Applied Nursing Research 26 (2013) 269–272

Contents lists available at ScienceDirect

Applied Nursing Research

journal homepage: www.elsevier.com/locate/apnr

Metric-driven harm: An exploration of unintended consequences of performance measurement

Betty Rambur, PhD, RN [a,*], Carol Vallett, EdD [b], Judith A. Cohen, PhD, RN [a], Jill Mattuck Tarule, EdD [b]

[a] University of Vermont, College of Nursing and Health Sciences, Burlington, VT 05405, USA
[b] University of Vermont, College of Education and Social Services, Burlington, VT 05405, USA

ARTICLE INFO

Article history:
Received 24 March 2013
Revised 14 September 2013
Accepted 20 September 2013

Keywords:
Metric-driven harm
Unintended consequences
Public reporting of health care outcomes
Performance measurement
Quality improvement

ABSTRACT

Performance measurement is an increasingly common element of the US health care system. Typically a proxy for high quality outcomes, there has been little systematic investigation of the potential negative unintended consequences of performance metrics, including metric-driven harm. This case study details an incidence of post-surgical metric-driven harm and offers Smith's 1995 work and a patient centered, context sensitive metric model for potential adoption by nurse researchers and clinicians. Implications for further research are discussed.

Public reporting of performance outcomes is an increasingly common element of US health care. More recently, pay-for-performance has emerged as an intuitively appealing, common sense approach to improving quality and decreasing costs, with some studies suggesting better outcomes associated with measurement (Calikoglu, Murray, & Feeney, 2012). Other studies report inconsistent and sometimes contradictory evidence on effectiveness and patient benefits (Gillam, Siriwardena, & Steel, 2012; Jha, Joynt, Orav, & Epstein, 2012; Parker, Schwamm, Fonarow, Smith, & Reeves, 2012; Woolhandler & Ariely, 2012), only modest improvements (Bardach et al., 2013), or raise new questions about sustainability over time (Petersen et al., 2013). Moreover, studies outside the US—where publically reported metrics have a longer history both within and outside of health care—cite some of the challenges that accompany public reporting of performance. These include gaming, synecdoche (taking a part to stand for a whole), measurement problems (Bevan & Hood, 2006) and negative behavior changes among managers (Ossege, 2012). Additionally, manipulation of any sort of quality data has been found to improve reported metrics while actually decreasing the quality of service and at a higher cost to consumers (Kuhn & Siciliani, 2009).

Closer to home, calls for exploration of unintended consequences of quality metrics span nearly a decade (Werner & Asch, 2005; Knot, 2012). Indeed, several contemporary studies raise intriguing questions. Fenton, Jerant, Bertakis, and Franks (2012), for example, in a large ($n = 51{,}946$) cohort study report that the seemingly positive patient outcome of higher patient satisfaction was associated with increased mortality. Although association is certainly not causation, other metrics have been more clearly linked to harm. A notable example was a metric intended to catalyze prompt treatment of community-acquired pneumonia for those presenting to emergency rooms with possible pneumonia. The result, however, was negative consequences that included misdiagnosis and inappropriate use of antibiotics (Kanwar, Brar, Khatib, & Fakih, 2007; Wachter, Flander, Fee, & Pronovost, 2008; Welker, Huston, & McCue, 2008).

Powell et al. (2012) reported several negative unintended outcomes for patients as a result of performance measurement systems. These included possible inappropriate clinical care, a lack of focus on patient concerns, and concessions in patient independence and education. In addition to patient impacts, Powell also provided examples of negative consequences on the members of primary care teams, especially nurses.

Despite notable exceptions such as the Powel et al study, the notion of unintended consequences of quality metrics—including metric-driven harm—is largely absent from the nursing literature. This gap in exploration is significant, as there is evidence that well-intended quality metrics in health care can actually lead to decreases in care and increases in mortality (Marshall, Shekelle, Brook, & Leatherman, 2000), while the publicly disclosed metrics show quality improvements. Clearly, the potential for such largely hidden impacts on patients, their families, and the nurses who treat these patients has not been factored into the quality metric equation.

At the same time, the empirical evidence that nurses' subjective impression of quality aligns with actual quality (McHugh & Stimpfel, 2012) suggests that nurses are well positioned to lead not only the

* Corresponding author. Tel.: +1 802 656 3401.
E-mail address: betty.rambur@uvm.edu (B. Rambur).

0897-1897/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.apnr.2013.09.001

investigation of metric-driven harm, but also devise corrective action. The purpose of this paper is to explore a case study of metric-driven behavior and metric-driven harm and posit a model for further exploration of this phenomenon by nurse clinicians and scientists.

1. Method

This case study was part of a larger study on moral distress, moral eustress, and organizational virtue. The researchers had examined stress in the workplace (Cohen, Tarule, Rambur, & Vallett, 2011) and had proposed a model, the Moral Cascade (Rambur, Vallett, Cohen, & Tarule, 2010), that depicted the process of moral decision making, which could lead to either moral distress (harmful stress) and/or moral eustress (beneficial stress) for the individual. Further, the model posits interplay among these individual outcomes and organizational values, practices, and trustworthiness, that is, organizational virtue as conceptualized by Cameron, Bright, and Caza (2004). In order to test The Moral Cascade, a pilot, mixed-methods design study was undertaken in both healthcare and higher education institutions. Subjects were recruited using snowball sampling. Semi-structured interviews were conducted with each subject, followed by measurement of the subject's moral development stage. The construct —moral development stage—was operationally defined as the score on the Defining Issues Test, an instrument with well-established reliability and validity (Rest, 1975). Additional data were collected that are not relevant to the case, and thus are not reported herein.

The interview data were analyzed using a grounded approach to identify themes. The case emerged from one such interview and, while indicative of a form of moral distress for the research on The Moral Cascade, it also presented the poignant and powerful vignette about an aspect of unintended consequences of quality measurement initiatives.

The interview began with the researcher asking the subject, in this case a staff registered nurse, to recount a time when faced with a situation or decision that involved or created moral distress. Moral distress was described by Corley (2002) as a time when the subject had a decision to make, knew what to do, but was thwarted from making that decision. Identifying data have been changed to ensure subject and patient confidentiality.

2. The case

I just feel at times that we are not looking at patients as people, but as statistics. But really that this person was a mother and a sister, and a daughter at one time, and that we were losing sight of that, and that I think that that happens frequently....

A patient in the surgical intensive care unit had an advanced directive in his chart that stated that if there was ever a time when he was unable to participate in his health decision making, he would want his care focused toward comfort. He had undergone surgery approximately 20 days before and had not recovered. In addition to several co-morbidities, the patient had been ventilated, intubated, had a feeding tube, and was receiving dialysis. The patient was non-communicative with all nurses, care-givers, physicians, and family except for a persistent request to “please help me and stay away from me.”

The patient's family members approached the nurse to discuss what could be done for comfort care and how to discontinue treatment. In spite of an attempt to engage the physician in this dialogue, the staff nurse stated “nothing happened.” When the staff nurse approached the head nurse about a consultation with the palliative care unit, the staff nurse was told that until the patient was 30 days post-surgery, they could not discuss discontinuing treatment. In the words of the staff nurse, the head nurse observed that there is a “30 day mortality rate that is a very important statistic” for the unit. She further noted that until the patient passed the 30 day mark, it was unlikely that the physician would allow palliative care to become involved.

Despite all attempts by the staff nurse to address continuation of treatment, which the staff nurse believed was not consistent with the patient's wishes and prognosis, the physicians continued treatment of the patient. In the words of the staff nurse: “Nothing altered the course of treatment until after those 30 days.”

After the 30 day window, there was a meeting with the patient's family and the physician, and the decision was made to remove the patient from the ventilator and allow the patient to die. The resident, not the physician, wrote the order. The physician never returned to see the patient or his family, despite the family's request that he do so. The staff nurse stated that the physician's refusal to see the family added to her moral distress since it seemed to confirm for her that the physician “really did care about the statistics” and not about a patient and family-centered approach to patient care. Although this staff nurse recounted the case as an incident of moral distress, she also added that it was not an isolated incident, observing that the number of patients on a ventilator, getting treatments that would not help them, is astounding.

The moral distress that emerged was recounted by the staff nurse as stemming from a perceived ethical conflict between 1) the patient's advanced directives, prognosis, and family wishes (all aligned with each other) and 2) organizational structure, a unit's 30 day post operative survival rate as a perceived success metric, and physician authority. Again, the latter were all aligned, but—in the perception of the staff nurse—in direct opposition to the patient's advanced directives, prognosis, and family wishes.

3. Discussion

Public dissemination of outcome data is intended to drive clinical decision-making, particularly when the outcomes are linked to reimbursement (Marshall, Shekelle, Leatherman, & Brook, 2000). Yet unintended consequences merit further consideration. Specifically, this case study suggests the need for further investigation of the extent to which futile treatment is directed to patients between days 1–30 post operatively, and the incidence of discontinuation shortly thereafter. While the widespread use of the 30 day post-surgical mortality metric is important, this case study illustrates that the human toll that can be exacted when trying to meet the 30 day goal takes precedence over other patient needs. Recent evidence also quantifies substantial financial cost of futile treatment in critical care (Huynh et al., 2013).

The findings of this case study add to the body of evidence that negative unintended consequences are occurring in American hospitals and medical care (Ganz et al., 2007; Mannion & Braithwaithe, 2012; Powell et al., 2012), and complement findings from the UK primary health care system (Lester, Hannon, & Campbell, 2011). While McHugh and Stimpfel (2012) demonstrated that nurses are in an ideal position to report quality care metrics, by implication these results also suggest that nurses are in an ideal position to identify negative unintended consequences, metric-driven cost, and metric-driven harm, and to direct corrective action, provided they have the tools and power to do so.

4. An organizing framework for exploring metric-driven harm

A foundational work in the United Kingdom about unintended consequences of publically reported performance metrics was promulgated by Smith (1995). Smith outlines eight categories of unintended consequences of public reporting of performance metrics that can be translated to health quality outcome metrics. These

categories are identified in Table 1, slightly modified for application to the contemporary US health care system.

The preceding case offers a clear example of measure fixation, and perhaps also tunnel vision, myopia, and gaming. Measure fixation is illustrated with the focus on the 30 day survival. Both tunnel vision and myopia could refer, in this case, to the physician and head nurse being unable to entertain any alternative or competing perspectives on appropriate care for the patient. Gaming could come in to play if the physician, who normally would be expected to accommodate to the aligned end-of-life directives of the patient/family, changes behavior to accommodate the quality metric.

Yet naming the potential elements that may spur negative unintended consequences, while an important first step, does not immediately or seamlessly lead to obvious, related corrective action. Smith's conceptual organization, however, does identify strategies for minimizing negative unintended consequences. Interestingly, these strategies align well with nurses' still unrecognized potential for meaningful participation in metric development and refinement. Smith, for example, identifies involvement of staff at all levels (p. 302) as a key strategy to mitigate the negative effects of all behaviors listed in the taxonomy. Nurses, as front line providers, see both the positive and negative implication of quality metrics, particularly if they are informed about the potential for such effects and do not suffer from their own version of tunnel vision, for example, a prioritization of measured elements of care over unmeasured elements of course.

A second strategy suggested in Smith's seminal overview is to "keep the system under constant review" (p. 302). This creates challenges in a system in which the metrics themselves are mandated by payer systems or accreditation bodies. Nurses are, however, in a unique position to empirically study the impact of quality indicators. This might include doing a careful audit of institutional processes and interactions among different processes and outcomes to measure the performance of the quality metric in producing the desired outcome. Moreover, Pronovost and Lilford (2011) offer five strategies for "improving the performance of performance metrics" (p. 569) to ensure that the metrics are valid, transparent, longitudinal, and applied in a standardized way across institutions. In addition, they recommend creation of an independent organization to monitor ongoing development and application of metrics.

As Smith notes, "most measures are proxies for outputs" (p.291). Thus, a more robust manner to assess the outputs in this case scenario and others like it could include elements such as incidence and timing of referrals to palliative care and review of hospital ethics committee's involvement, overlaying family satisfaction data. Furthermore, the case findings suggest that review of aggregate mortality data with particular attention to any spiking of mortality rates in the period immediately following 30 days is warranted.

This case also illustrates the context and antecedents to one nurse's moral distress. When nurses' knowledge of the patient's wishes and the context of those wishes are not integrated into the overall plan of care, or when nurses are thwarted from doing what they know to be right in the particular situation, moral distress ensues. Over time, moral distress can lead to personal negative outcomes of despair, frustration, anger, loss of personal meaning and organizational ineffectiveness. It can also lead to negative organizational outcomes such as higher employee turnover rates, absenteeism, disengagement and therefore decreased quality and productivity in the work, as well as patient/client dissatisfaction (Rambur et al., 2010).

Taken as a whole, the implications from this case study also suggest that the barriers to shared decision-making identified in primary care—harried physicians, insufficient provider training, and the lack of shared decision-making (Friedberg, Van Busum, Wexler, Bowen, & Schneider, 2013)—also exist in hospital settings. These authors suggest that the antidote will include clinical information systems to track patients through the decision-making process and broad, empowered, informed engagement of a provider team potentially inclusive of a "decision coach" rather than default reliance on the physician. In the case heretofore detailed, and using the example of advanced directive as an illustration only, this approach would include a clinical information system that retained the centrality of the patient's wishes, in this case an advanced directive, coupled with 1) explicit prompts requiring overt consideration and questioning of continued treatment and 2) engaged involvement of the entire patient care team. Moreover, a system in which the quality metric(s) crosscuts patient, nurse, and physician knowledge (see Fig. 2) rather than serving as a monolithic central target (Fig. 1) would serve patients, providers, and society-at-large by potentially decreasing metric-driven harm.

The case detailed illustrates how metric-centered behavior produced metric-driven harm. Moreover, although a patient's right to self-determination is called "universal principle of health care ethics," in the

Table 1
A taxonomy of unintended consequences, adapted from Smith (1995).

Tunnel vision refers to the prioritization of financially incentivized care over other valuable care, or a prioritization of measured elements of care over unmeasured care of equal or greater value.
Measure fixation is a focus on a particular measurement without reflection on how maximization of the particular outcome can miss or even be in opposition to the underlying objective of care. It also refers to focus on a particular measurement without attention to the distressed caused to a patient by maximization of the metric.
Acontextual actions or adverse selection refers to the potential to choose patients who can maximize positive measurement while deferring or refusing to treat more vulnerable, more seriously ill patients.
Misrepresentation refers to the deliberate manipulation of data so that the reported behavior differs from the actual behavior.
Gaming refers to manipulation of behavior to meet targets. It differs from misrepresentation in that it is a contortion of actual behavior, not merely reported data.
Myopia refers to a focus on short term, measurable performance at the expense of legitimate long term consequences and goals.
Suboptimization is the pursuit of narrow local objectives by managers at the expense of the objectives of the organization as a whole.
Ossification is the inhibition of innovation. It may include organizational paralysis brought about by an excessively rigid system of performance measures.

Fig. 1. Performance metric centric schematic.

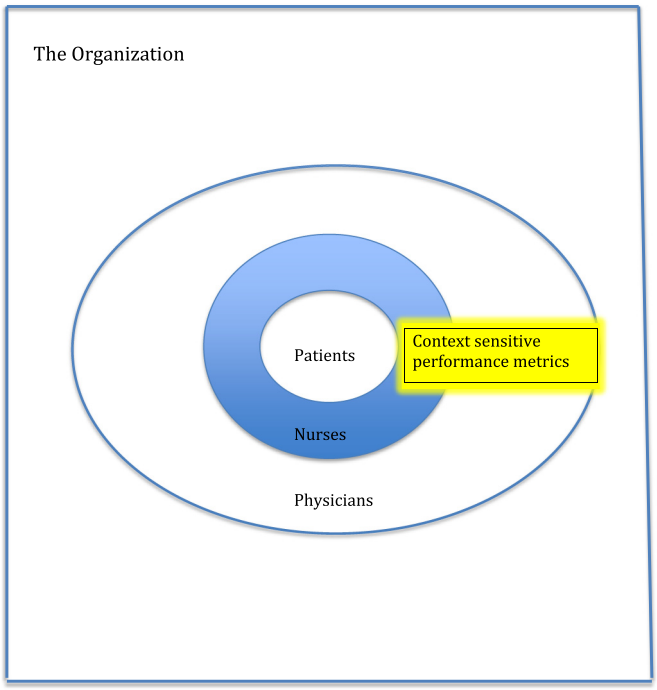

Fig. 2. Patient centered metrics.

actual practice setting, as the case illustrates, the patient and family are often powerless and voiceless in the face of metric-centered behavior. Fig. 2 illustrates a patient-centered model that is context sensitive within the organizational domain. Nurses have unique knowledge and perspective in this domain, in part because of their close proximity to patients and their wishes. Thus, to realize a vision of patient-centered, context sensitive care, nurses must truly be equal partners in clinical decision making. Furthermore, nurses are distinctly positioned to assess and, as necessary, redesign performance metrics as well as other aspects of health delivery and processes. This aim aligns with the Institutes of Medicine (2010) recommendation that health care organizations "support and help nurses in taking the lead in developing and adopting innovative, patient-centered models" (p.2). To that end, further exploration of the use of Smith's taxonomy is warranted, with particular attention to the potential for nurses to be key leaders in prevention of metric-driven harm.

Acknowledgments

The authors acknowledge the financial support of University of Vermont College of Nursing and Health Sciences Incentive Grant.

References

Bardach, N., Wang, J., DeLeon, S., Shih, S., Boscardin, W. J., Goldman, L. E., & Dudley, A. (2013). Effects of pay-for-performance incentives on quality of care in small practices with electronic health records: A randomized trial. *JAMA*, *310*(10), 1051–1059.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517–538.

Calikoglu, S., Murray, R., & Feeney, D. (2012). Hospital pay-for-performance programs in Maryland produced strong results, including reduced hospital-acquired conditions. *Health Affairs*, *31*(12), 2649–2658, http://dx.doi.org/10.1377/hlthaff.2012.0357.

Cameron, K., Bright, D., & Caza, A. (2004). Exploring the relationship between organizational virtuousness and performance. *American Behavioral Scientist*, *47*(6), 766–790.

Cohen, J. A., Tarule, J. M., Rambur, B. A., & Vallett, C. (2011). Stress and the workplace. *Handbook of Stress, Coping, and Health: Implications for Nursing Research, Theory, and Practice* (pp. 310).

Corley, M. C. (2002). Nurse moral distress: A proposed theory and research agenda. *Nursing Ethics*, *9*(6), 636–650, http://dx.doi.org/10.1191/0969733002ne557oa.

Fenton, J. J., Jerant, A. F., Bertakis, K. D., & Franks, P. (2012). The cost of satisfaction. *Archives of Internal Medicine*, *172*, 405–411.

Friedberg, J., Van Busum, K., Wexler, R., Bowen, M., & Schneider, E. (2013). A demonstration of shared decision making in primary care highlights barriers to adoption and potential remedies. *Health Affairs*, *32*, 268–275.

Ganz, D., Wenger, N., Rother, C., Kamberg, D., Change, J., & MacLean, C. (2007). The effects of a quality improvement initiative on the quality of other aspects of health care: The law of unintended consequences? *Medical Care*, *45*(1), 8–18.

Gillam, S. J., Siriwardena, A. N., & Steel, N. (2012). Pay-for-performance in the United Kingdom: Impact of the quality and outcomes framework—A systematic review. *The Annals of Family Medicine*, *10*(5), 461–468.

Huynh T, Kleerup E, Wiley J, Savitsky T, Guise D, Garger B, Wenger N. The frequency and cost of treatment perceived to be futile in critical care. JAMA Internal Medicine, doi: 10:1001/jamainternmed.2013:10261. Published online September 9, 2013

Institutes of Medicine (2010). The future of nursing: Leading change, advancing health. Washington, DC: Author.

Jha, A. K., Joynt, K. E., Orav, E. J., & Epstein, A. M. (2012). The long-term effect of premier pay for performance on patient outcomes. *New England Journal of Medicine*, *366*(17), 1606–1615.

Kanwar, M., Brar, N., Khatib, R., & Fakih, M. G. (2007). Misdiagnosis of community-acquired pneumonia and inappropriate utilization of antibiotics: Side effects of the 4 h antibiotic administration rule. *Chest*, *131*, 1865–1869.

Knot, U. (2012). Exploring the risk of unintended consequences of quality improvement efforts. *Journal of the American College of Cardiology*, *60*(9), 812–813.

Kuhn, M., & Siciliani, L. (2009). Performance indicators for quality with costly falsification. *Journal of Economics & Management Strategy*, *18*(4), 1137–1154.

Lester, H. E., Hannon, K. L., & Campbell, S. M. (2011). Identifying unintended consequences of quality indicators: A qualitative study. *BMJ Quality and Safety*, *20*, 1057–1060.

Mannion, R., & Braithwaithe, J. (2012). Unintended consequences of performance measurement in healthcare; 20 salutary lessons from the English National Health Service. *Internal Medicine Journal*, *569–574*.

Marshall, M., Shekelle, P., Brook, R., & Leatherman, S. (2000). Dying to know public release of information about quality of health care. DTIC document. Retrieved from. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA384942.

Marshall, M., Shekelle, P., Leatherman, S., & Brook, R. (2000). The public release of performance data: What do we expect to gain? A review of the evidence. *Journal of the American Medical Association*, *283*(14), 1866–1874.

McHugh, M. D., & Stimpfel, A. W. (2012). Nurse reported quality of care: A measure of hospital quality. *Research in Nursing & Health*, *35*(6), 566–575, http://dx.doi.org/10.1002/nur.21503.

Ossege, C. (2012). Accountability–Are we better off without it?: An empirical study on the effect of accountability on public managers' work behavior. *Public Management Review*, *14*(5), 585–607.

Parker, C., Schwamm, L. H., Fonarow, G. C., Smith, E. E., & Reeves, M. J. (2012). Stroke quality metrics systematic reviews of the relationships to patient-centered outcomes and impact of public reporting. *Stroke*, *43*(1), 155–162.

Petersen, L., Simpson, K., Pietz, K., Urech, T., Hysong, S., Profit, J., Conrad, D., & Dudley, A. (2013). Effects of individual physician-level and practice level financial incentives on hypertension: A randomized trial. *JAMA*, *310*(10), 1042–1050.

Powell, A. A., White, K. M., Partin, M. R., Halek, K., Christianson, J. B., Neil, B., et al. (2012). Unintended consequences of implementing a national performance measurement system into local practice. *Journal of General Internal Medicine*, *27*(4), 405–412.

Pronovost, P., & Lilford, R. (2011). A roadmap for improving the performance of performance measures. *Health Affairs*, *30*(4), 569–573.

Rambur, B. A., Vallett, C. M., Cohen, J. A., & Tarule, J. M. (2010). The moral cascade: Distress, eustress, and the virtuous organization. *Journal of Organizational Moral Psychology*, *1*(1), 1–14.

Rest, J. (1975). Longitudinal study of the defining issues test of moral judgment: A strategy for analyzing developmental change. *Developmental Psychology*, *11*(1), 738–748.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, *18*(2–3), 277–310.

Wachter, R., Flander, S., Fee, C., & Pronovost, P. (2008). Public reporting of antibiotic timing in patients with pneumonia: Lessons from a flawed performance measure. *Annals of Internal Medicine*, *149*(1), 20–32.

Welker, J., Huston, M., & McCue, J. (2008). Antibiotic timing an errors in diagnosing pneumonia. *Archive of Internal Medicine*, *168*(4), 351–356.

Werner, R., & Asch, D. (2005). The unintended consequences of publicly reporting quality information. *Journal of the American Medical Association*, *293*(10), 1239–1244.

Woolhandler, S., & Ariely, D. (2012). Will pay for performance backfire? Insights from behavioral economics. Health affairs (blog). October 20, 2012 retrieved from. http://healthaffairs.org/blog/2012/10/11will-pay-for-performance-backfire-insightsfrom-behavioraleconomics.