Stakeholders in safety: patient reports on unsafe clinical behaviors distinguish hospital mortality rates

LSE Research Online URL for this paper: http://eprints.lse.ac.uk/103949/

Version: Accepted Version

Article:

Stakeholders in Safety: Patient Reports on Unsafe Clinical Behaviors Distinguish Hospital Mortality Rates

Abstract

Patient safety research has adapted concepts and methods from the workplace safety literature (safety climate, incident reporting) to explain why patients experience unintentional harm during clinical treatment in hospital (*adverse events*). Consequently, patient safety has primarily been studied through data generated by healthcare staff. However, because adverse events relate to patient injuries, it is suggested that patients and their families may also have valuable insights for investigating patient safety in hospitals. We conceptualized this idea by proposing that patients are stakeholders in hospital safety who, through their experiences of treatments and independence from institutional culture, can provide valid and supplementary data on unsafe clinical care. In 59 UK hospitals we investigated whether patient evaluations of care ($n = 23,287$ surveys) and the safety information contained in healthcare complaints ($n = 2,017$, containing 2.5 million words) explained variance in excess patient deaths (hospital mortality) beyond staff evaluations of care ($n = 49,302$ surveys) and incident reports ($n = 242,859$). The severity of reports on unsafe clinical behaviors (error and neglect) communicated in patient' healthcare complaints explained additional variance in hospital-level mortality rates beyond that of staff-generated data. The results indicate that patients provide valid and supplementary data on unsafe care in hospitals. Generalized to other organizational domains, the findings suggest that non-employee stakeholders should be included in assessments of safety performance if they experience or observe unsafe behaviors. Theoretically, it is necessary to further examine how concepts such as safety climate can incorporate the observations and outcomes of stakeholders in safety.

Key words: Safety performance, complaints, patient safety, incident reports, safety climate

Approximately 10% of patients experience an adverse event (unintended harm during treatment) in hospital, with half of such events being preventable and 14% resulting in disability or death (de Vries, Ramrattan, Smorenburg, Gouma, & Boermeester, 2008). To improve patient safety and understand why adverse events occur, researchers have adapted concepts (e.g., safety climate) and methods (e.g., employee surveys, incident reporting) used by applied psychologists to explain and reduce workplace accidents (Flin, 2007; Vincent, 2011). As a consequence, patient safety in hospitals has primarily been studied through the collection and analysis of data generated by healthcare staff. However, because the target of patient safety is patient rather than employee outcomes, patients and their families (hereinafter 'patients') may also have valuable insight on the safety of care (Davis, Sevdalis, Neale, Massey, & Vincent, 2013; Papanicolas & Figueroa, 2019). The idea that patients can provide information to explain and monitor patient safety in hospitals is significant because, to date, applied psychological research on safety in organizations has mostly used employee-data to build theory and study accidents, and it suggests that non-employees may also be able to provide safety data. To develop and establish this idea, we investigate the validity and added value of using patient-generated data (collected through patient surveys and healthcare complaints) to investigate the safety of hospital care.

Safety in Organizations

Patient safety is one of several research domains (e.g., workplace safety, process safety) in which psychologists explore the causes of accidents and physical harm in organizations (Beus, McCord, & Zohar, 2016; Hopkins, 2009). The application of psychological constructs (e.g., safety climate) to explain safety outcomes in different domains is argued as beneficial, because the cross-fertilization of ideas, methods, and results between distinct fields of inquiry facilities a holistic "understanding of how to manage the full range of safety issues in organizations"

(Hofmann, Burke, & Zohar, 2017, p. 384). Building on the idea that integrating safety research

from different domains can lead to both a more holistic analysis of safety outcomes in

organizations and new research questions, we compare and integrate the literature on patient

safety and workplace safety (see Figure 1). This provides the theoretical basis for investigating

patient safety through patient-generated data, and reveals that external (i.e., non-employee) reports

on safety behaviors in organizations may be useful in domains beyond patient safety.

INSERT FIGURE 1 HERE

Workplace Safety and Patient Safety: Similarities

Research on employee and patient safety assumes shared determinants. For both, safety

climate and leadership demonstrate the organizational prioritization of safety, which in turn

predicts employee safety performance (i.e., safety compliance and participation for avoiding

worker injury, and error-free and high-quality care for avoiding patient harm) and occupational or

medical accidents (Agnew, Flin, & Mearns, 2013; Beus et al., 2016; Christian, Bradley, Wallace,

& Burke, 2009; Clarke, 2013; Flin, 2007; Griffin & Hu, 2013; Griffin & Neal, 2000; Katz-Navon,

Naveh, & Stern, 2005; Singer, Lin, Falwell, Gaba, & Baker, 2009). The knowledge, skills, and

motivations of staff to behave safely mediate the link between safety climate and unsafe acts, with

contextual factors (e.g., policies) also being key antecedents (Flin, 2007; Griffin & Neal, 2000;

Leroy et al., 2012; Wakefield, McLaws, Whitby, & Patton, 2010; Weaver et al., 2013).

In terms of measurement and empirical findings, the workplace safety and patient safety

literatures are also analogous. Safety climate is measured through employee surveys, and both

employee harm (e.g., nursing injuries) and patient harm are predicted by similar constructs (e.g.,

employee perceptions of management commitment to safety) and instruments (Hofmann & Mark,

2006; Taylor et al., 2012). Furthermore, to investigate safety behaviors and safety incidents,

researchers in both fields collect incident and safety reports from employees, co-workers, and

supervisors; due to their expertise, institutional roles, and proximity to safety management, these

organizational members can provide insight into unsafe behaviors, the reasons why such

behaviors occur, and their consequences (Agnew et al., 2013; Christian et al., 2009; Katz-Navon,

Naveh, & Stern, 2009; Vincent, 2011; Xia, Griffin, Wang, Liu, & Wang, 2018).

Workplace Safety and Patient Safety: Differences

Although initially similar, the different foci of the workplace safety and patient safety

literatures (employees and patients) has led to divergences in conceptualization and measurement.

The focus of safety climate in patient safety research is the prioritization of safe care delivery to

patients rather than the avoidance of employee workplace accidents (Flin, 2007; Halligan &

Zecevic, 2011; Olsen, 2010). Furthermore, while safe patient care is also determined by the

knowledge, skills, and motivations of employees, these factors tend to relate to clinical practices

(Singer et al., 2009; Vincent, 2011; Vogus, Sutcliffe, & Weick, 2010), with the behaviors crucial

for avoiding patient harm differing from the safety compliance and safety participation behaviors

important for avoiding workplace accidents (Griffin & Neal, 2000). Specifically, because patient

safety is integral to taskwork (Motowidlo & Van Scotter, 1994) in clinical care and is not a

parallel activity (e.g., following safety rules while operating machinery), researchers have focused

on erroneous or neglectful behaviors in healthcare delivery that are proximal (e.g., misdiagnosis

leading to death) or distal (e.g., ignoring hygiene rules, which leads to infection) causes of harm

(Dixon-Woods, Suokas, Pitchforth, & Tarrant, 2009; Reader & Gillespie, 2013; Vincent, 2011).

A further key difference is that research on workplace safety explains occupational harm

through data (e.g., safety climate, incidents, behaviors) provided by those who potentially cause

and/or experience accidents (i.e., employees). By contrast, adverse events are understood through

data provided by the staff who strive to provide safe care, and the experiences of those (i.e.,

patients) who experience harm have mostly been neglected (Harrison et al., 2015; Sahlström,

Partanen, & Turunen, 2018). This neglect is due to data validity concerns, minimal theorization of

the value of patient data, and the absence of a parallel research stream in the workplace safety

literature. However, in the context of limited success in reducing adverse event rates, patient

reports on unsafe care may provide supplementary insight into why unintended patient harm

occurs and how it can be avoided (O'Hara et al., 2018; Shojania & Thomas, 2013; Wachter, 2010;

Walton et al., 2017). A growing literature on patient experiences of hospital safety (e.g., using

incident reports, surveys, complaints) has emerged and indicates that patients can reveal distinct

safety problems (Levtzion-Korach et al., 2010), for instance, errors or neglectful acts unseen or

unreported by staff (e.g., diagnosis errors, continuity issues, not cleaning wounds) (Armitage et

al., 2018; Davis et al., 2013; Gillespie & Reader, 2018; Walton et al., 2017). Nevertheless, the

value of patient-generated safety data is unconceptualized, untested, and not examined in terms of

significance for patient safety or the wider safety literature.

Patients as Stakeholders in Safety

To conceptualize and investigate the role of patients in patient safety, the organizational

stakeholder literature is instructive. Because patients are participants (Ocloo & Matthews, 2016)

rather than formal members of healthcare organizations (i.e., they are service users, often without

clinical training, or contracted behavioral obligations), they can be considered organizational

stakeholders: "any group or individual who can affect or is affected by the achievement of the

organization's objectives" (Freeman, 1984, p. 46). According to this literature, stakeholders can

have important information for optimizing and legitimizing managerial decision-making due to

their alternative experiences and independent insight into an institution's performance (Beierle,

2002; Fang, Palmatier, & Evans, 2008; Smith & Ingram, 2002). This observation is significant for patient safety, as it suggests that patients might be conceptualized as "stakeholders in safety," whose experience and stake in the provision and consequences of healthcare delivery can provide distinct and useful information on the unsafe behaviors that lead to unintended patient harm.

Concretely, patient stakeholder observations on safety are valuable for two reasons. First, due to their alternative perspective on clinical treatments (i.e., as recipients), patients may observe and report different safety problems to healthcare employees. Supporting this assertion, research has shown that while clinicians tend to report on unambiguous medical errors, patients report on everyday behaviors that are important for safety, yet challenging to monitor reliably: for instance, mistakes in note-taking, neglectful care, ignoring patient information, medication errors, or ignoring alarms (Davis et al., 2013; O'Hara et al., 2018; Rathert, Brandt, & Williams, 2012; Walton et al., 2017). Second, patients are independent of the cultural factors that inhibit clinician safety reporting, for instance, reluctance to admit mistakes, fear of retribution, or lack of institutional learning (Pfeiffer, Manser, & Wehner, 2010; Waring, 2005). Investigations of major hospital failures (Francis, 2013) have found that healthcare complaints from patients reporting unsafe behavior (and not employee reporting) provide early signs of systemic failings in safety.

Thus, by conceptualizing patients as stakeholders in safety, the validity and importance of using patient reports about unsafe hospital care can be explained. This theorization arises from the adaptation of concepts and methods from the workplace safety literature to explain and measure patient safety, and the emergent incongruities. In a complementary fashion, the idea that patient-generated data can be used to study safety in healthcare institutions has implications for workplace safety research. For instance, research has shown that low safety climate and the poor or inconsistent implementation of safety systems negatively impacts employee incident reporting

(Probst, Brubaker, & Barsotti, 2008; Reason, 2000; Zohar, 2010). This has necessitated the use of

different data sources and perspectives (e.g., co-workers) to collect information on safety

behaviors (Beus et al., 2016; Griffin & Hu, 2013; Xia et al., 2018). Demonstrating that patients

provide valid and useful data on patient safety could yield a new source of data – external

stakeholders – for studying safety in settings where unsafe employee behaviors are consequential

for both non-employees and employees (e.g., public transport, food standards, building safety).

Current Study

To investigate the validity and added value of patient-generated information on patient

safety in healthcare institutions, we analyzed data pertaining to unsafe care contained within

patient experience surveys and healthcare complaints. We investigated the relationship of these

data with the Summary Hospital-level Mortality Indicator (SHMI) in UK National Health Service

(NHS) hospitals. SHMI is considered a proxy measure of patient safety (Bottle, Jarman, & Aylin,

2011) due to it capturing excess deaths potentially caused by neglectful care and medical errors.

Patient experience surveys are administered annually in the NHS, and are used to generate

patient-centered evaluations of hospital care (hereinafter *patient evaluations of care)*, with reports

of problems in healthcare delivery (e.g., treatment delays, poor responsiveness, feeding problems)

that can lead to unsafe outcomes (e.g., due to patients not receiving timely treatment, suffering

malnutrition) being solicited (Flott, Graham, Darzi, & Mayer, 2017; Raleigh, Frosini, Sizmur, &

Graham, 2012). Due to capturing data on unsafe clinical behaviors, patient evaluations of care are

theorized as a potential indicator of patient safety in hospitals (Doyle, Lennox, & Bell, 2013), and

thus might be expected to be associated with SHMI.

Healthcare complaints are unsolicited reports from patients on, among a range of issues,

unsafe care (Reader, Gillespie, & Roberts, 2014). Three different analyses are suggested for

leveraging the safety data in complaints. First, the number of complaints per admissions unit received by hospitals (hereinafter *complaint frequency*) is a possible safety indicator, as high complaint frequency may indicate safety problems (Taylor, Wolfe, & Cameron, 2004; The NHS Information Centre, 2014). Second, drawing on data science methods, unsafe care may be revealed through automated sentiment analyses of the complaint text (hereinafter *complaint sentiment*), with negative phrasing reflecting dissatisfaction due to unsafe care (Alemi, Torii, Clementz, & Aron, 2012; Greaves et al., 2014). Finally, complaints are shown to report clinical problems (hereinafter *complaint clinical severity*) pertaining to unsafe behaviors (error and neglect) observed during treatments (e.g., misdiagnoses, medication errors, ignoring hygiene rules); the average severity of these (e.g., minor mistakes in pain relief versus administering wrong cancer drugs) is a suggested indicator of patient safety (Gillespie & Reader, 2018).

To summarize, we examine the validity of data pertaining to unsafe healthcare delivery contained within patient experience surveys and healthcare complaints, and ask:

RQ1: Are patient-generated data on unsafe hospital care, captured through patient evaluations of care and healthcare complaints, associated with hospital-level mortality?

We theorize that the value of patient-generated data on patient safety in hospitals lies in its potential to provide additional information to employee-generated data. We investigate this for two variables. First, staff evaluations of the standard of care provided in their hospital (hereinafter *staff evaluations of care*). This is measured through the NHS staff survey (which is distributed annually to a million employees), and is used as a measure of patient safety in hospitals (Powell, Dawson, Topakas, Durose, & Fewtrell, 2014). Second, employee-generated incident reports captured by the NHS National Reporting and Learning System. These data are used to monitor adverse events and near-misses in the NHS (Howell et al., 2015), with the number of incidents

reported, especially for severe events, being theorized to reveal the population of safety events and thus indicate the safety of hospital care (Stavropoulou, Doherty, & Tosey, 2015).

In sum, to examine the added value of patient evaluations of care and healthcare complaints in relation to staff evaluations of care and incident reports, we ask:

RQ2: Do patient-generated data on unsafe hospital care, captured through patient evaluations of care and healthcare complaints, explain additional variance in hospital-level mortality beyond that of staff evaluations of care and incident reports?

Methods

Healthcare complaints were the primary data. We requested 50 irreversibly anonymized typed complaints (the first 25 received after 04/01/2013 and 10/01/2013) from 137 independently managed acute hospitals ("trusts"). The dates counteracted seasonal effects. Fifty-nine trusts provided 2,017 machine-readable complaints ($M = 34.19$, $SD = 12.25$, range: 20–63), containing 2,571,198 words ($M = 1,274.76$), which represented 14% of complaints received. The complaints received varied by trust due to variable resources for redaction (Appendix 1). All complaints were entered into NVivo11. Secondary data on patient experience, complaint frequency, staff surveys, incident-reports, SHMI, and provider spells were collected. The study was reviewed and approved by the institutional review board (ethics committee) at the London School of Economics.

Measurements

We used the following variables to test the research questions. Further description and information on them can be found in Table 1, Appendix 1, and Supplementary File 1.

INSERT TABLE 1 HERE

Patient evaluations of care. We used the 2013 NHS in-patient survey to measure patient care evaluations. Adhering to its dimensional structure (Dawson, 2018), we analyzed responses to

25 items (0–10 scale). There were 23,287 responses (Hospital $M = 394.69$). Principal component analysis (PCA) indicated a single-factor solution (Cronbach's alpha = 0.92). ICC(1) was significant ($p < 0.001$) and low, with a small effect size (0.014, 95% CI [0.009-0.022]). ICC(2) was significant ($p < 0.001$) and high, with a large effect size (0.785, 95% CI [0.709-0.853]).

Complaint frequency. This represents the number of complaints received by hospital per 1,000 provider spells (admissions to hospital: see control variable). Data were from NHS digital.

Complaint sentiment. The average sentiment for each complaint was computed in R with a widely used sentiment dictionary (Nielsen, 2011) consisting of 3,382 terms scored for sentiment from -5 (negative: e.g., "catastrophic") to +5 (positive: e.g., "thrilled"). For every hospital, we scored all words and used the mean sentiment of words in each complaint to calculate sentiment.

Complaint clinical severity. This was specified through the Healthcare Complaints Analysis Tool (HCAT), a psychometrically reliable and theoretically informed coding framework grounded in the typology and severity of problems reported in complaints (Gillespie & Reader, 2016). HCAT has been widely used to investigate complaints on unsafe care (e.g., Mack et al., 2017; Trbovich & Vincent, 2019; Wallace et al., 2018). Complaints are codified by the clinical problems (error and neglect) or non-clinical problems reported (management and relationships issues). Two trained MSc psychology graduates coded all problems reported in all complaints, with interrater reliability tested by a random sample of 101 letters (5%). Coders identified the presence of 13 clinical problems (e.g., misdiagnosis, poor hygiene) or non-clinical problems (e.g., car parking). Multiple problems could be coded per complaint. Clinical problems were graded for severity: low (1), medium (2), or high (3). See Table 3 in the results for the full list of clinical problem types. Non-clinical problems were coded as having "0" clinical severity. Clinical severity was calculated for each hospital using the mean severity score for all problems reported.

Staff evaluations of care. Drawing on the NHS staff survey (NHS England, 2013), we analyzed responses ($n = 49,302$, Hospital $M = 853.63$) to four items measuring staff evaluations of their hospital. PCA indicated a single-factor solution (Cronbach's alpha = 0.88). ICC(1) was significant ($p < 0.001$) and low, with a small effect size (0.053, 95% CI [0.038-0.078]). ICC(2) was significant ($p < 0.001$) and high, with a large effect size (0.939, 95% CI [0.916-0.959]).

Incident reports. Using the NHS National Reporting and Learning Service, we calculated "all safety incidents" (e.g., medication error) and "severe safety incidents" (e.g., deaths) reported by staff per 1,000 bed days for each trust. In total, there were 242,859 ($Mdn = 3,673$) safety incidents and 1,171 ($Mdn = 15$) severe safety incidents reported.

Summary Hospital-level Mortality Indicator (SHMI). This is the published ratio of observed patient deaths to expected deaths (i.e., controlling for underlying risks) at each trust.

Provider spells. To control for the size of each trust (i.e., patients treated), we controlled for "provider spells," which refers to the number of continuous patient stays using a hospital bed.

Analysis

We generated descriptive data and determined normality for all the study variables, and calculated intraclass correlations to test the reliability of the coding for complaint problem severity. We used a two-step model to test the study's research questions. First, to explore the associations among all study variables and examine the relationship between patient evaluations of care, healthcare complaints, and SHMI (RQ1), we used Spearman's rank correlations (because some variables were not normally distributed). Second, to test whether patient-generated data on unsafe care explained variance in SHMI beyond staff-generated data (RQ2), we used a stepwise multiple regression with provider spells as a control variable. We first added the staff-generated measures to the model, followed by the patient-generated measures. Because the analysis aimed to

predict SHMI (a hospital-level outcome), the secondary data were at hospital level (i.e., safety

incidents), and our hypothesis referred to hospital-level complaint profiles (frequency, sentiment,

clinical severity), linear regression was deemed more appropriate than multilevel model analysis.

Results

Table 2 reports the descriptive data for the study variables. The average hospital score was

7.18 for patient evaluations of care. On average, 3.96 complaints were submitted per 1,000 spells.

The average sentiment of the text was −0.32 (range: −0.66 to 0.09). Shapiro–Wilk tests identified

non-normal distributions for provider spells, all safety incidents, and severe safety incidents.

INSERT TABLE 2 HERE

In total, 1,715 clinical problems and 2,658 non-clinical problems were reported in the

complaints. Qualitative examples (by severity) are reported in Table 3 (see Supplementary File 2

for all clinical problems). Interrater reliability analysis was performed by two coders on 101

complaints, with an average ICC of 0.766 (95% CI [0.67–0.836], $F(100, 95.3) = 7.73, p < 0.001$).

On average, there were 2.16 problems contained in each complaint ($n = 2,017$), and 535 reported

high-severity clinical problems. The proportion of problems relating to high-severity clinical

issues ranged from 0% to 27% in the hospitals sampled, with a mean score of 0.84 (range: 0–3).

INSERT TABLE 3 HERE

RQ1: Are patient-generated data on unsafe hospital care, captured through patient evaluations of

care and healthcare complaints, associated with hospital-level mortality?

We used Spearman's rank correlation to investigate the first research question (Figure 2).

A nonparametric test was required because four of the study variables were not normally

distributed. No associations with SHMI were found for patient spells ($r_s = −0.087, p = 0.513$),

patient evaluations of care ($r_s = −0.01, p = 0.938$), complaint frequency ($r_s = 0.077, p = 0.563$), or

complaint sentiment ($r_s = -0.123$, $p = 0.353$). Clinical severity was significantly associated with higher SHMI rates ($r_s = 0.268$, $p = 0.04$). Qualified support was found for RQ1, with the clinical severity of problems reported in healthcare complaints being associated with hospital mortality.

INSERT FIGURE 2 HERE

RQ2: Do patient-generated data on unsafe hospital care, captured through patient evaluations of care and healthcare complaints, explain additional variance in hospital-level mortality beyond that of staff evaluations of care and incident reports?

This question was investigated through a stepwise multiple regression (Table 4) with SHMI as the dependent variable and provider spells as a control (model 1). Model 2 tested the staff-generated data (staff evaluations of care, all safety incidents, and severe safety incidents). Model 3 tested the patient-generated data (patient evaluations of care, complaint frequency, complaint sentiment, and clinical severity). Table 4 reports the three regressions.

INSERT TABLE 4 HERE

Model 1 found that a linear regression model with just provider spells was not significant ($r^2 < 0.001$, $F(1, 57) = 1.019$, $p = 0.317$). Model 2 found that adding staff evaluations of care and staff-reported incident data to Model 1 created a non-significant regression equation ($r^2 = 0.031$, $F(4, 54) = 1.462$, $p = 0.227$), although staff evaluations of care were negatively associated with SHMI ($p < 0.05$). Model 2 was not a significant improvement over Model 1 (r^2 change $= 0.03$, $p = 0.20$). Model 3 found that adding patient-generated data (patient evaluations of care, complaint frequency, complaint sentiment, and complaint clinical severity) to Model 2 led to a significant regression ($r^2 = 0.178$, $F(8, 50) = 2.566$, $p = 0.020$) and improvement over Model 1 (r^2 change $= 0.147$, $p = 0.015$). Higher complaint clinical severity ($p < 0.01$) and poor staff evaluations of care ($p < 0.01$) were associated with higher SHMI. Model 3 was broadly consistent with the

assumptions for multiple regressions; although there was weak evidence of a non-linear

relationship, there was no evidence of heteroscedasticity, the residuals were normally distributed,

and there were no cases outside of Cook's distance. Therefore, there was support for RQ2, with

the severity of clinical problems reported in healthcare complaints explaining the variance in

SHMI beyond staff-generated data.

Discussion

Our results indicate that patients – as stakeholders in safety – possess valid and

supplementary information for monitoring and explaining unsafe care in hospitals. However, the

validity of insight they provide depends on the data source and the analytical frame being used.

Although patient evaluations of care through surveys have been suggested as a potential

source of data on the safety of hospital care (Flott et al., 2017), poor experiences were not

associated with lower SHMI. This may be because patient evaluations of care provide insight on

average experiences pertaining to a range of care quality issues. Crucially, because the questions

in the patient survey do not explicitly capture unsafe behaviors (e.g., medical errors) they may not

capture the safety events that predict excess death rates. To better enable patients to engage as

stakeholders in safety, patient survey items focusing on unsafe events should be developed.

Furthermore, despite being used to indicate the safety of hospital care (The NHS

Information Centre, 2014), fewer complaints were not associated with lower SHMI. This may

reflect the lack of safety-specific information provided through analyzing complaint frequency,

with fewer complaints potentially reflecting (a) patient normalization of unsafe care, (b) beliefs

that complaints will not lead to change, (c) institutional barriers to complaining (i.e., discouraging

people), or (d) defensive processes for recategorizing low-level complaints as informal issues.

Complaint sentiment was also not associated with SHMI. Although sentiment is reliable to measure and is widely used (Alemi et al., 2012), it may be overly blunt for detecting safety issues. Words indicating clinical problems (e.g., death) may be lost amidst words and negative sentiment on non-safety issues. Machine learning trained against human-coded or safety outcome data may be a more promising route to automation (Bleaney, Kuzyk, Man, Mayanloo, & Tizhoosh, 2018).

Complaint clinical severity was associated with SHMI, indicating that useful insights into hospital safety can be generated when patient complaints are analyzed to yield specific data on unsafe behavior. The association with SHMI can be explained in two ways. First, and most simply, patient reports on clinical problems within healthcare complaints capture an accurate sample of the unsafe clinical behaviors that lead to patient harm and thus are associated with excess death rates. Second, and more subtly, healthcare complaints reveal the safety culture of a hospital. Many complaints contain information on severe safety problems; patients report on these through complaints as a last resort to obtain an institutional response. The fact that patients need to write a complaint about severe clinical issues indicates that unsafe clinical events are not being satisfactorily captured, resolved, or learned from. Accordingly, severe clinical complaints may reveal a poor safety culture, which is assumed to explain patient safety (Vogus et al., 2010).

In terms of employee-generated data, staff evaluations of care were associated with SHMI, whereas incident reports were not. This is constant with research showing safety climate to predict patient safety incidents (Singer et al., 2009), and reflects the observation that incident reports may reveal institutional norms for reporting and safety climate rather than accident rates (Probst, 2015; Probst & Estrada, 2010). The finding that patient reports of behavior in healthcare complaints explain variance in SHMI beyond surveys of staff evaluations of care indicates the value of combining data from patients and staff to evaluate the safety of hospital care.

Theoretical and Practical Implications

Demonstrating that patients can provide valid and supplementary information on unsafe hospital care through healthcare complaints has significance for patient safety research. Consistent with stakeholder theory (Freeman, 1984), it indicates that patients are stakeholders in safety within hospitals. Due to their experiences of unsafe events and independence from cultural factors that shape reporting, patients provide valid and detailed information on unsafe behaviors that may be unseen or unreported by clinical staff. In particular, the gravity of problems raised by patients in healthcare complaints (e.g., incorrect surgery) demands that complaints are treated with respect, and used to understand why patient harm occurs and ensure organizational learning.

Furthermore, drawing from the idea that research in different settings can lead to a more holistic understanding of safety in organizations (Hofmann et al., 2017), our findings have wider implications. While the behaviors for ensuring patient safety and workplace safety are distinct, demonstrating that patients can observe employee safety behaviors raises the question of whether stakeholders in other domains can provide similar insight; for example, in settings where stakeholders and employees are intertwined, such as public transport (e.g., where bus driver and passenger safety relies upon safe driving), policing (e.g., where officer and public safety hinge upon the safe use of force), or building maintenance (e.g., where compliance with health and safety rules impacts customers and staff). In light of the long-standing observation that incident reporting is a weak yet potentially powerful signal of safety problems (Macrae, 2009), and that safety climate and the enactment (rather than presence) of policies for capturing safety events determine incident reporting (Probst & Estrada, 2010; Zohar, 2010), stakeholders may represent a new and valuable channel for capturing information on both employee and stakeholder safety. Thus, for settings where stakeholders experience or observe unsafe behaviors, a more holistic

approach to investigating safety in organizations may be to integrate employee behaviors

pertaining to different safety outcomes, and include stakeholder data in assessments of safety.

Theoretically, concepts such as safety climate may be developed to provide an over-arching

explanation of the safety-related reports and outcomes for both employees and stakeholders.

By analyzing the unstructured textual data from patients to investigate employee safety

behaviors, we have also demonstrated the potential of using large-scale textual data to study safety

and supplement established measurements. This approach reflects the growing field of research

using unstructured text (e.g., customer/patient reviews) to predict healthcare audits (Griffiths &

Leaver, 2018), patient re-admissions (Glover et al., 2015), and faulty products (Bleaney et al.,

2018; Pierce et al., 2017). Advances in digitization and data accessibility mean that the range of

domains where textual data and stakeholder perspectives can be used to investigate safety will

likely increase, for instance, in analyzing complaint or social media data to identify reports of

unsafe behavior (e.g., in public services) or gathering insight into safety culture by investigating

employee online reviews of their company (e.g., references to safety in high-risk industries).

However, to harness these data, our findings indicate that specific (e.g., for detecting unsafe

behaviors) rather than generic (e.g., sentiment, satisfaction) measures should be fashioned.

Limitations

Even though we did not control for population attributes or trust specialties, our outcome

variable, SHMI, was weighted to take these aspects into account. SHMI, however, has been

critiqued (Manaseki-Holland et al., 2019) over accuracy concerns (e.g., on record keeping,

obscuration by expected deaths). Our reliance on secondary data has limitations. Due to

resourcing constraints, the number of complaints redacted and received by each hospital varied,

and there may have been variation in how hospitals gathered data (e.g., incident reports). We did

not construct the patient or staff survey items. While these both measured evaluations of the effectiveness of treatment delivery within a hospital, for which safety is integral, they did not directly map onto key constructs within the safety literature (e.g., safety climate).

Our analysis of complaint clinical severity was based on human coding, and although we found good coding reliability, there is scope for improvement. Future research should investigate using manually coded complaints to train a supervised machine learning model to detect and grade complaints pointing to clinical problems. Furthermore, although patients reported on the causes of clinical problems (e.g., mistakes, training), staff perspectives (e.g., on resourcing) are required to fully explain why unsafe care occurred. Finally, the sample of patient experiences that end up in formal complaints is both a strength and a limitation of complaint data. Only a subset of patients make complaints (Gillespie & Reader, 2018), and although these capture unsafe events, they cannot be used to generalize patient experience (patient surveys are more suited for this).

Conclusion

By conceptualizing patients as independent and informed stakeholders in safety, and investigating their reports of unsafe clinical behavior, we explained variance in hospital-level mortality rates beyond that of staff evaluations of care and incident reports. This supports the idea that, in addition to employee-generated data, patient-generated data may be used to monitor and understand unsafe care in hospitals. More broadly, it indicates that non-employee stakeholders may provide an alternative and independent source of data on safety in contexts where they observe and/or experience unsafe employee behaviors. Accordingly, we propose that a more holistic analysis of safety in organizations will be achieved through integrating behaviors pertaining to different safety outcomes, explaining them using safety climate, and studying them through both employee and stakeholder generated-data.

References

Agnew, C., Flin, R., & Mearns, K. (2013). Patient safety climate and worker safety behaviours in acute hospitals in Scotland. *Journal of Safety Research, 45*, 95–101.

Alemi, F., Torii, M., Clementz, L., & Aron, D. C. (2012). Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. *Quality Management in Healthcare, 21*(1), 9–19.

Armitage, G., Moore, S., Reynolds, C., Laloë, P.-A., Coulson, C., McEachan, R., . . . O'Hara, J. (2018). Patient-reported safety incidents as a new source of patient safety data: an exploratory comparative study in an acute hospital in England. *Journal of Health Services Research & Policy, 23*, 36–43.

Beierle, T. C. (2002). The quality of stakeholder-based decisions. *Risk Analysis, 22*(4), 739–749.

Beus, J. M., McCord, M. A., & Zohar, D. (2016). Workplace safety: A review and research synthesis. *Organizational Psychology Review, 6*(4), 352–381.

Bleaney, G., Kuzyk, M., Man, J., Mayanloo, H., & Tizhoosh, H. R. (2018). *Auto-detection of safety issues in baby products*. Paper presented at the Recent Trends and Future Technology in Applied Intelligence, Cham. http://arxiv.org/abs/1805.09772

Bottle, A., Jarman, B., & Aylin, P. (2011). Strengths and weaknesses of hospital standardised mortality ratios. *BMJ, 342*, c7116.

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology, 94*(5), 1103–1127.

Clarke, S. (2013). Safety leadership: A meta-analytic review of transformational and transactional

leadership styles as antecedents of safety behaviours. *Journal of Occupational and*

Organizational Psychology, 86(1), 22–49.

Davis, R., Sevdalis, N., Neale, G., Massey, R., & Vincent, C. A. (2013). Hospital patients' reports

of medical errors and undesirable events in their health care. *Journal of Evaluation in*

Clinical Practice, 19(5), 875–881.

Dawson, J. (2018). *Links between NHS staff experience and patient satisfaction: Analysis of*

surveys from 2014 and 2015. Retrieved from NHS England:

https://www.england.nhs.uk/wp-content/uploads/2018/06/01-018-edc03-staff-inpatient-

survey-report.pdf

de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A.

(2008). The incidence and nature of in-hospital adverse events: A systematic review.

Quality and Safety in Health Care, 17, 216–223.

Dixon-Woods, M., Suokas, A., Pitchforth, E., & Tarrant, C. (2009). An ethnographic study of

classifying and accounting for risk at the sharp end of medical wards. *Social Science and*

Medicine, 69, 362–369.

Doyle, C., Lennox, L., & Bell, D. (2013). A systematic review of evidence on the links between

patient experience and clinical safety and effectiveness. *BMJ Open, 3*, e001570.

Fang, E., Palmatier, R. W., & Evans, K. R. (2008). Influence of customer participation on creating

and sharing of new product value. *Journal of the Academy of Marketing Science, 36*(3),

322–336.

Flin, R. (2007). Measuring safety culture in healthcare: A case for accurate diagnosis. *Safety*

Science, 45(6), 653–667.

Flott, K. M., Graham, C., Darzi, A., & Mayer, E. (2017). Can we use patient-reported feedback to drive change? The challenges of using patient-reported feedback and how they might be addressed. *BMJ Quality & Safety, 26*(6), 502–507.

Francis, R. (2013). *Report of the mid-Staffordshire NHS Foundation Trust Public Inquiry*. Retrieved from Stationery Office: https://webarchive.nationalarchives.gov.uk/20150407084949/http://www.midstaffspublicinquiry.com/sites/default/files/report/Executive%20summary.pdf

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Cambridge: Cambridge University Press.

Gillespie, A., & Reader, T. (2016). The Healthcare Complaints Analysis Tool: Development and reliability testing of a method for service monitoring and organisational learning. *BMJ Quality & Safety, 25*, 937–946.

Gillespie, A., & Reader, T. (2018). Patient-centered insights: Using healthcare complaints to reveal hotspots and blindspots in quality and safety. *The Milbank Quarterly, 96*(3), 530–567.

Glover, M., Khalilzadeh, O., Choy, G., Prabhakar, A. M., Pandharipande, P. V., & Gazelle, G. S. (2015). Hospital evaluations by social media: A comparative analysis of Facebook ratings among performance outliers. *Journal of General Internal Medicine, 30*(10), 1440–1446.

Greaves, F., Laverty, A. A., Cano, D. R., Moilanen, K., Pulman, S., Darzi, A., & Millett, C. (2014). Tweets about hospital quality: A mixed methods study. *BMJ Quality & Safety, 23*(10), 838–846.

Griffin, M. A., & Hu, X. (2013). How leaders differentially motivate safety compliance and safety participation: The role of monitoring, inspiring, and learning. *Safety Science, 60*, 196–202.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*(3), 347–358.

Griffiths, A., & Leaver, M. P. (2018). Wisdom of patients: Predicting the quality of care using aggregated patient feedback. *BMJ Quality & Safety, 27*(2), 110–118.

Halligan, M., & Zecevic, A. (2011). Safety culture in healthcare: A review of concepts, dimensions, measures and progress. *BMJ Quality & Safety, 20*(4), 338–343.

Harrison, R., Walton, M., Manias, E., Smith–Merry, J., Kelly, P., Iedema, R., & Robinson, L. (2015). The missing evidence: A systematic review of patients' experiences of adverse events in health care. *International Journal for Quality in Health Care, 27*(6), 424–442.

Hofmann, D. A., Burke, M. J., & Zohar, D. (2017). 100 years of occupational safety research: From basic protections and work analysis to a multilevel view of workplace safety and risk. *Journal of Applied Psychology, 102*(3), 375–388.

Hofmann, D. A., & Mark, B. (2006). An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology, 59*(4), 847–869.

Hopkins, A. (2009). Thinking about process safety indicators. *Safety Science, 47*(4), 460–465.

Howell, A.-M., Burns, E. M., Bouras, G., Donaldson, L. J., Athanasiou, T., & Darzi, A. (2015). Can patient safety incident reports be used to compare hospital safety? Results from a quantitative analysis of the english national reporting and learning system data. *PLoS One, 10*(12), e0144107.

Katz-Navon, T., Naveh, E., & Stern, Z. (2005). Safety climate in health care organizations: A multidimensional approach. *Academy of Management Journal, 48*(6), 1075–1089.

Katz-Navon, T., Naveh, E., & Stern, Z. (2009). Active learning: When is more better? The case of

   resident physicians' medical errors. *Journal of Applied Psychology, 94*(5), 1200.

Leroy, H., Dierynck, B., Anseel, F., Simons, T., Halbesleben, J. R., McCaughey, D., . . . Sels, L.

   (2012). Behavioral integrity for safety, priority of safety, psychological safety, and patient

   safety: A team-level study. *Journal of Applied Psychology, 97*(6), 1273–1281.

Levtzion-Korach, O., Frankel, A., Alcalai, H., Keohane, C., Orav, J., Graydon-Baker, E., . . .

   Tomov, E. I. (2010). Integrating incident data from five reporting systems to assess patient

   safety: making sense of the elephant. *The Joint Commission Journal on Quality and

   Patient Safety, 36*(9), 402–410.

Mack, J. W., Jacobson, J., Frank, D., Cronin, A. M., Horvath, K., Allen, V., . . . Schrag, D. (2017).

   Evaluation of patient and family outpatient complaints as a strategy to prioritize efforts to

   improve cancer care delivery. *The Joint Commission Journal on Quality and Patient

   Safety, 43*(10), 498–507.

Macrae, C. (2009). Making risks visible: Identifying and interpreting threats to airline flight

   safety. *Journal of Occupational and Organizational Psychology, 82*(2), 273–293.

Manaseki-Holland, S., Lilford, R., Chen, Y., Gupta, K., Chilton, P., & Hofer, T. (2019). Ranking

   hospitals based on preventable hospital death rates: A systematic review with implications

   for both direct measurement and indirect measurement through standardized mortality

   rates. *The Milbank Quarterly, 97*(1), 228–284.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be

   distinguished from contextual performance. *Journal of Applied Psychology, 79*(4), 475–

   480.

NHS England. (2013). *Making sence of your staff survey data 2013*. Retrieved from:

    https://www.nhsstaffsurveys.com/Caches/Files/ST13.%20MAKING%20SENSE%20OF%

    20YOUR%20STAFF%20SURVEY%20DATA_FINAL.pdf

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in

    microblogs. *arXiv preprint arXiv:1103.2903*.

Ocloo, J., & Matthews, R. (2016). From tokenism to empowerment: Progressing patient and

    public involvement in healthcare improvement. *BMJ Quality & Safety, 25*(8), 626–632.

O'Hara, J. K., Reynolds, C., Moore, S., Armitage, G., Sheard, L., Marsh, C., . . . Lawton, R.

    (2018). What can patients tell us about the quality and safety of hospital care? Findings

    from a UK multicentre survey study. *BMJ Quality & Safety, 27*(9), 673–682.

Olsen, E. (2010). Exploring the possibility of a common structural model measuring associations

    between safety climate factors and safety behaviour in health care and the petroleum

    sectors. *Accident Analysis & Prevention, 42*(5), 1507–1516.

Papanicolas, I., & Figueroa, J. F. (2019). Preventable harm: Getting the measure right. *BMJ,

    366*(8208), I4185.

Pfeiffer, Y., Manser, T., & Wehner, T. (2010). Conceptualising barriers to incident reporting: A

    psychological framework. *Quality and Safety in Health Care, 19*(6), 1–10.

Pierce, C. E., Bouri, K., Pamer, C., Proestel, S., Rodriguez, H. W., Van Le, H., . . . Edwards, I. R.

    (2017). Evaluation of Facebook and Twitter monitoring to detect safety signals for medical

    products: An analysis of recent FDA safety alerts. *Drug Safety, 40*(4), 317–331.

Powell, M., Dawson, J., Topakas, A., Durose, J., & Fewtrell, C. (2014). *Staff satisfaction and

    organisational performance: Evidence from a longitudinal secondary analysis of the NHS*

staff survey and outcome data. Retrieved from National Institute for Heath Research:

https://www.ncbi.nlm.nih.gov/books/NBK263759/pdf/Bookshelf_NBK263759.pdf

Probst, T. M. (2015). Organizational safety climate and supervisor safety enforcement: Multilevel

explorations of the causes of accident underreporting. *Journal of Applied Psychology,*

100(6), 1899–1907.

Probst, T. M., Brubaker, T. L., & Barsotti, A. (2008). Organizational injury rate underreporting:

The moderating effect of organizational safety climate. *Journal of Applied Psychology,*

93(5), 1147–1154.

Probst, T. M., & Estrada, A. X. (2010). Accident under-reporting among employees: Testing the

moderating influence of psychological safety climate and supervisor enforcement of safety

practices. *Accident Analysis & Prevention, 42*(5), 1438–1444.

Raleigh, V. S., Frosini, F., Sizmur, S., & Graham, C. (2012). Do some trusts deliver a consistently

better experience for patients? An analysis of patient experience across acute care surveys

in English NHS trusts. *BMJ Quality & Safety, 21*(5), 381–390.

Rathert, C., Brandt, J., & Williams, E. S. (2012). Putting the 'patient'in patient safety: a

qualitative study of consumer experiences. *Health Expectations, 15*(3), 327–336.

Reader, T., & Gillespie, A. (2013). Patient neglect in healthcare institutions: A systematic review

and conceptual model. *BMC Health Services Research, 13*(1), 156.

Reader, T., Gillespie, A., & Roberts, J. (2014). Patient complaints in healthcare systems: A

systematic review and coding taxonomy. *BMJ Quality & Safety, 23*(8), 678–689.

Reason, J. (2000). Safety paradoxes and safety culture. *Injury Control and Safety Promotion, 7*(1),

3–14.

Sahlström, M., Partanen, P., & Turunen, H. (2018). Patient-reported experiences of patient safety incidents need to be utilized more systematically in promoting safe care. *International Journal for Quality in Health Care, 30*(10), 778–785.

Shojania, K. G., & Thomas, E. J. (2013). Trends in adverse events over time: Why are we not improving? *BMJ Quality & Safety, 22*, 273–277.

Singer, S., Lin, S., Falwell, A., Gaba, D., & Baker, L. (2009). Relationship of safety climate and safety performance in hospitals. *Health Services Research, 44*(2p1), 399–421.

Smith, S. R., & Ingram, H. (2002). Policy tools and democracy. In L. Salamon (Ed.), *The tools of government: A guide to the new governance* (pp. 565–584). New York: Oxford University Press.

Stavropoulou, C., Doherty, C., & Tosey, P. (2015). How effective are incident-reporting systems for improving patient safety? A systematic literature review. *The Milbank Quarterly, 93*(4), 826–866.

Taylor, D. M., Wolfe, R. S., & Cameron, P. A. (2004). Analysis of complaints lodged by patients attending Victorian hospitals, 1997-2001. *Medical Journal of Australia, 181*, 31–35.

Taylor, J., Dominici, F., Agnew, J., Gerwin, D., Morlock, L., & Miller, M. R. (2012). Do nurse and patient injuries share common antecedents? An analysis of associations with safety climate and working conditions. *BMJ Quality & Safety, 21*(2), 101–111.

The King's Fund. (2017, 23/03/2017). A mixed bag of results from the NHS Staff Survey. Retrieved from: https://www.kingsfund.org.uk/blog/2017/03/mixed-bag-staff-survey

The NHS Information Centre. (2014). *Data on written complaints in the NHS 2013-14*. Retrieved from: https://files.digital.nhs.uk/publicationimport/pub14xxx/pub14705/data-writ-comp-nhs-2013-2014-rep.pdf

Trbovich, P., & Vincent, C. (2019). From incident reporting to the analysis of the patient journey. *BMJ Quality & Safety, 28*(3), 169–171.

Vincent, C. (2011). *Patient safety*. London: John Wiley & Sons.

Vogus, T. J., Sutcliffe, K. M., & Weick, K. E. (2010). Doing no harm: Enabling, enacting, and elaborating a culture of safety in health care. *The Academy of Management Perspectives, 24*(4), 60–77.

Wachter, R. (2010). Patient safety at ten: Unmistakable progress, troubling gaps. *Health Affairs, 29*, 163–173.

Wakefield, J. G., McLaws, M.-L., Whitby, M., & Patton, L. (2010). Patient safety culture: Factors that influence clinician involvement in patient safety behaviours. *Quality and Safety in Health Care, 19*(6), 585–591.

Wallace, E., Cronin, S., Murphy, N., Cheraghi-Sohi, S., MacSweeney, K., Bates, M., & Fahey, T. (2018). Characterising patient complaints in out-of-hours general practice: A retrospective cohort study in Ireland. *British Journal of General Practice, 68*(677), e860–e868.

Walton, M. M., Harrison, R., Kelly, P., Smith-Merry, J., Manias, E., Jorm, C., & Iedema, R. (2017). Patients' reports of adverse events: A data linkage study of Australian adults aged 45 years and over. *BMJ Quality & Safety, 26*(9), 743–750.

Waring, J. J. (2005). Beyond blame: Cultural barriers to medical incident reporting. *Social Science & Medicine, 60*(9), 1927–1935.

Weaver, S. J., Lubomksi, L. H., Wilson, R. F., Pfoh, E. R., Martinez, K. A., & Dy, S. M. (2013). Promoting a culture of safety as a patient safety strategy: A systematic review. *Annals of Internal Medicine, 158*(5), 369–374.

Xia, N., Griffin, M. A., Wang, X., Liu, X., & Wang, D. (2018). Is there agreement between

worker self and supervisor assessment of worker safety performance? An examination in

the construction industry. *Journal of Safety Research, 65*, 29–37.

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions.

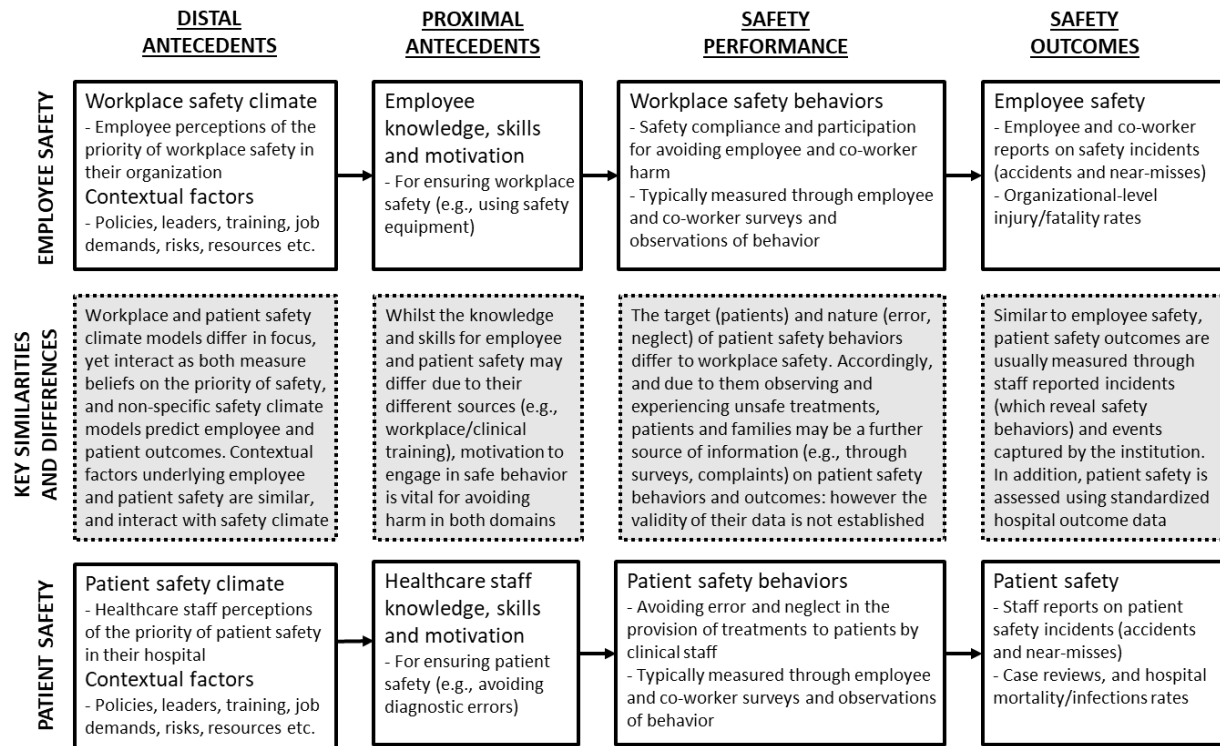*Accident Analysis & Prevention, 42*(5), 1517–1522.

	DISTAL ANTECEDENTS	PROXIMAL ANTECEDENTS	SAFETY PERFORMANCE	SAFETY OUTCOMES
EMPLOYEE SAFETY	Workplace safety climate - Employee perceptions of the priority of workplace safety in their organization **Contextual factors** - Policies, leaders, training, job demands, risks, resources etc.	Employee knowledge, skills and motivation - For ensuring workplace safety (e.g., using safety equipment)	Workplace safety behaviors - Safety compliance and participation for avoiding employee and co-worker harm - Typically measured through employee and co-worker surveys and observations of behavior	Employee safety - Employee and co-worker reports on safety incidents (accidents and near-misses) - Organizational-level injury/fatality rates
KEY SIMILARITIES AND DIFFERENCES	Workplace and patient safety climate models differ in focus, yet interact as both measure beliefs on the priority of safety, and non-specific safety climate models predict employee and patient outcomes. Contextual factors underlying employee and patient safety are similar, and interact with safety climate	Whilst the knowledge and skills for employee and patient safety may differ due to their different sources (e.g., workplace/clinical training), motivation to engage in safe behavior is vital for avoiding harm in both domains	The target (patients) and nature (error, neglect) of patient safety behaviors differ to workplace safety. Accordingly, and due to them observing and experiencing unsafe treatments, patients and families may be a further source of information (e.g., through surveys, complaints) on patient safety behaviors and outcomes: however the validity of their data is not established	Similar to employee safety, patient safety outcomes are usually measured through staff reported incidents (which reveal safety behaviors) and events captured by the institution. In addition, patient safety is assessed using standardized hospital outcome data
PATIENT SAFETY	Patient safety climate - Healthcare staff perceptions of the priority of patient safety in their hospital **Contextual factors** - Policies, leaders, training, job demands, risks, resources etc.	Healthcare staff knowledge, skills and motivation - For ensuring patient safety (e.g., avoiding diagnostic errors)	Patient safety behaviors - Avoiding error and neglect in the provision of treatments to patients by clinical staff - Typically measured through employee and co-worker surveys and observations of behavior	Patient safety - Staff reports on patient safety incidents (accidents and near-misses) - Case reviews, and hospital mortality/infections rates

Figure 1. Model summarizing key similarities and differences between the employee safety and patient safety literatures

Table 1. *Study Variables*

Variable*:	Measure	Indicative of:
Patient evaluations of care	Psychometrically validated annual survey, running since 2002, with 25 items measuring experiences of care for patients staying at least one night in hospital (Dawson, 2018). Items focused on: response times; staffing levels; waiting times; decision-making; help to eat; confidence in clinical staff. Scores were averaged for each hospital.	Lapses in the quality of treatment delivery by clinical staff that potentially lead to unsafe outcomes. The survey does not measure safety directly (e.g., errors) and relates to patient-specific experiences (i.e., not the hospital as a whole).
Complaint frequency	The number of written complaints per 1,000 admissions (patients who stay overnight) was calculated for each hospital trust.	High number of patients reporting problems with clinical safety. Can be shaped by other factors (e.g., barriers to complaining) and other issues (e.g., billing), and gives no insight on safety behavior.
Complaint sentiment	The average negativity of complaint text sent to each hospital trust.	Dissatisfaction due to unsafe care. Sentiment may also reflect upset with non-clinical issues, and provides no behavioral data.
Complaint clinical severity	The average severity of clinical problems (graded on a 1–3 scale, with non-clinical problems being graded 0) reported in the complaints sent to each hospital trust. Clinical problems consisted of perceived errors or neglectful care: see Table 3 for examples.	Unsafe behaviors in providing clinical care. Patients describe medical errors (e.g., for treatments, diagnoses) and neglectful care (e.g., for examinations, personal care) that lead to harm. Problems may not be accurately reported or may be misattributed (e.g., clinician medication error caused by equipment failure).
All safety incidents	The frequency of staff-reported safety incidents (potential or actual instances of patient harm) reported by staff per 1,000 bed days for	Staff-reported events that endanger patient safety. Reporting trends influenced by safety climate (which encourages reporting and

Variable*:	Measure	Indicative of:
	each hospital trust. Incidents focused on problems such as diagnostic, procedural, equipment, or medication errors.	learning), meaning that incident data provide a mixed indicator of safety (few reports can reveal poor safety climate or few incidents).
Severe safety incidents	The frequency of staff-reported severe safety incidents (e.g., wrong-site surgery) reported by staff per 1,000 bed days for each hospital trust.	Staff-reported events that led to patient death or disability. Due to the relatively low rates of such events, they are not optimal for revealing widespread problems in clinical delivery and may not be effectively reported in a highly unsafe hospital.
Staff evaluations of care	Annual employee survey, running since 2003 (The King's Fund, 2017), with four items on effectiveness of care provided to patients (e.g., hospital prioritization of care; satisfaction with care standards). The average overall score for each hospital trust was used.	Organizational conditions for providing safe care. Does not reveal clinical practices that endanger safety. However, it is an indicator because it shows staff to be concerned about the state of clinical care, which underpins safety.
SHMI	Taking account of patient background data (e.g., primary diagnosis, complicating factors), SHMI is the ratio of observed deaths at a hospital divided by the expected number (based on a calculation of the likelihood of patients dying from their admitting condition) (Campbell, Jacques, Fotheringham, Maheswaran, & Nicholl, 2012).	Potentially avoidable patient deaths that may have occurred due to problems (e.g., medical error, neglect) in the delivery of care.

* Please see Appendix 1 and Supplementary File 1 for statistical and descriptive information on the study variables.

Table 2. *Descriptive Data for All Study Variables*

Variable	N	Mean	St. Dev.	Min	Max	Shapiro–Wilk
Provider spells	59	63876.41	31077.98	17434	163970	0.913*
Patient evaluations of care	59	7.18	0.20	6.76	7.54	0.974
Complaint frequency (per 1,000 spells)	59	3.96	1.17	1.50	7.92	0.968
Sentiment (-5 to 5, negative to positive)	59	-0.32	0.15	-0.66	0.09	0.991
Complaint clinical severity (0 to 3, none to high)	59	0.84	0.18	0.43	1.20	0.982
Staff evaluations of care	59	3.67	0.21	3.06	4.16	0.990
All safety incidents (per 1,000 bed days)	59	7.97	2.45	3.50	17.10	0.937*
Severe safety incidents (per 1,000 bed days)	59	0.04	0.04	0.00	0.18	0.749*
Summary hospital mortality indicator (SHMI)	59	1.00	0.09	0.76	1.18	0.961

Note: * < .001, indicating a non-normal distribution

Table 3. *Distribution of Complaint Clinical Problems (n = 1,715), by Severity, with Examples*

Clinical problem	N	Low severity	N	Medium severity	N	High severity	N
Examination: neglect in examining and monitoring patients	335	Patient with high temperature not monitored for a day	39	Patient with fractures not given any X-rays at Accident and Emergency	180	Patient referred for cervical cancer examined without swabs, samples, or smear test	116
General care: neglect in providing clinical treatments	209	Hour delay in replacing catheter after it had been incorrectly put in	40	Doctor refused to administer more pain relief because he was too busy	117	Heavily bleeding patient in labor left in corridor for 4–5 hours	52
Diagnosis: errors in making diagnoses	192	Fractured wrist misdiagnosed as muscle damage; revised a week later	9	Severe pneumonia and sepsis misdiagnosed as slight (nonurgent) infection	96	Tumor in kidney misdiagnosed as cyst before donation of the other kidney	87
Treatment: errors in providing clinical treatments	140	Patient needing to take urine test twice; urine sample misplaced	20	Patient with contagious condition left in room with other (young) patients for 7.5 hours	69	Equipment left in patient during spinal surgery; not identified in post-surgery MRI scans	51
Training: mistakes by unqualified staff	138	Health-care assistant unable to find vein for blood sample	22	Nurse did not have experience for providing surgical aftercare	75	Locum doctor operated instead of a gallbladder surgeon	41
Medication: errors in administering medicines	134	Short delay in administration of antibiotics	24	Patient given steroid injections instead of anesthetic injections	72	Did not stop administering Epilim to young patient, resulting in pancreatitis and diabetes	38

| Category		Example 1		Example 2		Example 3	
Outcomes: procedures leading to adverse effects	124	Blood test causing a large hematoma	10	Knee operation resulting in nerve damage	63	Repair of aortic graft resulting in catastrophic infections	51
Personal care: neglecting patient hygiene	90	Dirty pajamas not changed	21	Patient in poor hygienic state (e.g., no cleaning, shaving)	48	Patient lying in own urine and bed sores for 10 days	21
Care plans: failing to develop care plans	86	Assistance increase not considered in patient discharge plan	10	Patient discharged without advice, home assessment, or care plan	60	Alzheimer's patient discharged to empty home	16
Responsiveness; ignoring patient calls for help	82	Nurse unable to come quickly when patient started vomiting and called for help	7	Staff not responsive to alarm bells: patient left in shower after fall	41	Patient left at A&E to wait while suffering from multiple seizures	34
Handling: treating patients roughly	81	Breathing mask applied too tightly	27	Rough hoisting of patient who just had stomach surgery	43	Bone marrow transplant patient required anesthetics for avoidable handling blisters	11
Teamwork: mistakes in coordination	50	Confusion between nurses on dressing patient's wound	16	Patient discharged by consultant when another established opposite	27	Failure to share stroke diagnosis between staff members	7
Nourishment: not providing food/water	54	Patient given cold meal	20	Patient not given any drinks for a full day	25	Patient did not receive any food for four days	9
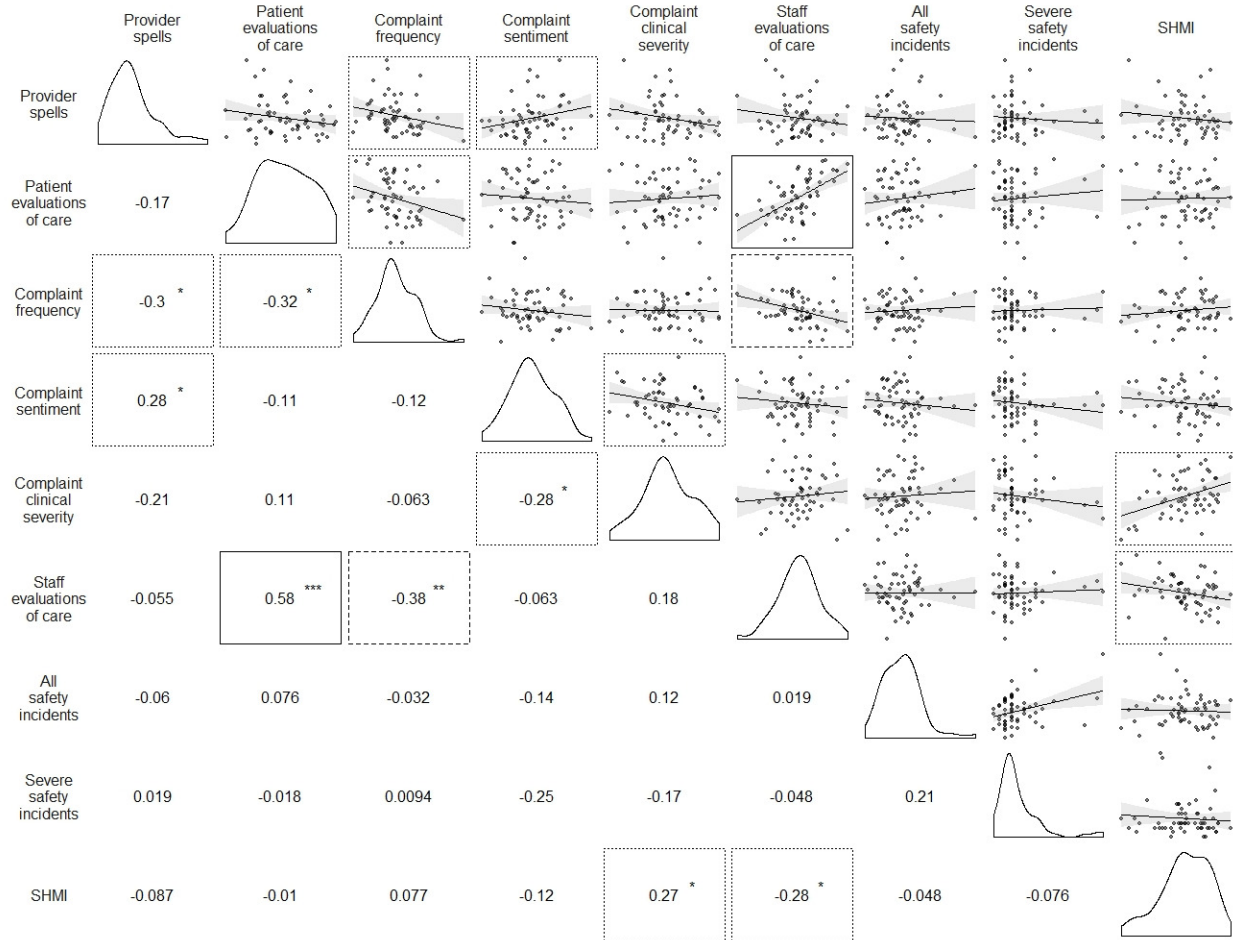
Figure 2. Spearman's rank correlations, scatterplots, and density. The lower-left cells are pairwise Spearman's rank correlations (*p < 0.05, **p < 0.01, ***p < 0.001). The upper-right cells are pairwise scatterplots with linear models overlaid. Boxes facilitate identifying the scatterplot that corresponds to statistically significant correlations: dotted (*p* < 0.05), dashed (*p* < 0.01), and solid (*p* < 0.001). The diagonal displays the distribution of each variable with density plots.

Table 4. *Multiple Regression for the Summary Hospital-level Mortality Indicator (SHMI)*

	Dependent variable:		
	SHMI		
	(1)	(2)	(3)
Provider spells	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Staff evaluations of care		-0.120*	-0.189**
		(0.058)	(0.067)
All safety incidents		-0.001	-0.005
		(0.005)	(0.005)
Severe safety incidents		-0.181	0.013
		(0.353)	(0.336)
Patient evaluations of care			0.122
			(0.072)
Complaint frequency			0.005
			(0.011)
Complaint sentiment			-0.024
			(0.079)
Complaint clinical severity			0.208**
			(0.067)
Constant	1.028***	1.496***	0.658
	(0.028)	(0.224)	(0.473)
Observations	59	59	59
R^2	0.018	0.098	0.291
Adjusted R^2	0.0003	0.031	0.178
Residual Std. Error	0.093 (df = 57)	0.091 (df = 54)	0.084 (df = 50)
F Statistic	1.019 (df = 1; 57)	1.462 (df = 4; 54)	2.566* (df = 8; 50)
Note:			*p < 0.05; **p < 0.01; ***p < 0.001

Appendix 1: Sampling and Survey Analyses

Sampling

Healthcare complaints. The study sample was determined through a pilot study of healthcare complaints submitted to NHS Hospitals in 2011–2012. The pilot data contained 720 complaints from 23 acute NHS hospital trusts, which showed a correlation between clinical severity and SHMI ($r = 0.42$, $n = 23$, $p = 0.061$). Based on these data, measuring clinical severity ($SD = 0.71$) from a trust with an average number of complaints ($n = 245$) with a 0.3% margin of error (95% confidence) would require 20 complaints per trust. A power calculation for a positive correlation of 0.4 (power = 0.8, $p = 0.05$) indicated a sample size of 37 trusts. For the 2013–2014 sample, and to ensure we overshot the minimum threshold for the power calculation, we used a Freedom of Information (FOI) request to collect 50 irreversibly anonymized complaints from 137 hospital trusts (the first 25 typed complaints immediately received after April 01, 2013 and the first 25 typed complaints received after October 01, 2013). For the fifty-nine hospitals that responded, there was variance in the final number of complaints received (M = 34.19, SD = 12.25), for example due to some hospitals having limited resources for redaction. Consistency in sampling was ensured through, regardless of numbers, hospitals redacting and sampling healthcare complaints received immediately after the two targets dates. Eight hospitals were removed prior to the analysis because they did not submit the minimum of 20 complaints required to estimate mean clinical severity (CI 0.6, 95% confidence).

Secondary data. The secondary data were independently collected by several NHS services. In each case, we sought to obtain data nearest to 01/07/13 because this was the midway point between the two timepoints for sampling complaints (i.e., 01/04/13 and 01/10/13). The 59 trusts included in the study were not significantly different from nonparticipating trusts ($n = 81$)

for the following variables: hospital provider spells ($Mdn_{included}$ = 58,568, $Mdn_{excluded}$ = 51,607, W = 2,651, p = 0.271), the inpatient survey ($Mdn_{included}$ = 7.941, $Mdn_{excluded}$ = 7.924, W = 2,549, p = 0.502), complaint frequency ($Mdn_{included}$ = 218, $Mdn_{excluded}$ = 209, W = 2,543.5, p = 0.517), the staff survey ($Mdn_{included}$ = 3.47, $Mdn_{excluded}$ = 3.733, W = 2,477, p = 0.62), safety incidents ($Mdn_{included}$ = 8.1, $Mdn_{excluded}$ = 7.2, W = 2,710, p = 0.177), severe safety incidents ($Mdn_{included}$ = 0.03, $Mdn_{excluded}$ = 0.03, W = 2,194.5, p = 0.407), and hospital mortality ($Mdn_{included}$ = 1.003, $Mdn_{excluded}$ = 1.014, W = 2,421, p = 0.896).

Survey Analyses

Patient evaluations of care. We conducted a principal component analysis for the 25 items underlying the survey scale specified by Dawson et al. (2018) and calculated Cronbach's alpha. The Kaiser-Meyer-Olkin measure established the sampling adequacy as excellent (KMO = 0.96, min = 0.87). A one-factor solution was viable (parallel analysis = 3, optimal coordinates = 3, acceleration factor = 1, supported by visual inspection of the scree plot, which indicated a very sharp hinge at the second eigenvalue) and explained 39.73% of the variance. Cronbach's alpha was 0.92 (95% CI [0.92–0.93]). ICC(1) was low and had a small effect size (0.014, 95% CI [0.009-0.022]), indicating isolated patient survey scores provided reduced information for distinguishing hospitals. ICC(2) was high and had a large effect size (0.785, 95% CI [0.709-0.853]), indicating that when aggregated the scores measured reliable differences between hospitals (F(58,23218) = 6.573, p < 0.001).

Staff evaluations of care. We conducted a principal component analysis for the four items underlying the survey scale specified by The National NHS Staff Survey Co-ordination Centre and calculated Cronbach's alpha. The Kaiser-Meyer-Olkin measure established the

sampling adequacy as very good (KMO = 0.80, min = 0.79). A one-factor solution was indicated

(parallel analysis = 1, optimal coordinates = 1, acceleration factor = 1) and explained 78.48% of

the variance. Cronbach's alpha was 0.88 (95% CI [0.87–0.88]). ICC(1) was low and had a small

effect size (0.053, 95% CI [0.038-0.078]), indicating that isolated staff survey scores provided

reduced information for distinguishing hospitals. ICC(2) was high and had a large effect size

(0.939, 95% CI [0.916-0.959]), indicating that when aggregated the scores measured reliable

differences between hospitals ($F(58, 49243) = 73.069$, $p < 0.001$).