BMJ Open

Identifying positively deviant elderly medical wards using routinely collected NHS Safety Thermometer data: an observational study

Ruth Baxter,[1,2] Natalie Taylor,[3] Ian Kellar,[2] Victoria Pye,[4] Mohammed A Mohammed,[1,5] Rebecca Lawton[1,2]

[1]Yorkshire Quality and Safety Research Group, Bradford Institute for Health Research, Bradford, UK
[2]School of Psychology, University of Leeds, Leeds, UK
[3]Cancer Research Division, Cancer Council, Sydney, New South Wales, Australia
[4]Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia
[5]Faculty of Health Studies, University of Bradford, Bradford, UK

Correspondence to
Dr Ruth Baxter;
ruth.baxter@bthft.nhs.uk

ABSTRACT

Objective The positive deviance approach seeks to identify and learn from exceptional performers. Although a framework exists to apply positive deviance within healthcare organisations, there is limited guidance to support its implementation. The approach has also rarely explored exceptional performance on broad outcomes, been implemented at ward level, or applied within the UK. This study develops and critically appraises a pragmatic method for identifying positively deviant wards using a routinely collected, broad measure of patient safety.

Design A two-phased observational study was conducted. During phase 1, cross-sectional and temporal analyses of Safety Thermometer data were conducted to identify a discrete group of positively deviant wards that consistently demonstrated exceptional levels of safety. A group of matched comparison wards with above average performances were also identified. During phase 2, multidisciplinary staff and patients on the positively deviant and comparison wards completed surveys to explore whether their perceptions of safety supported the identification of positively deviant wards.

Setting 34 elderly medical wards within a northern region of England, UK.

Participants Multidisciplinary staff (n=161) and patients (n=188) clustered within nine positively deviant and comparison wards.

Results Phase 1: A combination of analyses identified five positively deviant wards that performed best in the region, outperformed their organisation and performed consistently well over 12 months. Five above average matched comparator wards were also identified. Phase 2: Staff and patient perceptions of safety generally supported the identification of positively deviant wards using Safety Thermometer data, although patient perceptions of safety were less concordant with the routinely collected data.

Conclusions This study tentatively supports a pragmatic method of using routinely collected data to identify positively deviant elderly medical wards; however, it also highlights the various challenges that are faced when conducting the first stage of the positive deviance approach.

Trial registration number UK Clinical Research Network Portfolio (reference-18050).

Strengths and limitations of this study

► There is limited guidance to support the identification of positive deviants in healthcare settings. This study develops a method for identifying positive deviants using routinely collected data.

► A combination of four different analyses (including performance rankings, comparisons with organisational level performances and Statistical Process Control methods) were conducted to provide a pragmatic yet robust method for identifying a discrete group of positively deviant wards that performed exceptionally well on a broad outcome of safety.

► Staff and patient perceptions of safety were measured using validated surveys to explore whether they supported the identification of positively deviant wards using routinely collected data.

► Due to the small sample size (n=9 wards) it was not possible to statistically assess whether staff and patient perceptions of safety supported the identification of positive deviants using routinely collected Safety Thermometer data.

► The study was conducted on elderly medical wards and so further research is required to explore whether the methods can be generalised to identify positive deviants in other healthcare settings.

BACKGROUND

Despite extensive efforts to improve, patient safety continues to be a pervasive problem across the globe.[1 2] Traditionally, these efforts have focused on past errors and harm, but there are increasing calls to also explore how '*safe*' patient care is delivered.[3 4] Positive deviance provides an asset-based approach to improving the quality and safety of healthcare.[5 6] The approach seeks to identify and learn from those who demonstrate exceptional performance on an outcome of interest.[6] It assumes that solutions to problems already exist within communities and
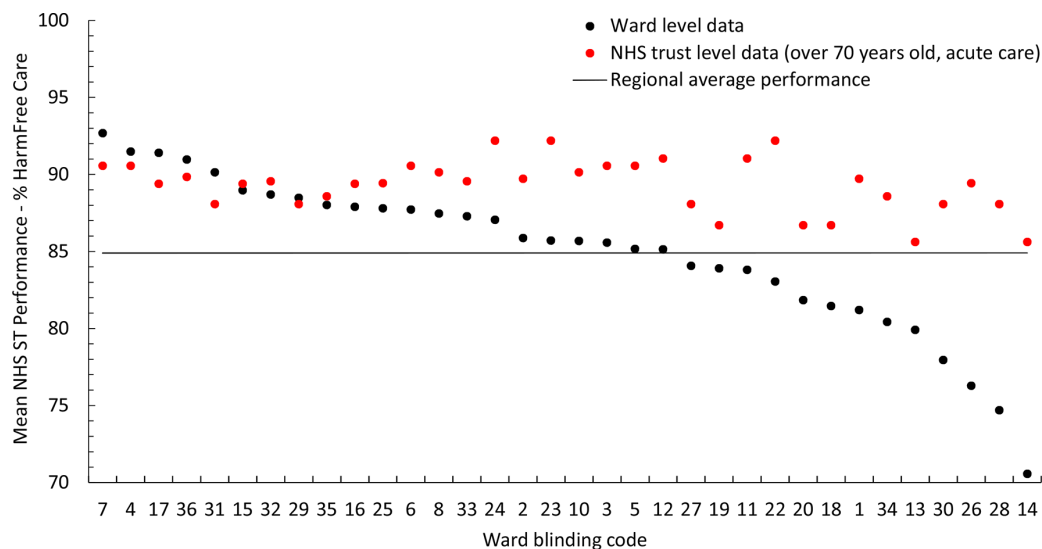
BMJ

Figure 1 Scatterplot comparing average ST harm-free care performances at ward and NHS Trust (organisation) levels. NHS, National Health Service; ST, Safety Thermometer.

that positive deviants (individuals, teams or organisations) identify these solutions and succeed despite facing similar constraints as others.

Bradley *et al*[7] have proposed a four-stage framework to implement the positive deviance approach within healthcare organisations. Positive deviants, who display exceptionally high performance, are identified using routinely collected data (stage 1). Hypotheses about how they succeed are generated using qualitative methods (stage 2). These hypotheses are quantitatively tested in larger, more representative samples (stage 3), and then disseminated to others with the help of key stakeholders (stage 4). Despite the increasing popularity of the positive deviance approach,[8] there is little evidence or practical guidance to support its application within healthcare organisations.[9 10]

Identifying positive deviants who demonstrate exceptional performance

Identifying positive deviants may be the most crucial stage of Bradley *et al's*[7] framework as subsequent stages hinge on its perceived legitimacy. Misidentification could lead to the generation of hypotheses that do not capture the factors that facilitate exceptional performance. Bradley *et al*[7] suggest that positive deviants should be identified by ranking routinely collected data, and previous healthcare applications have, for example, identified three of the top 10 clinics with the best anticoagulation control[11] or the top quintile of primary care medical homes with the most improved diabetes outcomes.[12] However, performance rankings can differ depending on the rating systems that are used, creating confusion and contradiction about who demonstrates high and low performances.[13–15] Positive deviants are also supposed to demonstrate *exceptional* rather than just *good* performances on the outcome of interest,[7] but rankings or league tables simply appraise performances along a continuum without

differentiating a distinct group of 'outliers' or positive deviants from the rest of a population.

Previous healthcare applications of the approach have predominantly identified positively deviant organisations (eg, hospitals) or individuals.[9] This is despite greater amounts of variation existing at the level of a hospital ward or unit[16–18] and the majority of frontline care being delivered by the multidisciplinary teams that work within these clinical microsystems.[19] Previous applications have also typically focused on specific processes or outcomes of care such as hand hygiene compliance and the incidence of healthcare-associated infections.[9] Although it is relevant to explore positive deviance in this way, factors that contribute to safety often operate across various levels of the system and affect multiple outcomes.[20 21] If ward teams succeed on broad outcomes of care there are likely to be some underlying, latent factors that facilitate their success. Understanding these factors and spreading the associated strategies may generate more far-reaching improvements in quality and safety. However, it remains unknown whether positively deviant wards or units can be identified accurately using a routinely collected, broad and multidimensional measure of patient safety.

The shortcomings of using routinely collected data, such as publication lags, coding differences and data gaming, are well documented.[22–24] Nonetheless, if positive deviance is to become a useful improvement approach, its methods must be pragmatic and accessible for healthcare organisations, networks and frontline improvers to use. Routinely collected and publicly available data are therefore required, especially when applying the approach across several different wards, units or organisations. In preparation for this study, various routinely collected measures of safety within the UK's National Health Service (NHS) were identified (eg, Hospital Episode Statistics[25] and the NHS Staff Survey[26]), but the NHS Safety Thermometer (ST) was identified as
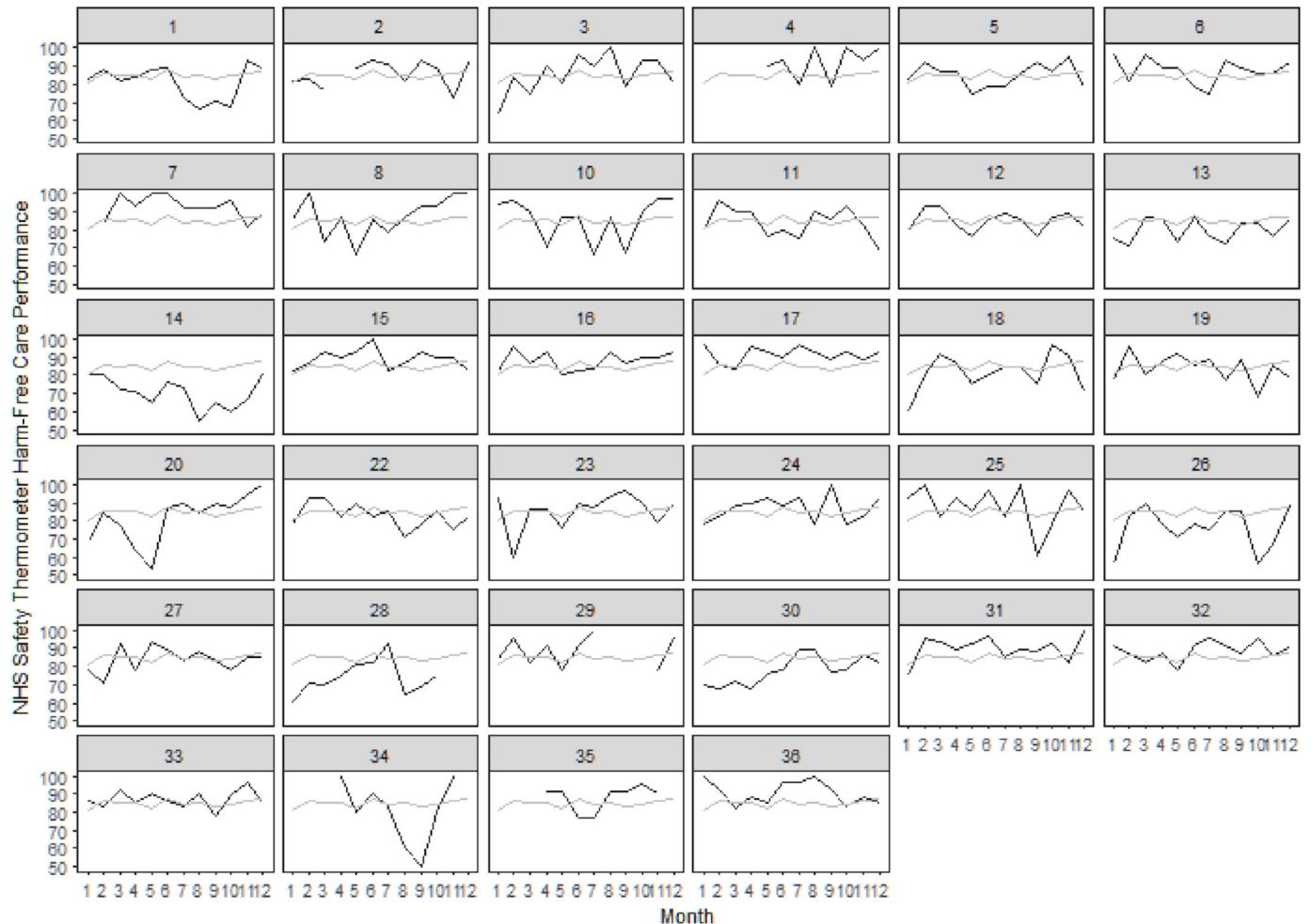
Figure 2 Run charts comparing ward and regional level monthly ST harm-free care performance across a 12-month period. Each square represents an individual ward within the population. Wards are numbered consecutively according to their pseudonym (from top left to bottom right across the rows). Black lines represent a ward's monthly harm-free care performance. Grey lines represent the region's average monthly performance. NHS, National Health Service; ST, Safety Thermometer .

the only routinely collected, broad measure of safety that is publicly available at ward level. Each month, the NHS ST is used to measure four commonly occurring patient harms—falls, pressure ulcers, venous thromboembolism and urinary tract infections. These data are used to report on the proportion of 'harm-free care' that is delivered at ward, specialty and organisational levels.[27] During this study, NHS ST data will be used to identify positively deviant ward teams that deliver exceptional levels of safe patient care.

One of the key challenges to improving the quality and safety of healthcare is convincing people to adopt a chosen solution.[28] The positive deviance approach seeks to identify solutions from within with the assumption that these solutions will be acceptable to others, feasible to implement and sustainable over time.[5 6] However, if staff and/or patients do not perceive positive deviants to be performing exceptionally well, they may not engage with the positively deviant strategies that are disseminated during stage 4 of the Bradley *et al*[7] framework. As this study is one of the first applications of the approach to identify positive deviants at ward level using a broad

measure of safety, this study also sought to assess the extent to which staff and patient perceptions of safety supported the identification of positively deviant wards via ST data. A number of validated surveys exist to measure staff perceptions of 'safety culture'—the shared values, beliefs, norms and attitudes that guide how healthcare staff behave in order to maintain safety.[29] Furthermore, there is increasing evidence to suggest that patients can be involved in maintaining their own safety—patients are able to identify adverse events and can provide a unique perspective on the safety of care.[30–32]

Study design and aims

Based on these gaps in the literature, the overarching aim of this observational study was to develop and critically appraise a robust yet pragmatic method for identifying positive deviants at ward level using a routinely collected, broad outcome of safety. The study was conducted in two phases. During phase 1, we sought to apply a rigorous and robust analysis (compared with simply ranking data) to the ST's harm-free care data to identify a distinct group of positively deviant wards that demonstrated exceptional
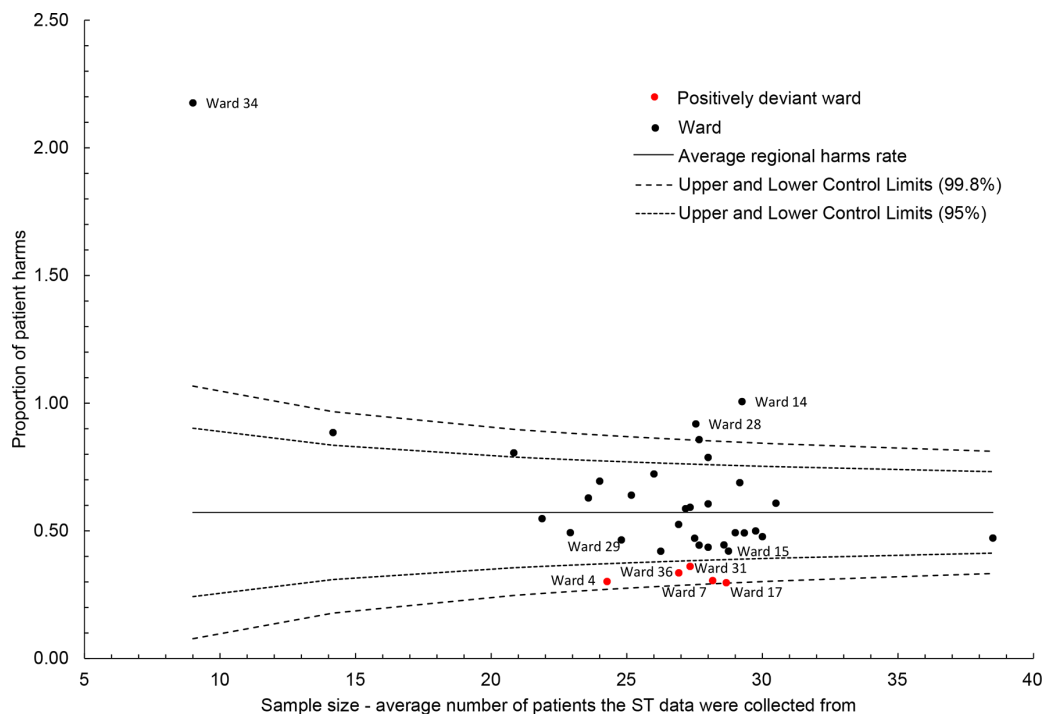
Figure 3 Funnel plot of average ST harm-free care performance and average sample size. ST, Safety Thermometer.

levels of safety over a 12-month period. To try and identify positively deviant ward teams that were delivering exceptionally safe patient care under challenging circumstances, elderly medical wards were sampled because older patients are particularly vulnerable to safety incidents and harms such as falls and pressure ulcers.[33 34] Phase 2 of this study aimed to explore whether staff and patient perceptions of safety (which were measured using surveys) were similar to, or at odds with, the routinely collected NHS ST data that had been used to identify the positively deviant wards.

Although this particular study focuses solely on stage 1 of the Bradley *et al*[7] framework, it also contributes to a wider application of the positive deviance approach[35] which seeks to generate hypotheses about *how* the positively deviant ward teams deliver exceptionally safe patient care. The qualitative findings from this wider application (stage 2 of the framework) will be published separately. To the authors' knowledge, this study is one of the first applications of positive deviance within the English NHS.

PHASE 1 METHODS

This paper adheres to the Strengthening the Reporting of Observational Studies in Epidemiology statement.[36] A complete checklist can be found in online supplementary file 1.

Sample

Elderly medical wards (n=36) were identified from within 13 acute NHS Trusts (healthcare organisations) in a northern region of England, UK. This represented 10% of all acute NHS Trusts in England. Elderly medical wards were defined as those providing 24-hour, acute,

medical care for elderly patients (>65 years); with dedicated multidisciplinary teams; and patient stays typically exceeding 48 hours. Specialty wards (eg, stroke, rehabilitation and assessment units) were excluded to maximise homogeneity within the sample.

Data extraction

For all wards ST data were extracted from a publicly accessible website[37] for a 12-month period—August 2013 to July 2014. Data were extracted for the ST's 'harm-free care' measure and all of the individual ST harms at two different levels: ward level for all patients; and Trust level for acute patients over 70 years.[35] Double blinding during both study phases ensured that researchers (RB, RL, NT and IK), staff and patients were not aware of how wards compared with each other on their ST performances.

Analysis

Pragmatic cross-sectional and temporal analyses were conducted to identify a distinct group of positively deviant elderly medical wards that displayed exceptionally high performance on the ST harm-free care measure. Initially, in line with guidance[7] and previous applications,[9] wards were ranked on their average harm-free care performances and then three further analyses were conducted. First, to ensure that positive deviants' success was not simply a function of organisational performance, a scatterplot compared wards with their respective NHS Trust level data. Second, as small sample sizes increase the likelihood of variability being attributable to chance,[38] a funnel plot compared ward level performances against their average sample sizes. Third, performances were assessed over the 12-month period using run charts.[39] Monthly ward level data were compared with the monthly

Table 1 Ward level descriptive statistics for all staff and patient survey measures

	ST harm-free care (Phase 1) Mean %	Patient Measure of Safety Mean (SD)*	Friends and Family Test Mean (SD)*	CQUIN Mean (SD)†	Patient Safety Grade Mean (SD)*
Positively deviant wards					
Ward 1	90.14	4.33 (0.45)	4.71 (0.56)	2.48 (0.45)	4.29 (0.56)
Ward 3	92.68	4.21 (0.34)	4.55 (0.67)	2.58 (0.47)	4.21 (0.70)
Ward 6	91.48	3.94 (0.37)	4.14 (1.15)	2.45 (0.32)	4.09 (0.54)
Ward 10	90.97	4.52 (026)	4.65 (0.49)	2.53 (0.48)	4.13 (0.78)
Comparison wards					
Ward 2	88.48	4.11 (0.53)	4.26 (0.75)	2.25 (0.50)	3.50 (1.15)
Ward 4	87.72	4.09 (0.39)	4.26 (1.00)	2.43 (0.45)	4.07 (0.48)
Ward 5	85.17	3.96 (0.39)	4.15 (1.23)	2.18 (0.58)	4.05 (0.52)
Ward 8	87.90	4.51 (0.27)	4.75 (0.44)	2.48 (0.33)	3.69 (0.79)
Ward 9	88.01	4.30 (0.36)	4.46 (0.88)	2.50 (036)	3.29 (1.16)
Average					
PD group	91.33	4.24 (0.41)	4.51 (0.78)	2.51 (0.43)	4.18 (0.67)
Comparison group	87.46	4.20 (0.43)	4.38 (0.92)	2.38 (0.46)	3.71 (0.91)

*Measured on a 0–5 Likert scale.
†Measured on a 0–3 Likert scale. Higher scores represent safer perceptions of patient care on all measures.
CQUIN, Commissioning for Quality and Innovation; PD, positive deviant; ST, Safety Thermometer.

regional average, and run charts were visually assessed to identify consistent outperformers. Statistical Process Control (SPC) methods such as run charts and funnel plots are increasingly being promoted to assess variation within healthcare.[40 41] Performance variations exist within any stable system and SPC methods can be used to distinguish between variation that occurs by chance (ie, noise in the system) and variation that has an assignable cause (ie, a signal of positive deviance).[42]

Identification of wards

Analyses were compared to identify a distinct group of positively deviant elderly medical wards. Individual harms data were also assessed to ensure that positive deviants performed well across all measures in the ST's harm-free care composite. In preparation for phase 2 of this study, comparison wards with slightly above average ST performances were identified. As is the case for many sources of publicly available data, it was not possible to conduct any case mix adjustments on ST data. Consequently, comparator wards were matched to positive deviants on three key variables to increase homogeneity within the sample: patient gender—mixed, female or male; NHS Trust type—teaching and/or foundation trusts; and a routinely collected measure of deprivation.[43] To ensure that positive deviants did not simply care for younger, and thus comparatively more healthy patients,[33 34] administrative average patient age data were analysed post hoc.

PHASE 1 RESULTS

Data were analysed for 34 elderly medical wards clustered within 13 NHS Trusts. Two wards with over 50% missing data were excluded. Average harm-free care performances ranged from 70.56% to 92.68%. (Online supplementary file 2 presents all ward rankings.) Wards 7, 4, 17, 31, 36 and 29 were the only wards to outperform their respective Trusts on the ST harm-free care data (figure 1). The first five of these also ranked the highest in the sample while ward 29 ranked eighth. Visual assessment of the run charts (figure 2) indicated that wards 7, 17, 31, 36, 4 and 15 consistently outperformed the regional average over 12 months, with greater certainty held for those wards listed first. Although none of the wards exceeded the funnel plot's three SE control limits, wards 7, 4, 17, 31 and 36 exceeded them at two SEs (figure 3).

Using a combination of these four analyses, five wards (7, 4, 17, 31 and 36) were identified to form a distinct group of positive deviants. These wards demonstrated the best performances (rankings); outperformed their respective NHS Trusts (scatterplot); consistently outperformed over 12 months (run charts); and their performance variation was attributable to more than chance alone (funnel plot). They also performed well—around/above average—for each individual ST harm (online supplementary file 2). Wards 29 and 15, which were identified through the scatterplot and run charts respectively, did not exceed the funnel plot control limits and so were not deemed to be positively deviant.

Table 2 A visual representation of how positively deviant and comparison wards, which were identified using routinely collected ST data, ranked on patient and staff perceptions of safety

Rank	Routinely collected ST harm-free care (phase 1)	Patient and staff perceptions of safety (phase 2)			
		Patient Measure of Safety	Friends and Family Test	CQUIN	Patient Safety Grade
1 (High)	Ward 3	Ward 10	Ward 8	Ward 3	Ward 1
2	Ward 6	Ward 8	Ward 1	Ward 10	Ward 3
3	Ward 10	Ward 1	Ward 10	Ward 9	Ward 10
4	Ward 1	Ward 9	Ward 3	Ward 1 and Ward 8	Ward 6
5	Ward 2	Ward 3	Ward 9		Ward 4
6	Ward 9	Ward 2	Ward 4 and Ward 2	Ward 6	Ward 5
7	Ward 8	Ward 4		Ward 4	Ward 8
8	Ward 4	Ward 5	Ward 5	Ward 2	Ward 2
9 (Low)	Ward 5	Ward 6	Ward 6	Ward 5	Ward 9

Positively deviant wards (as identified by the routinely collected ST data) are shaded in colour. Comparison wards are represented in white. Higher ranks represent safer perceptions of patient care on all measures.
CQUIN, Commissioning for Quality and Innovation; ST, Safety Thermometer.

In total, five matched comparison wards with slightly above average ST harm-free care performances were also identified. Independent samples t-tests indicated that positively deviant wards (M=91.33, SD=0.92) significantly differed from comparators (M=87.46, SD=1.31, t_8=5.42, P=0.001) and all other wards in the region (M=83.85, SD=4.57, t_{32}=3.61, P=0.01) for average ST harm-free care performance. Online supplementary file 3 presents the key characteristics of the positively deviant and comparison wards. Although it was not sufficiently powered due to the small sample size (n=9), a post hoc analysis of administrative data for average patient age suggests that positively deviant wards (M=85.1, SD=2.11) did not care for younger, and thus more healthy patients, than the comparison wards (M=84.92, SD=1.42, t_7=0.15, P=0.88).

PHASE 2 METHODS
Participants and recruitment
The positively deviant and comparison wards identified during phase 1 were invited to participate in phase 2 of the study. One positively deviant ward was unable to take part. Up to 20 patients and a minimum of 50% of the multidisciplinary ward team were recruited opportunistically to complete surveys assessing their perceptions of safety on the ward. Eligible patients were 65 years or older, were deemed to have capacity and were considered to be physically well enough. Staff could hold any job role and be of any professional grade. The patient sample size of 20 was determined by previous research; recruiting beyond 20 participants only minimally narrows the CIs for the main measure in the patient survey.[44 45] A 30%–50% response rate for the staff survey has also previously been reported.[46] All recruitment was conducted between February and August 2015

(due to publication lag of ST data and the time taken to gain ethical approvals and NHS permissions). Double blinding was retained.

Data collection tools
Patient survey
Patients completed the Patient Measure of Safety (PMOS) which gathers feedback from hospitalised patients about the safety of their care and assesses perceptions about factors that contribute to safety.[30 47] A total of 44 items are scored on a 5-point Likert scale ranging from strongly agree to strongly disagree. These items measure nine safety domains: communication and team working; organisation and care planning; access to resources; ward type and layout; information flow; staff roles and responsibilities; staff training; equipment (design and functioning); and delays. A stand-alone item measures dignity and respect. The PMOS has been validated, is reliable and considered acceptable to patients.[30 47] Patients also completed the NHS Friends and Family Test (FFT)—a single-item measure of patient experience used in the UK,[48] and three items that had previously been a part of the NHS Commissioning for Quality and Innovation (CQUIN) payment framework.[49]

Staff survey
Multidisciplinary staff completed the Patient Safety Grade (PSG) which asks them to grade their ward on overall safety using a 5-point Likert scale (excellent to failing). The PSG is one of four outcomes within the validated Hospital Survey on Patient Safety Culture where non-required outcomes can be removed.[46 50 51] This single item was used to maximise response rates.[52] The patient and staff surveys were published with the study protocol.[35]

Procedure

Eligible patients were identified by clinical members of the ward team. Researchers provided written and verbal information about the study and patients gave written informed consent. The patient survey was completed electronically and patients chose whether to do this independently or with the researcher's support. Where support was requested (eg, due to frailty) the researcher simply read the questions and recorded the patient's answers for them.

Paper copies of the staff surveys were distributed by the researcher and ward managers to multidisciplinary staff at convenient times (eg, clinical handovers). Staff placed their completed surveys into a 'drop box' which was stored securely on the ward. Staff participation was incentivised by a prize draw (£20 gift voucher per ward).

Analyses

Blinding was removed. Items within the PMOS and CQUIN measures were aggregated to create an overall PMOS and overall CQUIN score for each individual patient. Individual staff and patient-level data were then aggregated to ward level by calculating an average ward level score for all measures—the PMOS, FFT, CQUIN and PSG. To assess whether staff and patient perceptions of safety supported the identification of positive deviants, wards were ranked and z-scores were compared in a scatterplot. The small sample size (n=9 wards) and dichotomised performance groups prevented statistical analysis.

PHASE 2 RESULTS

Data were collected from 188 patients and 161 multidisciplinary staff, clustered within nine participating elderly medical wards. (Online supplementary file 4 reports all recruitment data.) On average, patients were 84.53 years old (SD=5.45), and staff were predominantly nurses or support workers. Table 1 reports the ward level descriptives for all of the measures. (Minimum and maximum values are presented in online supplementary file 4.) Positively deviant wards performed better than comparators across all four measures, although differences between the groups were small.

Ranked performances (table 2) highlight that staff on positively deviant wards perceived care to be safer than staff on comparison wards as measured by the PSG. This was also largely true for patients, although their perceptions were less concordant with the ST data. Positively deviant ward 6 displayed the lowest PMOS and FFT scores, and two comparators (wards 8 and 9) performed better than some positive deviants on certain patient measures. The scatterplot of z-scores (online supplementary file 5) compared performances across different normal distributions. Positively deviant wards predominantly performed above the mean and, as a group, generally performed better on all measures than the comparison wards supporting their identification using ST data.

DISCUSSION

This study developed and critically appraised a method for conducting stage 1 of the Bradley et al[7] positive deviance framework—identifying positive deviants. More specifically, we applied a rigorous and robust analysis to the ST harm-free care data to identify positively deviant wards that demonstrated sustained exceptional levels of safety, and established the extent to which survey-based staff and patient perceptions of safety aligned with these ST harm-free data. Previous applications of the approach have typically identified positively deviant individuals or organisations who demonstrate exceptional performance on narrow processes or outcomes of care.[9] However, performance variation also exists between wards/units.[18] Although there is merit to focusing on specific aspects of care, this can divert attention away from other important aspects of safety reducing opportunities for wider improvement, for example, by implementing cultural changes that improve several different outcomes (eg, improving multidisciplinary teamwork).

During phase 1 of this study, a robust yet pragmatic analysis successfully identified a discreet group of five statistically different positively deviant elderly medical wards with exceptionally high ST harm-free care performances. Although these wards did rank top of the region, rankings alone did not differentiate between positively deviant wards and those that performed well. This study therefore advanced the previous methods that have been used to identify positive deviants by identifying a distinct and statistically different group of wards that ranked best within the region and outperformed their NHS Trust (organisation), performed consistently over 12 months and demonstrated success beyond what would be expected through chance. SPC methods such as funnel plots and run charts are increasingly promoted for assessing performance variation within healthcare.[40 41] They combine statistical rigour with sensitive measurement to differentiate between variation that is expected by chance and variation that has an assignable cause.[42 53] The methods are also considered to be relatively intuitive and pragmatic enough for use by improvers on the frontline.[41]

However, the extent to which wards truly demonstrated exceptional performance can be questioned. Wards did not exceed the funnel plot's three SE control limits and so their exceptional, outlier status was limited. Furthermore, although positively deviant wards differed statistically from others, the minimal performance differences between them and the 'next best' wards highlight the importance of considering clinical significance—that is, whether differences meaningfully affect patient treatment.[54] Positive deviants are supposed to demonstrate 'exceptionally' high performance,[7] but there is little consensus in the literature about how to differentiate between high performance and positively deviant performances. Control limits are considered to be conservative and can be made more or less stringent depending on the context.[39] In the absence of extreme outliers, it may still

be possible to conduct stage 2 of the Bradley *et al* framework[7] in order to generate useful learning from 'positive deviants' who simply perform *well*. However, regardless of whether one learns from outlying positive deviants or not, it is important to note that there is currently a lack of evidence on the methods that should be used to (A) effectively uncover positively deviant success strategies, and (B) disseminate them to others (stages 2 and 4 of the Bradley framework).[9 10]

During phase 2 of this study, staff and patient perceptions of safety did, in the main, corroborate the routinely collected ST data, providing tentative support for the methods used to identify positively deviant wards. Patients on positively deviant wards, though, did not uniformly perceive their care to be safer than those on comparator wards. This could be explained by wards adopting different approaches to delivering safe patient care; for example, if staff emphasise guideline compliance over patient-centred care this may influence patient perceptions of safety. Furthermore, patients may have unique perspectives of safety, which perhaps encompass the culture of a ward rather than just the outcomes that are reflected in the routinely collected data.[31 55] The lack of agreement between the FFT, CQUIN and ST measures may also have arisen by measuring two associated but distinct quality domains—patient experience versus safety.[56] Fundamentally though, inconsistent patient perceptions highlight that different positively deviant wards may have been identified had a different broad measure of safety (other than the ST) been used.

In addition to these considerations, there are various overarching considerations and challenges that are faced when applying stage 1 of the Bradley *et al* framework[7] to identify positive deviants. First, there are few sources of routinely collected data within the UK's NHS or further afield that broadly measure safety and are publicly available at ward level.[16] This makes it difficult to adopt a pragmatic approach and identify positive deviants across different organisations. Although this study provides tentative support for using ST data within a UK healthcare setting, the harms measured within this tool are particularly pertinent to older people. Consequently, this measure may lack relevance to applications of the approach which seek to identify positive deviants in different healthcare settings, for example, paediatric wards or emergency departments.

Second, positive deviants are assumed to succeed *despite facing the same constraints as others*[6] and so it is critical to identify them from within a homogenous population to ensure that, as far as possible, one is comparing like for like. This study increased homogeneity by defining elderly medical wards as stringently as possible and by sampling matched comparators. However, numerous factors are known to contribute to patient safety incidents,[57] case mix adjustments are notoriously difficult[58] and one can never fully control for all confounding variables when identifying positive deviants. The complexity of healthcare means that all ward teams deliver patient care within their own unique organisational contexts and so it will never be possible to sample a fully homogenous population or identify positive deviants that face *exactly* the same constraints as others. This challenge is likely to be especially pertinent when comparing performance on broad outcomes of safety, when using publicly available data and when adopting a pragmatic approach.

Third, although the problems associated with routinely collected data are well documented,[22–24] there are also wider implications of using routine data to identify positive deviants. Performance variation can arise because measurement is conducted in a social context—staff do not make decisions about the same things nor do they decide things in the same way.[59] This is problematic when the positive deviance approach seeks to compare performances across several different healthcare providers. Furthermore, healthcare organisations retrospectively measure the absence rather than the presence of safety,[60] and measurement and monitoring systems say nothing about how safe patient care currently is or how safe it will be in the future.[61] This compounds the ability to accurately identify positive deviants and thus the ability to reliably conduct subsequent stages of the approach.

Study limitations

Various study limitations have already been highlighted including measuring statistical rather than clinical differences between positively deviant and comparison wards. Due to resource constraints, it was also not possible to assess staff and patient perceptions of safety across all 34 wards that were sampled during phase 1 of the study. The resulting small and dichotomised sample during phase 2 meant that the associations between staff and patient perceptions of safety and the ST data could not be assessed statistically. Furthermore, some of the differences in performance between positively deviant and comparison wards were small. Had the comparison group comprised negative deviants (the worst performers) rather than above average performers, the differences between the two groups on each of the quantitative measures may have been starker. Many previous applications of stage 1 of the positive deviance approach have identified positive and negative deviants.[35] However, this specific comparison group was chosen with the wider application of positive deviance in mind,[35] so that when we qualitatively explore *how* positive deviants succeed (stage 2 of the framework[7]) we can strive to distinguish between *exceptional* and good performances, not just explore how teams differ from the worst in the population.

Finally, as with many routinely collected measures of quality and safety, the reliability and validity of the ST has been questioned. ST data are collected opportunistically at a single monthly time point,[27] harm definitions are subject to interpretation[62 63] and data collection was previously incentivised. However, the ST is used to measure performance in most acute NHS Trusts and it is the only routinely collected, broad measure of safety that is publicly available at ward level in the

UK. Additional ST tools, such as the Medications Safety Thermometer, have been developed[64 65] and so, if these can also be used to identify positively deviant wards, then the methods tested in this study could have greater impact across the NHS.

CONCLUSIONS

This study has shown that a distinct group of positively deviant wards that perform exceptionally well on a routinely collected, broad measure of safety can be identified using a robust yet pragmatic method, and that staff and patient perceptions of safety do, in the main, support their identification. It has highlighted the challenges faced when selecting a source of routinely collected data that provides a valid and reliable measure at the appropriate level in order to facilitate performance comparisons across wards or units in several organisations. Many of these challenges are applicable to a variety of different settings and applications of the approach and so this study may provide generalisable guidance on the methods that can be used to effectively apply stage 1 of the Bradley *et al* framework[7] and identify positive deviants.

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Patient consent Not required.

Ethics approval South East Scotland Research Ethics Committee (ref 14/SS/1085). NHS permissions were gained from all five NHS Trusts involved.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Anonymised data are available on request.

REFERENCES

1. Hogan H, Healey F, Neale G, et al. Preventable deaths due to problems in care in english acute hospitals: a retrospective case record review study. *BMJ Qual Saf* 2012;21:737–45.
2. Landrigan CP, Parry GJ, Bones CB, et al. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;363:2124–34.
3. Hollnagel E, Braithwaite J, Wears RL. *Resilient health care*. UK: Ashgate, 2013.
4. Vincent C, Amalberti R. *Safer healthcare: strategies for the real world*: Springer Open, 2016. http://link.springer.com/book/10.1007%2F978-3-319-25559-0
5. Lawton R, Taylor N, Clay-Williams R, et al. Positive deviance: a different approach to achieving patient safety. *BMJ Qual Saf* 2014;23:880–3.
6. Marsh DR, Schroeder DG, Dearden KA, et al. The power of positive deviance. *BMJ* 2004;329:1177–9.
7. Bradley EH, Curry LA, Ramanadhan S, et al. Research in action: using positive deviance to improve quality of health care. *Implement Sci* 2009;4:25.
8. Dixon-Woods M, Martin GP. Does quality improvement improve quality? *Future Hospital Journal* 2016;3:191–4.
9. Baxter R, Taylor N, Kellar I, et al. What methods are used to apply positive deviance within healthcare organisations? A systematic review. *BMJ Qual Saf* 2016;25:190–201.
10. Rose AJ, McCullough MB. A practical guide to using the positive deviance method in health services research. *Health Serv Res* 2017;52:1207–22.
11. Rose AJ, Petrakis BA, Callahan P, et al. Organizational characteristics of high- and low-performing anticoagulation clinics in the veterans health administration. *Health Serv Res* 2012;47:1541–60.
12. Gabbay RA, Friedberg MW, Miller-Day M, et al. A positive deviance approach to understanding key features to improving diabetes care in the medical home. *Ann Fam Med* 2013;11(Suppl 1):S99–107.
13. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Aff* 2015;34:423–30.
14. Healthcare Association of New York State. *HANYS' report on report cards: understanding publicly reported hospital quality measures*. US, New York: HANYS Quality Institute, 2013.
15. Rothberg MB, Morsi E, Benjamin EM, et al. Choosing the best hospital: the limitations of public quality reporting. *Health Aff* 2008;27:1680–7.
16. Pannick S, Wachter RM, Vincent C, et al. Rethinking medical ward quality. *BMJ* 2016;355:i5417.
17. Rose JS, Thomas CS, Tersigni A, et al. A leadership framework for culture change in health care. *Jt Comm J Qual Patient Saf* 2006;32:433–42.
18. Schwendimann R, Zimmermann N, Küng K, et al. Variation in safety culture dimensions within and between US and Swiss Hospital Units: an exploratory study. *BMJ Qual Saf* 2013;22:32–41.
19. Nelson EC, Batalden PB, Huber TP, et al. Microsystems in health care: part 1. Learning from high-performing front-line clinical units. *Jt Comm J Qual Improv* 2002;28:472–93.
20. Reason J. Understanding adverse events: human factors. *Qual Health Care* 1995;4:80–9.
21. Vincent C, Taylor-Adams S, Stanhope N. Framework for analysing risk and safety in clinical medicine. *BMJ* 1998;316:1154–7.
22. Lilford R, Mohammed MA, Spiegelhalter D, et al. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147–54.
23. Shaw J, Taylor R, Dix K. *Uses and abuses of performance data in healthcare*. UK: Dr Forster, 2015.
24. Zhan C, Miller MR. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;12(Suppl 2):58ii–63.
25. Digital N. Hospital episode statistics 2017. http://content.digital.nhs.uk/hes (cited 13th Oct 2017).
26. The Picker Institute. The NHS staff survey 2017. http://www.nhsstaffsurveys.com/Page/1056/Home/NHS-Staff-Survey-2017/ (cited 13th Oct 2017).
27. Power M, Stewart K, Brotherton A. What is the NHS safety thermometer? *Clin Risk* 2012;18:163–9.
28. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Qual Saf* 2012;21:876–84.
29. Scott T, Mannion R, Davies H, et al. The quantitative measurement of organizational culture in health care: a review of the available instruments. *Health Serv Res* 2003;38:923–45.
30. Giles SJ, Lawton RJ, Din I, et al. Developing a patient measure of safety (PMOS). *BMJ Qual Saf* 2013;22:554–62.
31. Weingart SN, Pagovich O, Sands DZ, et al. What can hospitalized patients tell us about adverse events? Learning from patient-reported incidents. *J Gen Intern Med* 2005;20:830–6.

32. Weissman JS, Schneider EC, Weingart SN, *et al*. Comparing patient-reported hospital adverse events with medical record review: do patients know something that hospitals do not? *Ann Intern Med* 2008;149:100–8.

33. Sari AB, Cracknell A, Sheldon TA. Incidence, preventability and consequences of adverse events in older people: results of a retrospective case-note review. *Age Ageing* 2008;37:265–9.

34. Thomas EJ, Brennan TA. Incidence and types of preventable adverse events in elderly patients: population based review of medical records. *BMJ* 2000;320:741–4.

35. Baxter R, Taylor N, Kellar I, *et al*. Learning from positively deviant wards to improve patient safety: an observational study protocol. *BMJ Open* 2015;5:e009650.

36. von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet* 2007;370:1453–7.

37. Health and Social Care Information Centre. NHS safety thermometer. 2014 http://www.hscic.gov.uk/thermometer (cited 29th Apr 2014).

38. Mayer EK, Bottle A, Rao C, *et al*. Funnel plots and their emerging application in surgery. *Ann Surg* 2009;249:376–83.

39. Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf* 2011;20:46–51.

40. Mountford J, Wakefield D. From stoplight reports to time series: equipping boards and leadership teams to drive better decisions. *BMJ Qual Saf* 2017;26:9–11.

41. Schmidtke KA, Poots AJ, Carpio J, *et al*. Considering chance in quality and safety performance measures: an analysis of performance reports by boards in English NHS trusts. *BMJ Qual Saf* 2017;26:61–9.

42. Mohammed MA. Using statistical process control to improve the quality of health care. *Qual Saf Health Care* 2004;13:243–5.

43. Office of the Deputy Prime Minister. *The english indices of deprivation 2004: summary (revised)*. UK: Office of the Deputy Prime Minister, 2004.

44. Lawton R, O'Hara JK, Sheard L, *et al*. Can staff and patient perspectives on hospital safety predict harm-free care? An analysis of staff and patient survey data and routinely collected outcomes. *BMJ Qual Saf* 2015;24:369–76.

45. Sheard L, O'Hara J, Armitage G, *et al*. Evaluating the PRASE patient safety intervention - a multi-centre, cluster trial with a qualitative process evaluation: study protocol for a randomised controlled trial. *Trials* 2014;15:2282.

46. Sorra J, Nieva V. *Hospital survey on patient safety culture*. USA, Rockville: Westat, AHRQ, 2004. Publication No. 04-0041.

47. McEachan RR, Lawton RJ, O'Hara JK, *et al*. Developing a reliable and valid patient measure of safety in hospitals (PMOS): a validation study. *BMJ Qual Saf* 2014;23:565–73.

48. NHS England. Friends and family test. 2014 http://www.england.nhs.uk/statistics/statistical-work-areas/friends-and-family-test/ (cited 29th Apr 2014).

49. Department of Health. Using the Commissioning for Quality and Innovation (CQUIN) payment framework – Guidance on national goals for 2011/12. 2010 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215049/dh_133859.pdf (cited 3rd Apr 2016).

50. Sarac C, Flin R, Mearns K, *et al*. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. *BMJ Qual Saf* 2011;20:842–8.

51. Sorra JS, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 2010;10:199.

52. Edwards P, Roberts I, Clarke M, *et al*. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;324:1183.

53. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;12:458–64.

54. Oberst MT. Clinical versus statistical significance. *Cancer Nurs* 1982;5:475–6.

55. Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. *BMJ Qual Saf* 2016;25:626–32.

56. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;3:e001570.

57. Lawton R, McEachan RR, Giles SJ, *et al*. Development of an evidence-based framework of factors contributing to patient safety incidents in hospital settings: a systematic review. *BMJ Qual Saf* 2012;21:369–80.

58. Shahian DM, Normand SL. What is a performance outlier? *BMJ Qual Saf* 2015;24:95–9.

59. Dixon-Woods M, Leslie M, Bion J, *et al*. What counts? An ethnographic study of infection data reported to a patient safety program. *Milbank Q* 2012;90:548–91.

60. Hollnagel E, Wears RL, Braithwaite J. *From Safety-I to Safety-II: a white paper*. USA, Australia: Published simultaneously by the University of Southern Denmark, University of Florida, and Macquarie University, The Resilient Health Care Net, 2015.

61. Vincent C, Burnett S, Carthey J. *The measurement and monitoring of safety*. UK: The Health Foundation, 2013.

62. Buckley C, Cooney K, Sills E, *et al*. Implementing the safety thermometer tool in one NHS trust. *Br J Nurs* 2014;23:268–72.

63. Power M, Fogarty M, Madsen J, *et al*. Learning from the design and development of the NHS safety thermometer. *Int J Qual Health Care* 2014;26:287–97.

64. NHS Quality Observatory. NHS safety thermometer 2013. https://www.safetythermometer.nhs.uk/index.php?option=com_content&view=article&id=1&Itemid=101 (cited 12th May 2017).

65. Rostami P, Power M, Harrison A, *et al*. Learning from the design, development and implementation of the medication safety thermometer. *Int J Qual Health Care* 2017;29:301–9.