute to the relatively stable nature