BMJ

RESEARCH

# Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation

Amirta Benning, programme manager,[1] Maisoon Ghaleb, lecturer in pharmacy practice/patient safety,[2,3] Anu Suokas, translational research facilitator,[4] Mary Dixon-Woods, professor of medical sociology,[5] Jeremy Dawson, research fellow,[6] Nick Barber, professor of the practice of pharmacy,[2] Bryony Dean Franklin, professor of medication safety and director, centre for medication safety and service quality,[2,7] Alan Girling, senior research fellow,[1] Karla Hemming, senior research fellow,[1] Martin Carmalt, consultant physician,[8] Gavin Rudge, data scientist,[1] Thirumalai Naicker, honorary research associate,[1] Ugochi Nwulu, senior research associate/coordinator,[9] Sopna Choudhury, research associate,[1] Richard Lilford, professor of clinical epidemiology[1]

[1]School of Health and Population Sciences, University of Birmingham, Edgbaston, West Midlands B15 2TT, UK

[2]Department of Practice and Policy, School of Pharmacy, University of London, London WC1N 1AX

[3]School of Pharmacy, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB

[4]Arthritis Research UK Pain Centre, Academic Rheumatology, Clinical Sciences Building, City Hospital, Nottingham NG5 1PB

[5]Department of Health Sciences, University of Leicester, Leicester LE1 7RH

[6]Work and Organisational Psychology Group, Aston Business School, Aston University, Aston Triangle, Birmingham B4 7ET

[7]Imperial College Healthcare N HS Trust, St Mary's Hospital, London W2 1NY

[8]Royal Orthopaedic Hospital, Bristol Road South, Northfield, Birmingham B31 2AP

[9]Clinical Investigation Unit, University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital, Queen Elizabeth Medical Centre, Birmingham B15 2TH

**Correspondence to: R J Lilford**
**r.j.lilford@bham.ac.uk**

Cite this as: *BMJ* 2011;342:d195
doi:10.1136/bmj.d195

## ABSTRACT

**Objectives** To conduct an independent evaluation of the first phase of the Health Foundation's Safer Patients Initiative (SPI), and to identify the net additional effect of SPI and any differences in changes in participating and non-participating NHS hospitals.

**Design** Mixed method evaluation involving five substudies, before and after design.

**Setting** NHS hospitals in United Kingdom.

**Participants** Four hospitals (one in each country in the UK) participating in the first phase of the SPI (SPI1); 18 control hospitals.

**Intervention** The SPI1 was a compound (multi-component) organisational intervention delivered over 18 months that focused on improving the reliability of specific frontline care processes in designated clinical specialties and promoting organisational and cultural change.

**Results** Senior staff members were knowledgeable and enthusiastic about SPI1. There was a small (0.08 points on a 5 point scale) but significant (P<0.01) effect in favour of the SPI1 hospitals in one of 11 dimensions of the staff questionnaire (organisational climate). Qualitative evidence showed only modest penetration of SPI1 at medical ward level. Although SPI1 was designed to engage staff from the bottom up, it did not usually feel like this to those working on the wards, and questions about legitimacy of some aspects of SPI1 were raised. Of the five components to identify patients at risk of deterioration—monitoring of vital signs (14 items); routine tests (three items); evidence based standards specific to certain diseases (three items); prescribing errors (multiple items from the *British National Formulary*); and medical history taking (11 items)—there was little net difference between control and SPI1 hospitals, except in relation to quality of monitoring of acute medical patients, which improved on average over time across all hospitals. Recording of respiratory rate increased to a greater degree in SPI1 than in control hospitals; in the second six hours after admission recording increased from 40% (93) to 69% (165) in control hospitals and from 37% (141) to 78% (296) in SPI1 hospitals (odds ratio for "difference in difference" 2.1, 99% confidence interval 1.0 to 4.3; P=0.008). Use of a formal scoring system for patients with pneumonia also increased over time (from 2% (102) to 23% (111) in control hospitals and from 2% (170) to 9% (189) in SPI1 hospitals), which favoured controls and was not significant (0.3, 0.02 to 3.4; P=0.173). There were no improvements in the proportion of prescription errors and no effects that could be attributed to SPI1 in non-targeted generic areas (such as enhanced safety culture). On some measures, the lack of effect could be because compliance was already high at baseline (such as use of steroids in over 85% of cases where indicated), but even when there was more room for improvement (such as in quality of medical history taking), there was no significant additional net effect of SPI1. There were no changes over time or between control and SPI1 hospitals in errors or rates of adverse events in patients in medical wards. Mortality increased from 11% (27) to 16% (39) among controls and decreased from 17% (63) to 13% (49) among SPI1 hospitals, but the risk adjusted difference was not significant (0.5, 0.2 to 1.4; P=0.085). Poor care was a contributing factor in four of the 178 deaths identified by review of case notes. The survey of patients showed no significant differences apart from an increase in perception of cleanliness in favour of SPI1 hospitals.

**Conclusions** The introduction of SPI1 was associated with improvements in one of the types of clinical process studied (monitoring of vital signs) and one measure of staff perceptions of organisational climate. There was no additional effect of SPI1 on other targeted issues nor on other measures of generic organisational strengthening.

## INTRODUCTION

Since the publication of two key reports in 2000,[1 2] increased efforts, attention, and resources have been given to patient safety, but by 2005 progress was reported as frustratingly slow.[3] There is now keen interest in finding ways to make more sustained improvements. Approaches to patient safety that aim to intervene at an organisational level and that actively engage workers in the management of risk are argued to be especially promising for providing solutions to the problems of patient safety.[4 5] The Health Foundation's Safer Patients Initiative (SPI) was an important example of such an approach.

### The Safer Patients Initiative programme

The Health Foundation selected four hospitals (table 1), one in each country of the United Kingdom, to participate in the first phase of SPI (SPI1).[6] The Health Foundation (a British charity dedicated to improving the quality of healthcare) invested £775 000 (€900 000, $1.2m) in each hospital. SPI1 ran from January 2005 to September 2006 inclusive and was intended to embed and spread thereafter. The Health Foundation issued a request for applications to conduct an independent evaluation of the SPI1 intervention. We report on the results of this evaluation.

SPI1 set out to secure improvements in patient safety[7-9] and covered a range of aims (box 1), which included a 50% reduction in adverse events.[10] Mentored by the Institute for Healthcare Improvement (IHI), based in the United States, it had several features similar to the well publicised US "Saving 100,000 lives campaign."[11 12] The intervention consisted of many components (box 1), some generic (aimed at strengthening the organisation as a whole) and others targeted at specific high risk clinical issues. The generic components sought to promote patient safety as an organisational priority, increase the effectiveness of senior leadership in relation to safety, engender a culture of safety, and instil knowledge of the principles of safe practice among staff. The targeted components of SPI1 focused on improving the reliability of specific frontline care processes in designated clinical areas (general wards, critical care, perioperative care, and management of medicines).

From each hospital, 15-20 "change agents" participated in four learning sessions run by the IHI to gain knowledge and expertise in how to change practice, measure the effects, and sustain safety improvements.[7-9] Change agents were charged with leading change by facilitating the implementation of SPI1 in their hospitals. They also formed a virtual community that shared data, tools, expertise, and experience; participated in conference calls and electronic mailing lists; and uploaded monthly progress reports through a project intranet.

SPI1 hospitals received site visits and technical assistance from the IHI over the course of the programme, and staff members in participating hospitals were asked to take part in surveys of safety culture managed by the IHI. Changes in clinical processes were supported by learning materials (such as care "bundles," collections of carefully packaged evidence based standards directed at a particular condition or clinical scenario) (box 1). The intention was to encourage compliance with the tenets of safe and effective care by using Plan-Do-Study-Act (PDSA) cycles, where changes are first implemented on a small scale and tested and refined before being rolled out on a larger scale. The programme strongly emphasised the role of measurement in improving patient safety. Staff members were expected to upload time series process data (such as compliance with prophylaxis for deep vein thrombosis) to the project intranet to create run charts—graphs to help people visualise changes on selected measures over time. Monitoring of care processes through these charts was seen as important for guiding local improvement activities and evaluating their impact.

### Study aims

We evaluated changes in a range of end points relevant to patient safety over time and whether these changes could be attributed to SPI1 in the four participating hospitals. We used a controlled design to identify whether SPI1 had any additional effect (that is, net of any overall trends in hospital safety over the same period).

We have reported here on the methods used and the main findings based on SQUIRE guidelines.[13] Further details are available in the full online report.[6] The evaluation of the rollout of SPI to other hospitals (SPI2) is available in a companion paper.[14]

## METHODS

### Framework for evaluation

The study was a "mixed method" evaluation[15 16] because it combined qualitative and quantitative data and involved many "levels" in the organisation, "from boardroom to patient."[15-19] This type of evaluation is particularly suitable for service delivery/management interventions that are not likely to yield the type of conclusive results characteristic of evaluations of treatments based on measurement of outcomes on patients.[20] Mixed method evaluation draws on the idea of "triangulation," where confidence in the findings increases when observations of one type are corroborated by other types of evidence. This follows a tradition in the philosophy of science.[15-19] Figure 1 shows the framework for the evaluation.

**Table 1 | NHS hospitals* that participated in phase one of Safer Patients Initiative (SPI1)**

| Hospital number | Rural/ urban | No of beds | Teaching status | A and E | ICU | FTE consultant (specialists) |
|---|---|---|---|---|---|---|
| 1 | Urban | 625 | Associate teaching hospital | Yes | Yes | 112 |
| 2 | Rural | 750 | No | Yes | Yes | 120 |
| 3 | Urban | 903 | Principal teaching hospital | Yes | Yes | 242 |
| 4 | Rural | 280 | No | Yes | No | 36 |

A and E=accident and emergency department; FTE=full time equivalent; ICU=intensive care unit.
*National Health Service hospital, with no private beds. Figures provided as of October 2004.

**Box 1: Key generic and specific elements of the Safer Patients Initiative**

General aim: "to avoid unnecessary harm, pain or suffering as a result of error in medical interventions"

**Generic improvement to reduce adverse events**

***Aims***
- Build a culture of safety and good leadership
- Train to enable organisations to identify problems and develop and evaluate methods to reduce risk
- Foster an understanding of the principles of safe practice

***Approach used in SPI***
- Collaborative residential learning sessions with the Institute for Healthcare Improvement (IHI) faculty to train "change agents" who would lead implementation in their hospitals
- Instruction in use of control charts
- Web based learning and site visits from IHI expert faculty
- Know-how for Plan-Do-Study-Act (PDSA) cycles
- Electronic facility for information sharing—for example, to share results of PDSA cycles
- Leadership projected in part by management walk rounds
- Participation in safety culture surveys
- Formation of a collaborative learning community

**Specific interventions**

***Aim***
- Identify and respond to deteriorating patients to reduce need for "crash calls" and avoidable mortality

***Approach used in SPI***
- Review of 50 deaths carried out by hospital staff
- Tools for monitoring patients' condition and for triggering action, such as forms to record vital signs and other salient information (such as early warning score system (EWSS))
- Promote the use of risk (severity) scores
- Establish a rapid response team

***Aim***
- Reduce medication errors

***Approach used in SPI***
- Assessment of medication safety by involving staff in failure mode effects analysis (FMEA); educating staff to identify and remedy weak links in medication practice from prescribing to administration and monitoring
- Tool to identify and measure rates of adverse events
- Tool to reduce adverse events from treatment with anticoagulants
- Education to improve "medicine reconciliation," to ensure that medicines that patients were taking before admission were not inadvertently omitted or altered after admission

***Aim***
- Communication between staff to reduce adverse events/mortality

***Approach used in SPI***
- Situation background assessment recommendation (SBAR) tool to ensure that information is communicated in a structured way
- Safety briefings; briefings at shift changes to ensure staff members are aware of relevant information for patients

***Aim***
- Infection control including for meticillin resistant *Staphylococcus aureus* (MRSA)

***Approach used in SPI***
- Perioperative antibiotics to reduce infection at surgical site
- Catheter insertion and maintenance drill to prevent central line infections in intensive care
- Following the tenets of a package of ventilator guidelines (bundles) to reduce ventilator acquired pneumonia, venous thromboembolism, and stress ulcers in intensive care units
- Improvements in hand hygiene—for example, by training and use of posters

### Substudies

The evaluation comprised five substudies to investigate key aims of the initiative (table 2).

*Improvement of leadership and engagement of leaders and staff*—Semistructured interviews were conducted to assess knowledge and enthusiasm for the initiative among 60 senior members of staff in the four SPI1 hospitals.

*Improvement in culture of safety and "transformative" effects*—Before and after surveys of staff attitudes in control and SPI1 hospitals were conducted by means of a validated questionnaire to assess staff morale, attitudes, and aspects of culture (the NHS National Staff Survey).

*Informing, educating, and motivating frontline clinical staff*—Qualitative studies comprising ethnographic observations on acute medical wards, interviews, and focus groups in SPI1 hospitals were conducted to obtain direct access to staff behaviour and views, and to add depth of understanding.

*Impact on processes of clinical care*—To identify any improvements, we measured error rates in control and SPI1 hospitals by means of explicit (criterion based) and separate holistic reviews of case notes. The study group comprised patients aged 65 or over who had been admitted with acute respiratory disease: this is a high risk group to whom many evidence based guidelines apply and hence where significant effects were plausible.

*Improving outcomes of care*—We reviewed case notes to identify adverse events and mortality and assessed any improvement in patients' experiences by using a validated measure of patients' satisfaction (the NHS patient survey).

### Study sample

Data were collected in the four SPI1 intervention hospitals and 18 control hospitals. For economy, the control hospitals comprised nine that would later participate in the second phase of SPI (SPI2) but were pre-intervention at the time of this study. A further nine controls were added that would also form the controls for the SPI2 hospitals.[14] All 18 control hospitals were located in England (thereby providing access to the NHS staff and patient surveys).

Use of before and after observations across control and SPI1 sites enabled us to compare rates of change across control and SPI1 hospitals over time. This approach, known as the "difference in difference" method,[18] is particularly suitable for identifying temporal variations in outcomes that are not due to exposure to the intervention that is being evaluated, and is used to take into account the effect of events other than the particular intervention of interest that occur over the intervention period.

### Interviews with strategic/senior staff

Using semistructured telephone interviews with 60 hospital staff members in strategic/senior positions across the four hospitals, we investigated how far participants understood and expressed enthusiasm for the SPI1. A researcher (JW, see acknowledgments)

Table 2 | Summary of substudies comprising evaluation of phase one of Safer Patients Initiative (SPI1)

| Substudy and topic | Data source | Location | Unit of analysis (quantitative studies) |
|---|---|---|---|
| **Interviews with senior staff** | | | |
| Impact of SPI at senior management level | Semistructured interviews with senior hospital staff | SPI1 hospitals | NA |
| **Staff survey** | | | |
| Staff morale, culture, and opinion | Questionnaire as used in NHS National Staff Survey | Control and SPI1 hospitals | Staff member |
| **Qualitative study** | | | |
| Impact of SPI on practitioners at ward level | Ethnographic observations, interviews, and focus groups in acute medical wards | SPI1 hospitals | NA |
| **Quality of care: acute medical care** | | | |
| Quality of care of patients aged >65 with acute respiratory disease | Case note reviews (both explicit and holistic) | Control and SPI1 hospitals | Patient |
| **Outcomes** | | | |
| Adverse events in patients aged >65 with acute respiratory disease | Holistic case note review | Control and SPI1 hospitals | Patient |
| Hospital mortality in patients aged >65 with acute respiratory disease | Case note review | Control and SPI1 hospitals | Patient |
| Patient satisfaction | Questionnaire as used in NHS patient surveys | Control and SPI1 hospitals | Hospital |

NA=not applicable.

conducted the interviews, which were tape recorded and transcribed. Data analysis was first based on the constant comparative method that was used to generate thematic categories into which participants' accounts could be categorised.[21] Three independent reviewers then scored the responses of hospital participants on a 1-10 scale for level of knowledge and level of enthusiasm.

### Staff survey

The National NHS Staff Survey questionnaire was used at two time points to measure variables such as staff morale, attitudes, and aspects of "culture" that might be affected by the generic strengthening of organisational systems that SPI1 intended to achieve. Eleven of the 28 survey questions (table 3)[22-26] were identified as likely to be of relevance. Questionnaires were sent to all staff members in the four SPI1 hospitals, including the three hospitals outside England that did not routinely participate in the National NHS Staff Survey. In each of the 18 control hospitals, a simple random sample of 850 staff members was used; at this sample size, an average 60% response rate would yield 95% confidence intervals of no greater than 10% for all scores in a single organisation. Further detail about the questionnaire is available from www.nhsstaffsurveys.com, and further detail about the evaluation is in the full report.[6] Statistical analyses were adjusted for age, sex, ethnic background, occupational group, length of service, and management status.

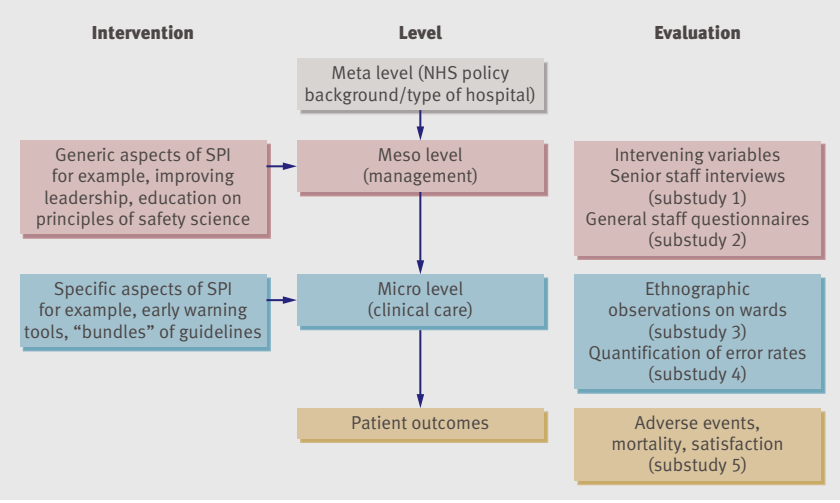


Fig 1 | General scheme for evaluation of first phase of Safer Patients Initiative

### Qualitative study

One author (AS) undertook three rounds of data collection. Between April and September 2006, she visited one medical ward in each of the four hospitals for one week. She undertook around 150 hours of ethnographic observations and 47 interviews with different types of ward staff, focusing on general issues relating to patient safety and SPI1. The wards were selected on the basis that they would treat many of the patients whose case notes would be reviewed (see below), but not on the basis that SPI1 "change agents" (see box 1) were necessarily active on these wards. From April to June 2007, shortly after the intervention phase, a second week-long visit to each ward involved a further round of ethnographic observations (around 150 hours) and 41 interviews. A third visit involving three focus groups at each site (one at study ward level, one involving people with responsibilities for patient safety/SPI1, and one at strategic level) was used to feed back preliminary findings and to ask staff for their reflections on SPI1. These were carried out in

**Table 3** | Staff survey scores in control and SPI1 hospitals at two periods* in evaluation of phase one of Safer Patients Initiative (SPI1)

| | Control hospitals | | | | | SPI1 hospitals | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Survey 1 | | Survey 2† | | | Survey 1 | | Survey 2† | | | | | |
| Survey question‡ | No of responders | Score (SE) | No of responders | Score (SE) | Absolute % change | No of responders | Score (SE) | No of responders | Score (SE) | Absolute % change | Range at baseline§ | Effect of SPI1 (99% CI)¶, P value | Point estimate favours SPI1 |
| Well structured appraisals within previous 12 months[22 23] | 8046 | 39 (1) | 7260 | 28 (1) | −10 | 6111 | 34 (1) | 3993 | 27 (1) | −7 | 27-46 | 3 (−2 to 8), 0.095 | Yes |
| Working in well structured teams[24] | 8052 | 40 (1) | 7279 | 37 (1) | −3 | 6088 | 38 (1) | 3961 | 35 (1) | −2 | 34-52 | 1 (−4 to 6), 0.510 | Yes |
| Witnessed potentially harmful errors or near misses in previous month | 8236 | 47 (1) | 7520 | 39 (1) | −8 | 6242 | 48 (1) | 4133 | 44 (1) | −4 | 41-56 | 5 (−2 to 11), 0.068 | No |
| Work related injury in previous 12 months | 8286 | 22 (0) | 7372 | 19 (0) | −3 | 6246 | 21 (1) | 4023 | 18 (1) | −3 | 18-26 | 0 (−4 to 3), 0.854 | NA |
| Work related stress in previous 12 months | 8368 | 34 (1) | 7457 | 33 (1) | −1 | 6340 | 34 (1) | 4094 | 30 (1) | −4 | 29-39 | −5 (−12 to 0), 0.013 | Yes |
| Physical violence from patients/ relatives in previous 12 months | 8283 | 13 (0) | 7482 | 11 (0) | −2 | 6290 | 15 (0) | 4117 | 11 (0) | −3 | 9-18 | −3 (−8 to 0), 0.026 | Yes |
| Intention to leave[25] | 8263 | 3.36 (0.01) | 7437 | 3.29 (0.01) | −0.08 | 6282 | 3.29 (0.01) | 4111 | 3.21 (0.01) | −0.09 | 2.99-3.51 | 0.04 (−0.03 to 0.10), 0.139 | No |
| Staff job satisfaction[25] | 8357 | 3.47 (0.01) | 7495 | 3.37 (0.01) | −0.10 | 6366 | 3.42 (0.01) | 4122 | 3.34 (0.01) | −0.09 | 3.25-3.57 | 0.03 (−0.02 to 0.08), 0.132 | Yes |
| Quality of work-life balance[25] | 8249 | 2.63 (0.01) | 7436 | 2.72 (0.01) | 0.10 | 6330 | 2.61 (0.01) | 4106 | 2.67 (0.02) | 0.06 | 2.45-2.80 | −0.05 (−0.12 to 0.03), 0.106 | No |
| Support from supervisors[25] | 8310 | 3.45 (0.01) | 7477 | 3.41 (0.01) | −0.05 | 6334 | 3.36 (0.01) | 4109 | 3.32 (0.01) | −0.04 | 3.18-3.55 | 0.02 (−0.04 to 0.08) 0.358 | Yes |
| Organisational climate[25 26] | 8302 | 3.11 (0.01) | 7424 | 2.89 (0.01) | −0.22 | 6267 | 3.08 (0.01) | 4070 | 2.93 (0.01) | −0.15 | 2.82-3.32 | 0.08 (0.02 to 0.13), <0.01 | Yes |

NA=not applicable.
*Survey periods do not map precisely to epochs described in methods section.
†After intervention.
‡First six scores are percentages, simply reflecting proportion of respondents who answered "yes" to single question or set of questions. The five others are on scale of 1-5 and are based on mean of between three and six questions, each of which was scored between 1 and 5 for each respondent. For four of these five scores (quality of work-life balance, staff job satisfaction, support from supervisors, and organisational climate), higher scores are better, except for intention to leave.
§Indicates range of scores across intervention and control hospitals in first survey to give some context for level of change shown. Difference in change and corresponding confidence interval does not necessarily reflect difference in absolute change because of inclusion of covariates in models tested.
¶Values differ from simple subtraction due to rounding.

**Table 4 | Explicit case note review in phase one of Safer Patients Initiative (SPI1): areas of review, source of criteria, method for assessment of errors, and relevant SPI target**

| Area of review | Source of criteria | How it was assessed | Specific SPI goal |
|---|---|---|---|
| Clinical observations and signs of patient deterioration | British Thoracic Society (BTS)[31 32] | Completeness with which patients' vital signs were recorded (table 5) evaluated on admission and then for first and subsequent 6 hour time periods. Vital sign data recorded in case notes constituted numerator, while all vital signs that should have been recorded constituted denominator | Yes |
| Quality of medical history taking | Expert opinion for patient admitted with acute respiratory disease | 11 items (table 8) constituting ideal history each marked as inquired into or not | No |
| Compliance with evidence based care specific classes of patients | Expert opinions, BTS, or *British National Formulary* (BNF), or both as specified | Proportion of patients for whom urea and electrolytes, chest x ray, and full blood count ordered within 6 hours of admission (expert opinion) | No |
| | | Proportion of patients with pneumonia for whom CURB score was calculated (BTS, BNF) | Yes |
| | | Proportion of patients with COPD and asthma receiving systemic steroids (BTS) | No |
| | | Proportion of patients with asthma in whom peak flow was measured (expert opinion) | No |
| Prevalence of prescribing errors | Previously developed forms based on BNF[33] | Errors were identified and categorised according to stage of drug use process.[34] Denominator was number of drugs used | Yes |

CURB=confusion/urea/respiratory rate/blood pressure score; COPD=chronic obstructive pulmonary disease.

May-July 2008 by two of the authors (AS and UG) and one researcher (JW).

Data analysis of interviews was based on the constant comparative method,[21] facilitated by the use of NVivo software. For focus groups and ethnographic field notes, simple coding procedures were used to categorise the data thematically.

### Quality of care/error rates in patients with acute respiratory disease on medical wards

Using case note review, we assessed quality of care using standards based on established guidelines wherever possible. We reviewed case notes from patients with similar clinical conditions across the SPI1 and control hospitals so that we could compare like with like. Case note review is resource intensive and thus must be selective. We focused on patients aged over 65 with acute respiratory disease (community acquired pneumonia, exacerbation of asthma, or chronic obstructive pulmonary disease) admitted to medical wards (one in each participating hospital) (box 2). The review assessed processes targeted by SPI1 (such as quality of observations to detect deteriorating patients) and those that might be expected to improve if there was an overall shift in organisational systems and culture toward patient safety (such as improved medical histories).

**Box 2: Rationale for focus on patients aged over 65 admitted with acute respiratory disease**

- It was important to focus on an issue where all four SPI1 hospitals would be implementing a specific SPI intervention (intensive care was excluded partly for this reason)
- There is a high incidence of comorbidities in people aged over 65, making this a high risk and hence a potentially error-rich population in which an effective organisational patient safety intervention might yield detectable improvements. Improving recognition and response to acute deterioration was a specific target of the SPI (see box 1), and patients admitted with acute respiratory disease are at high risk of such deterioration[27 28]
- There was evidence that monitoring and medication practice was suboptimal in NHS hospitals, thus providing sufficient room for improvements to be detected with samples of affordable size[29]
- A single set of case notes could be used to assess end points targeted by several SPI interventions (see box 1), specifically including management of patients at risk of deterioration and prescribing errors

Case notes from the 18 control and four SPI1 hospitals were collected from two periods of six months, which we termed epochs. Epoch 1 (October 2003 to March 2004) was before the intervention and epoch 2 (October 2006 to March 2007) was after.

We aimed to analyse 100 sets of case notes from each SPI1 hospital per epoch (800 in total) and 15 from each control hospital per epoch (540 in total). This would give 80% power to detect a 13 percentage point improvement in error rates from a baseline of 70% at P=0.05 (see full report for further details).[6] For each set of case notes, the admissions of interest were photocopied, anonymised, and digitised and the year of admission was removed to ensure blinding to epoch. These processes were completed before reviewing. Review was both explicit (criterion based) and implicit (holistic) because each method identifies a different spectrum of errors.[30]

*Criterion based review*—Qualified pharmacists (MG and BDF) conducted explicit reviews according to predefined criteria relating to deterioration of the patient; medical history taking; compliance with evidence based care; and prescribing errors (table 4). To control for any learning or fatigue effects, or both, in reviewers, case notes were scrambled to ensure that they were not reviewed entirely in series. Agreement on prescribing error between observers[35] was evaluated by assigning one in 10 sets of case notes to both reviewers, who assessed cases in batches, blinded to each other's assessments, but compared and discussed results after each batch. Data were collected on the adequacy of masking of patients' names, hospital, and epoch. All analyses were adjusted for patient level covariates of age and sex. We used cubic polynomials in the time of review to adjust for learning/fatigue effects in the review process.

*Holistic review*—Holistic review was not part of the original protocol (for either SPI1 or SPI2) but was carried out after a suggestion by the late Vin McLoughlin (a director of the Health Foundation) in November

**Table 5** | Vital signs and routine investigations before (epoch 1) and after (epoch 2) phase one of Safer Patients Initiative (SPI1). Figures are percentage compliance (binomial standard error (SE)) and odds ratios

| | Control hospitals | | SPI1 hospitals | | OR (99% CI), P value | |
|---|---|---|---|---|---|---|
| | Epoch 1 (n=236) | Epoch 2* (n=240) | Epoch 1 (n=381) | Epoch 2* (n=380) | Changes in controls | Effect of SPI1† |
| **On admission** | | | | | | |
| Temperature | 98 (1) | 99 (1) | 99 (1) | 99 (1) | 5.1 (0.3 to 89.5), 0.144 | 0.2 (0.01 to 8.5), 0.289 |
| Respiratory rate | 96 (1) | 99 (1) | 91 (2) | 98 (1) | 4.7 (0.6 to 36.5), 0.052 | 1.5 (0.2 to 16.0), 0.677 |
| Cyanosis/oxygen saturation | 99 (1) | 99 (1) | 98 (1) | 99 (1) | 1.7 (0.2 to 18.2), 0.578 | 2.7 (0.1 to 55.2), 0.385 |
| Confusion/mental state | 58 (3) | 65 (3) | 67 (2) | 69 (2) | 1.2 (0.7 to 2.1), 0.307 | 1.2 (0.6 to 2.6), 0.437 |
| Pulse | 99 (1) | 99 (1) | 99 (1) | >99 (<1) | 3.3 (0.2 to 68.7), 0.306 | 0.5 (0.01 to 23.7), 0.614 |
| Blood pressure | 99 (1) | 99 (1) | 99 (1) | >99 (<1) | 3.3 (0.2 to 68.7), 0.306 | 0.5 (0.01 to 23.7), 0.614 |
| **At 6 hours** | | | | | | |
| Temperature | 62 (3) | 74 (3) | 76 (2) | 86 (2) | 1.7 (1.0 to 2.9), 0.008 | 1.0 (0.5 to 2.2), 0.976 |
| Respiratory rate | 44 (3) | 73 (3) | 43 (3) | 82 (2) | 3.6 (2.1 to 6.2), <0.001 | 2.0 (1.0 to 4.2), 0.015 |
| Pulse | 67 (3) | 77 (3) | 83 (2) | 88 (2) | 1.7 (1.0 to 3.0), 0.012 | 1.0 (0.4 to 2.3), 0.973 |
| Oxygen saturation | 61 (3) | 75 (3) | 77 (2) | 88 (2) | 1.9 (1.1 to 3.3), 0.002 | 1.3 (0.6 to 2.8), 0.425 |
| **At 12 hours** | | | | | | |
| Temperature | 59 (3) | 70 (3) | 71 (2) | 82 (2) | 1.8 (1.0 to 3.0), 0.005 | 1.3 (0.6 to 2.8), 0.314 |
| Respiratory rate | 40 (3) | 69 (3) | 37 (3) | 78 (2) | 3.7 (2.2 to 6.2), <0.001 | 2.1 (1.0 to 4.3), 0.008 |
| Pulse | 62 (3) | 74 (3) | 75 (2) | 83 (2) | 1.9 (1.1 to 3.2), 0.002 | 1.2 (0.5 to 2.5), 0.618 |
| Oxygen saturation | 56 (3) | 73 (3) | 64 (3) | 82 (2) | 2.3 (1.4 to 3.9), <0.001 | 1.4 (0.7 to 2.9), 0.234 |
| **Routine investigations** | | | | | | |
| Urea and electrolytes | >99 (<1) | 99 (1) | 99 (1) | 99 (1) | 0.6 (0.02 to 14.1), 0.665 | 0.8 (0.02 to 39.5), 0.865 |
| Chest x ray | 97 (1) | 98 (1) | 95 (1) | 94 (1) | 2.4 (0.5 to 11.8), 0.164 | 0.5 (0.1 to 3.0), 0.291 |
| Full blood count | 99 (1) | 98 (1) | 99 (1) | 98 (1) | 1.2 (0.1 to 10.5), 0.789 | 0.2 (0.01 to 5.4), 0.223 |

*After intervention.
†OR >1 favours SPI1.

2007 at a project progress meeting. A specialist in general medicine (MC) holistically evaluated each set of case notes to identify errors in the care of patients and adverse events. To measure reliability between observers,[35] an experienced trainee in respiratory medicine (TN) independently re-evaluated a subset of these case notes (n=91). Error was defined, according to Reason, as "a failure to complete a planned action as it was intended or adoption of an incorrect plan."[36] The results, adjusted for age and sex, are presented as the average number of errors per 100 patients; a patient could have more than one error. Errors were categorised as relating to diagnosis on admission; hospital acquired infection; technical/management; medication/maintenance/test results; clinical reasoning; and discharge information.[37-41] Reviewers also gave an overall rating of the quality of care.[6]

## Outcomes

*Adverse events*—The holistic review of case notes identified adverse events related to healthcare, which were reported as a rate of total adverse events per 100 patients. An adverse event was defined as "an unintended injury or complication which results in disability, death or prolonged hospital stay and which is caused by healthcare management."[38] Reviewers classified adverse events (including those resulting in death) by degree of preventability (on the balance of probabilities). Adjustments were made for age and sex.

*Mortality*—Overall adjusted (or crude) hospital mortality rates for all four SPI1 hospitals were not available because of difficulty accessing this information from the non-English countries. Overall mortality was unlikely to be a useful measure for this study as it is an insensitive marker of avoidable mortality,[42] creating a high risk of a false negative result in a study with only four intervention hospitals.[43] We therefore compared mortality rates across epochs among patients in medical wards whose case notes had been included in the case note review. Adjustments were made for age, sex, and number of comorbidities.

*Patients' satisfaction*—An emphasis on safety can make staff and patients feel valued.[44] We used a survey

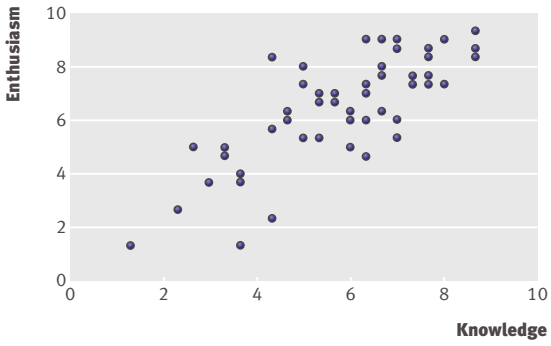


**Fig 2** | Correlation between knowledge and enthusiasm for SPI1 among senior strategic hospital staff (some points represent results for more than one interviewee)

Table 6 | Compliance with particular standards (such as use of steroids when indicated) before (epoch 1) and after (epoch 2) phase one of Safer Patients Initiative (SPI1). Figures are numbers (percentage, SE) and odds ratios for effect of SPI1

| | Control hospitals | | SPI1 hospitals | | |
|---|---|---|---|---|---|
| | Epoch 1 | Epoch 2* | Epoch 1 | Epoch 2* | OR (99% CI)†, P value |
| Asthma or COPD: given systemic steroids within 24 hours | 129 (88, 3) | 135 (93, 2) | 224 (91, 2) | 199 (88, 2) | 0.7 (0.2 to 2.9), 0.500 |
| Asthma peak flow record‡ | 34 (79, 7) | 29 (83, 7) | 78 (82, 4) | 37 (65, 8) | 0.6 (0.04 to 8.2), 0.570 |
| Pneumonia: CURB score recorded in notes | 102 (2, 1) | 111 (23, 4) | 170 (2, 1) | 189 (9, 2) | 0.3 (0.02 to 3.4), 0.173 |

COPD=chronic obstructive pulmonary disease; CURB=confusion/urea/respiratory rate/blood pressure score.
*After intervention.
†OR >1 favours SPI1.
‡Item with significant between hospital components of variation within arms of study (P<0.01).

based on the questionnaire used in the National NHS Acute Inpatient Survey in England (www.nhssurveys.org) to obtain evidence of such a "halo effect" on patients' experiences. The dates of the surveys of patients were aligned with the dates of staff surveys.[6] Five scores were identified for analysis: three related to overall satisfaction scores and two to cleanliness. Data in the control arm were available only at the level of the organisation (hospital), not at the level of the individual clinical directorate (medicine, surgery, etc). Thus the analysis was conducted at the organisational level throughout. In the SPI1 arm, organisation level scores were formed by averaging all respondents' scores in each hospital.

### Statistical methods

We used generalised linear mixed models for formal statistical analyses. Fixed effects were included for the difference in levels before the intervention between control and SPI1 hospitals ("baseline comparisons"); the change over time in control hospitals between the period before (epoch 1) and after (epoch 2) the intervention; and the effect of SPI1, interpreted as the discrepancy between the temporal changes experienced over the two epochs in the control and SPI1 hospitals. The impact of intraclass correlation in before and after studies is lower than for cross sectional studies.[18 45] Nevertheless, we took into account the effect of hospital level clustering by including random effects for each hospital.

We used logistic models for binary responses; Poisson models for adverse events and general care errors; negative binomial models for medication errors (per recorded prescription); and normal models in all other cases. No adjustments were made for multiple comparisons, but we used 99% confidence intervals throughout and significance was set at $P<0.01$. Agreement between observers was tested with the intraclass correlation coefficient for rating scales and the $\kappa$ statistic for dichotomous judgments. All models were fitted in Stata v10.

## RESULTS

### Senior/strategic staff interviews

Interviews with senior/strategic hospital staff in the SPI1 hospitals showed that they were mainly positive about SPI1 and had considerable enthusiasm for the initiative and its principles. Only seven of the 60 participants were unable to describe SPI1 accurately or in detail.[8] Even though there was a broadly shared understanding of the programme's theory of change (what the intervention was intended to achieve and with what methods), there was less evidence of an explicit and shared organisational theory of change (what was needed to make the programme work at an organisational level and how programme implementation could be optimised). Despite their enthusiasm, hospital staff members expressed multiple concerns about the ambitious reach of the programme, whether resources would be equal to the demands, and whether resistance might be encountered at the "sharp end"[46] where staff care for patients. Strategies for overcoming such difficulties were not clear in the accounts of stakeholders.

The quantitative analysis found that 73% scored above 5 (out of 10) on the knowledge scale and 83% scored above 5 (out of 10) on the enthusiasm scale (fig 2). Correlations between knowledge and enthusiasm across the three raters were 0.61, 0.69, and 0.91. Reliability between raters was medium to high, with intraclass correlations of 0.55 for knowledge and 0.65 for enthusiasm.

### Staff survey

There was no significant difference in response rates between control and SPI1 hospitals at baseline, and the hospitals were broadly similar with respect to morale, attitudes, and culture of the hospital staff at baseline. Table 3 shows the changes in control and SPI1 hospitals on each of 11 scores, along with the differences between groups. The difference was significant for only one of the 11 scores (organisational climate) ($P<0.01$), favouring the SPI1 hospitals over controls. There was a modest effect size for the difference in change between the control and SPI1 hospitals after covariates were taken into account: 0.08 points on a 5 point scale where there was a range at baseline of 0.50 points between hospitals.

### Qualitative study

The focus groups in four medical wards across the four sites agreed that the senior staff members in the hospitals were committed and enthusiastic about SPI1, made a significant strategic contribution, gave weight to the initiative, and generally set a good example for staff.

**Table 7 | Analysis of prescribing errors before (epoch 1) and after (epoch 2) phase one of Safer Patients Initiative (SPI1)**

| | Control hospitals | | SPI1 hospitals | |
|---|---|---|---|---|
| | Epoch 1 | Epoch 2* | Epoch 1 | Epoch 2* |
| No of patients† | 233 | 239 | 381 | 378 |
| No of prescriptions | 5482 | 6207 | 10664 | 11538 |
| Prescriptions per patient | 23.5 | 26.0 | 28.0 | 30.5 |
| Total No of errors | 596 | 564 | 1157 | 1530 |
| Error rate (SE)‡ | 0.14 (0.02) | 0.11 (0.01) | 0.15 (0.02) | 0.15 (0.01) |

*After intervention.
†With medication charts available for review.
‡Adjusted for date of review.

> "If these guys aren't behind it very quickly your clinical directors and . . . other directors, you know [other] senior people start to fall by the wayside and so I think that's absolutely paramount having the top guys leading the way, so I think that has been one of the big successes."

The involvement of the IHI in SPI1 was seen as crucial in lending credibility and support to the initiative and was much valued as a source of knowledge and expertise.

Ethnographic observations suggested that medical wards were pressurised settings, often coping with multiple demands and limited resources.[47] The impact of SPI1 at the sharp end of medical wards was mostly difficult to detect, except in relation to improved monitoring of patients' vital signs and use of early warning score track-and-trigger systems to detect and respond to possible deterioration.

Despite the enthusiasm and support at the strategic level or "blunt end," medical ward staff, for the most part, tended either to know relatively little about SPI1 procedures, practices, and principles, or viewed them as handed down (top down) rather than something they were involved in developing (bottom up). There was little evidence that frontline medical ward staff perceived a sense of ownership over the initiative. The perception existed among some that SPI1 had allowed a small number of people to become an elite group with enhanced career prospects, while others were left feeling excluded.

> "SPI was a select group of twenty people . . . you're starting off in small areas and of course the by-product of that is that you've got a small group dealing with those small areas so there is, although we may not like it, there is a perception in some parts of the organisation that SPI is a, an elite entity."

The gap between the strategic level view and what was happening at the sharp end was evident in several different ways. For example, "leadership walk rounds" were discussed enthusiastically in the leadership focus groups, but some staff members at the sharp end thought that the process was disappointing and might even have undermined SPI1 because it sometimes seemed to show a failure to connect senior management with the wards in meaningful ways.

> "Well he came around and spoke to a few people and just asked about any concerns. He said he was interested to know how the nursing staff felt and he wanted to know one thing that he could take back to the rest of the board about any issues that nursing staff had. Afterwards they sent a letter to say thanks but you never hear any . . . well we haven't heard anything more than that so."

Several PDSA "success stories" were reported in the focus groups. Few of the frontline ward staff members interviewed, however, seemed aware of PDSA cycles. Somewhere between the blunt end and the sharp end, the model of participative engagement on which SPI1 was based had got rather lost, at least in relation to medical wards. There were several important influences that determined the extent to which SPI1 interventions became embedded on these wards. One example was intensity of the "dose" of SPI1 given to the wards. The activities of SPI1 change agents were dispersed throughout hospitals, and they might have focused their efforts on the more well defined clinical areas (such as the intensive care unit) that were the subject of clinical processes targeted by SPI1.

> "The ward aspect was one of the more difficult areas to implement because it wasn't as sort of specific, as confined, whereas the other areas perhaps were confined to theatres or ICU. [The ward] was much more . . . nebulous and covered a much bigger area . . . than perhaps other parts of the SPI."

A further issue was legitimacy. Sometimes staff simply did not see particular interventions as being scientifically legitimate.

> "Something that appears on the surface very simple, like the definition of a surgical infection, caused an absolute riot."

Scientific legitimacy debates were, perhaps paradoxically, exacerbated by the use of PDSA cycles. Some clinical staff members reported seeing the run-chart data collected during the cycles as unreliable and lacking in credibility and therefore as not providing enough of a prompt for change. Claims of problematic evidence, however, might have been used strategically as a means of resisting change and reinforcing inertia. Legitimacy issues also arose when staff members did not recognise that the problem being tackled was a "real" one requiring a particular response, or they did not consider that the resources that would be required to implement the intervention were legitimate in light of other priorities and pressures. We observed further barriers to adopting safety initiatives, including the instability of teams caused by rotating and agency staff, meaning it was difficult to sustain a collective knowledge and faith in SPI1 over time.

Positive impacts of SPI1 included increased managerial recognition and focus on patient safety and the promotion of a systematic approach. One of the major

**Table 8** | Medical history taking (% of patients asked required questions) before (epoch 1) and after (epoch 2) phase one of Safer Patients Initiative (SPI1) and effect of SPI. Figures are percentages (binomial standard errors (SE)) and odds ratios (99% confidence intervals) and P values for effect of SPI1

| | Control hospitals | | SPI1 hospitals | | |
|---|---|---|---|---|---|
| | Epoch 1 (n=236) | Epoch 2* (n=240) | Epoch 1 (n=381) | Epoch 2* (n=380) | OR (99% CI)†, P value |
| Duration of "presenting" symptom | 95 (2) | 95 (2) | 95 (1) | 96 (1) | 1.6 (0.4 to 6.9), 0.414 |
| Normal exercise tolerance | 33 (3) | 35 (3) | 33 (2) | 37 (3) | 1.5 (0.7 to 3.0), 0.178 |
| Presence/absence shortness of breath | 90 (2) | 92 (2) | 93 (1) | 94 (1) | 0.9 (0.3 to 3.1), 0.843 |
| Presence/absence orthopnoea | 28 (3) | 29 (3) | 24 (2) | 20 (2) | 1.0 (0.5 to 2.2), 0.966 |
| Presence/absence cough | 90 (2) | 90 (2) | 85 (2) | 89 (2) | 1.9 (0.7 to 5.3), 0.129 |
| If cough present, was it productive | 83 (3) | 86 (2) | 82 (2) | 88 (2) | 1.3 (0.5 to 3.3), 0.453 |
| History of smoking taken | 76 (3) | 80 (3) | 80 (2) | 82 (2) | 0.8 (0.4 to 1.9), 0.519 |
| Presence/absence of haemoptysis | 24 (3) | 26 (3) | 26 (2) | 27 (2) | 0.9 (0.4 to 1.9), 0.769 |
| Chest pain (of any type) | 61 (3) | 69 (3) | 75 (2) | 72 (2) | 0.7 (0.3 to 1.5), 0.230 |
| Occupation/previous occupation | 40 (3) | 38 (3) | 64 (3) | 64 (3) | 1.1 (0.6 to 2.2), 0.622 |
| Pets at home | 3 (1) | 3 (1) | 2 (1) | 1 (1) | 0.6 (0.1 to 6.1), 0.571 |
| % over all items | 57 | 59 | 60 | 61 | — |

*After intervention.
†OR >1 favours SPI1.

lessons learnt was the scale of resources and organisational support required to make patient safety initiatives work. There was a perception that hospitals had underestimated this and been too ready to assume that something that worked in a defined clinical area (such as intensive care units) would easily transfer to other environments or would spontaneously spread. Hospitals reported that they had begun to devise strategies for future implementation, which included engaging senior clinicians and encouraging local ownership.

### Quality of care/error rates in patients with acute respiratory disease on medical wards

*Explicit review of case notes*

The smallest SPI1 hospital in our sample could not provide the target numbers of case notes, leading to a slight shortfall in the intended SPI1 sample size of 400 sets of case notes in each epoch: 381 in epoch 1 and 380 in epoch 2. Control hospitals yielded 236 sets of case notes in epoch 1 and 240 sets in epoch 2. Baseline comparisons across all 1237 sets showed no significant differences between the control and SPI1 hospitals for any of the explicit review measures (table 4) assessed against the predetermined criteria.

Compliance in recording patients' observations at both six and 12 hours after admission improved markedly between the two epochs, and this was significant for all but one of eight possible items. Both control and SPI1 hospitals improved, though the improvement was greater in SPI1 hospitals (table 5); this difference in difference was significant only for the recording of respiratory rate at 12 hours. There was also a considerable increase over time in use of the CURB (confusion/urea/respiratory rate/blood pressure) score in patients with pneumonia, but these differences were not significant between control and SPI1 hospitals, and the point estimate favoured controls (table 6).

There were many prescribing errors, but these were mostly minor (table 7). There were no significant effects of SPI1 either over time or in favour of SPI in quality of prescribing (error rate ratio (estimated from population averaged negative binomial model) 1.2, 0.9 to 1.8; P=0.138).

The quality of observations, risk assessment, and prescribing were SPI1 targets (box 1), but we also sought evidence of a "halo effect" that might be evidence of a general strengthening of the system. There was no evidence of improvement in adherence to various tenets of safe evidence based practice (table 6) either over time or in SPI1 hospitals in particular. Compliance with certain standards (such as use of steroids when indicated) was already high (over 85%) at baseline in both sets of hospitals, leaving little room for further improvement. The quality of history taking remained stable over time in both control and SPI1 hospitals (table 8).

A comparison of the prescribing error results showed substantial[35] agreement between the two observers, with a κ of 0.71 and 0.70 in epochs 1 and 2, respectively. For some items—notably prescribing errors (table 7) and recording of exercise tolerance and chest pain (table 8)—the time in the sequence when the review of the case notes was conducted made a significant contribution to the recorded outcome. The propensity to record prescribing error first improved (taken as learning) and then declined (taken as fatigue). Adjustment for these effects was made in the analysis of these and all other items (see full report for further information on tests for homogeneity of end points among control and SPI1 hospitals, and effects of patient age on the quality of care[6]). Despite masking of all case notes in the hospital records department, and again at the Birmingham "centre" (see methods), the reviewer was able to discern the patient's name in 4%

**Table 9 | Mortality rates before (epoch 1) and after (epoch 2) phase one of Safer Patients Initiative (SPI1)**

| | Control hospitals | | SPI1 hospitals | |
|---|---|---|---|---|
| | Epoch 1 | Epoch 2* | Epoch 1 | Epoch 2* |
| No of Patients | 236 | 240 | 381 | 380 |
| Deaths | 27 | 39 | 63 | 49 |
| % Mortality (SE) | 11.4 (2.1) | 16.3 (2.4) | 16.5 (1.9) | 12.9 (1.7) |
| Mean (SD) age (years) | 77.6 (7.6) | 79.7 (7.7) | 77.4 (7.6) | 78.2 (8.0) |
| % women | 58.5 | 52.1 | 50.4 | 51.8 |
| Mean No of comorbidities | 2.8 | 3.1 | 3.3 | 3.8 |

*After intervention.

(42) of cases; hospital of origin in 1% (11) of cases; and epoch in 14% (158) of cases.

*Holistic review of case notes*

In the four SPI1 hospitals, 390 and 381 sets of case notes from epoch 1 and epoch 2, respectively, were holistically reviewed. For the 18 control hospitals, numbers were 243 and 246, respectively (range 8-15 cases per hospital). The total number of case notes (1260) is slightly different to the number in the explicit review because not all notes had been sent for both types of review when the study "closed." Both reviewers assessed 91 sets of case notes; measures of reliability between them were, as expected for holistic reviews,[30] low (κ=0.15, SE 0.08).

A total of 425 errors were identified. The most common categories of errors related to diagnosis, assessment, or admission or to poor clinical reasoning. From epoch 1 to 2, error rates declined from 44.4 to 42.3 in control hospitals and from 29.7 to 24.4 in the SPI1 hospitals. The difference in changes across control and SPI1 hospitals was not significant (rate ratio 0.87, 99% confidence interval 0.52 to 1.44; P=0.48).

**Box 3: Preventable adverse events identified as being strongly* or certainly preventable in the 1260 case notes reviewed in the holistic review**

**Epoch 1**

***Control hospitals***

- Patient given oxygen and became unrousable from $CO_2$ retention, required admission to intensive care unit

***SPI1 hospitals***

- Patient lost consciousness because of hypoglycaemia caused by overdose of insulin to control hyperkalaemia (patient died)
- Supraventricular tachycardia in patient with untreated hypokalaemia (patient died)
- Wrong choice of antibiotic for severe community acquired pneumonia (patient died)
- Patient's breathlessness deteriorated because nurses omitted scheduled use of nebuliser
- Patient sent home with severe uninvestigated anaemia†
- Patient started on treatment for hyperthyroidism, despite equivocal test result (and in wrong dose)
- Bronchospasm could have been avoided or lessened had β blocker been stopped

**Epoch 2**

***Control hospitals***

- Patient lost consciousness because of hypoglycaemia caused by overdose of insulin to control hyperkalaemia*
- Delay in administration of vitamin K leading to haematoma
- Patient's breathlessness increased, requiring transfer to high dependency unit, after failure to administer prescribed antibiotics

***SPI1 hospitals***

- Patient's collapse caused by adrenal crisis because corticosteroids were not prescribed for patient with known Addison's disease (patient died)
- Failure to treat MRSA; general practitioner not informed on discharge. No absolute evidence of harm but high risk
- Severe anaemia not investigated and general practitioner not informed†
- Bronchospasm could have been avoided or lessened had β blocker been stopped
- Failure to inform general practitioner of the risk of $CO_2$ retention†

*More likely than not (>50%) on balance of probabilities.
†No absolute evidence of harm in these cases but patients were discharged in clear danger and this influenced reviewer.

## Outcomes

*Adverse events*

The holistic review of the 1260 case notes identified 56 adverse events across all epochs and hospitals, giving an overall adverse event rate of about four per 100 patients treated. Agreement between raters in identification of adverse events was somewhat higher than agreement in identification of errors by holistic review (κ=0.25). The adverse event rate increased from 2.9 to 4.8 per 100 patients in control hospitals and declined from 6.2 to 3.7 among SPI1 hospitals. This difference in changes across control and SPI1 hospitals was not significant (rate ratio 0.40, 0.09 to 1.84; P=0.12).

In about a quarter (16) of the adverse events, there existed strong or certain evidence that the event was preventable (box 3). At around 1.3% (16/1260), this is a somewhat lower rate of preventable adverse events than reported for hospital inpatients elsewhere.[40]

*Mortality*

The mortality rate among patients admitted to medical wards was high, at over 14% (178 deaths across 1237 sets of case notes included in the explicit review), but there were no significant differences between control and SPI1 hospitals in the adjusted mortality rates. The baseline comparisons showed no significant differences between control and SPI1 hospitals nor was there significant evidence of temporal change in the control hospitals (table 9). When we adjusted figures for age, sex, and number of comorbidities in the analysis of mortality rates the odds ratios were 1.9 (99% confidence interval 0.6 to 5.6; P=0.149) for the baseline comparison, 1.4 (0.7 to 2.9; P=0.274) for the change in control hospitals, and 0.5 (0.2 to 1.4; P=085) for the effect of SPI1. There was no significant effect of sex or number of comorbidities on mortality, but the odds of death increased by 8% (5% to 11%) per year of patient age (P<0.001). Few of the deaths were attributable to error. In two cases, the reviewer thought that the death was definitely caused by the error (untreated documented hyperkalaemia, failure to recognise adrenal crisis) and in two further cases that it was more likely than not that death could have been averted (wrong choice of antibiotic, insulin overdose).

Table 10 | Patient survey scores* in control and SPI1 hospitals before (survey 1) and after (survey 2) phase one of Safer Patients Initiative (SPI1)

| | Control hospitals | | | | SPI1 hospitals | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No of patients | Survey 1 | Survey 2† | Absolute % change | No of patients | Survey 1 | Survey 2† | Absolute % change | Range at baseline | Effects of SPI1‡ (99% CI), P value |
| Overall, how would you rate the care you received? | 9209 | 79 | 77 | −2 | 2325 | 77 | 76 | −1 | 71-83 | 1 (−1 to 3), 0.330 |
| Overall, did you feel you were treated with respect and dignity while you were in hospital? | 9213 | 89 | 87 | −1 | 2370 | 88 | 86 | −2 | 84-91 | −1 (−2 to 1), 0.269 |
| How would you rate how well doctors and nurses worked together? | 9161 | 78 | 76 | −2 | 2360 | 76 | 75 | −1 | 72-82 | 1 (−1 to 3), 0.135 |
| In your opinion, how clean was hospital room or ward that you were in? | 9323 | 82 | 80 | −1 | 2395 | 79 | 79 | 0 | 71-89 | 1 (−1 to 4), 0.288 |
| How clean were toilets and bathrooms that you used in hospital? | 9147 | 79 | 77 | −2 | 2302 | 74 | 76 | 1 | 69-86 | 3 (0 to 6), 0.009 |

*Mean values of the five survey scores in control and SPI1 hospitals for first and second patient surveys, rated between 0-100.
†After intervention
‡Values >1 favour SPI1.

*Patients' satisfaction*

The response rate for the first survey was 54% (1961/3624) in the four SPI1 hospitals; for the second it was 51% (1720/3397). In the 18 control hospitals there was a greater fall, from 63% (9563/15274) to 56% (8590/15300). At baseline there were no significant differences between control and SPI1 hospitals on any of the scores. Only one of the patient satisfaction scores (cleanliness of bathrooms) showed a significantly different change (favouring SPI1) (table 10).

## DISCUSSION

### Main findings

Phase one of the Safer Patients Initiative (SPI1) was generally greeted enthusiastically at the strategic level (blunt end) in participating hospitals. Ethnographic observations and interviews, however, suggested that staff at the sharp end of medical wards generally had only a vague idea of the intervention, and that, with the exception of patient monitoring, few had direct experience of most of its components. Only one of the 11 dimensions on the staff survey changed significantly between groups; this occurred on an item relating to organisational climate, which favoured the SPI. Quantitative evaluation showed a significant effect in favour of SPI1 in respect of only one clinical process. This exception was in relation to the aim of improving observation of acutely ill patients, where recording of respiratory rate improved significantly more in SPI1 than in control hospitals. This quantitative finding was consistent with the qualitative observations.

During the study there was a general improvement in the quality of the monitoring of vital signs among both control and SPI1 hospitals; there was improvement in all eight of the required measurements of vital signs over the first 12 hours after admission (and this was significant in all but one case). The magnitude of the improvement in monitoring respiratory rate from before to after the intervention was greater than the magnitude of the difference between control and SPI1 hospitals. Likewise there was a sharp improvement over time in use of formal risk scoring for patients with pneumonia. These findings are all consistent with a general (temporal) improvement in the quality of monitoring for sick patients in the NHS over the study period.

We observed several clinical processes that were not specific SPI1 targets but might have been expected to improve if the overall goal of strengthening the system and achieving cultural and organisational realignments around safety had been achieved on medical wards. There was no significant difference between control and SPI1 hospitals over time. For some measures—such as use of corticosteroids in patients with chronic obstructive pulmonary disease and asthma—practice was already good at baseline and there was little room for further improvement. There was, however, also no change in other measures of quality of care, even when there was greater room for improvement. The holistic review corroborated findings from the criterion based explicit review, showing no consistent trends in either errors or adverse events in medical wards treating patients with acute respiratory disease in control versus SPI1 hospitals.

### Strengths and weaknesses

Our study had several strengths, including use of a predefined protocol, selected competitively, for evaluating SPI1. We quantified safety practices and used independent reviewers who made observations across multiple hospitals. Breaking the link between reviewer and hospital removes vested interest and ensures that any measurement error is systematic, thereby reducing risk of bias in comparisons between institutions. Reviewers were masked (as further insurance against observer bias), and we quality assured the masking process. We measured learning/fatigue effects (not just variation between observations)[30] to ensure that comparisons were not biased over time. Qualitative and quantitative observations across the different levels of hospitals enabled "triangulation"[48] of data collection and interpretation. We used a before and after controlled design and "difference in difference" approach to quantitative measurements.[15] With some notable exceptions[49,50] most quality improvement reports lack contemporaneous controls, thus denying

## WHAT IS ALREADY KNOWN ON THIS TOPIC

There is keen interest in finding ways of making healthcare safer for patients, and there are many examples of successful interventions targeted at specific clinical problems, such as hospital acquired infections

There are few formal summative evaluations of attempts to generate systematic organisation-wide improvements in the safety of care

Most evaluations of patient safety and quality improvement efforts lack contemporaneous controls

## WHAT THIS STUDY ADDS

Quality of monitoring of sick patients improved in control hospitals and in hospitals participating in the Safer Patients Initiative, but control hospitals did not improve as much as intervention hospitals; there was a small improvement in staff attitudes to organisational climate in intervention hospitals

On a range of other measures and outcomes related to patient safety, there was no additive effect attributable to the Safer Patients Initiative

Qualitative and quantitative findings of the evaluation of the Safer Patients Initiative were convergent

the ability to be confident in determining whether any changes can be attributed to the intervention.

These strengths mean that this evaluation adds to the science of evaluating large scale, complex quality improvement interventions that defy rigorous assessment with single method or single measure designs. Our approach dealt with the problems of data quality and other biases associated with using hospitals' own SPI1 run-charts, which are not designed for research purposes. A separate study has shown that data are not collected in a consistently defined way and tend to have missing values.[51]

One limitation of our study is that it was non-randomised. Results might have been biased against SPI1 because SPI1 hospitals might have had less room for improvement, and controls might have had higher than average performance, particularly as half were also selected as future SPI2 intervention sites. The staff survey and explicit review, however, showed similar performance at baseline, while adverse events and error rates detected on holistic review were equivocal (favouring the control hospitals in one domain and the SPI1 hospitals in another).

Results might have been biased in favour of SPI1 because SPI1 sites were selected, not chosen at random (the reverse of the possible bias mentioned above), and agreement to participate in the evaluation could have had a differentially motivating effect in SPI1 hospitals than in control hospitals (a form of the Hawthorne effect that could not be avoided by randomisation). These potential biases against controls would have been scientifically more worrisome had the results not been mostly null.

We thank Michael D L Morgan, Martyn R Partridge, and Philip W Ind for their expertise and contribution in the development of the forms for the explicit case note review; Janet Willars for conducting strategic/senior interviews and for helping to facilitate focus groups; Dale Webb, Louise Thomas, and Simona Arena for their insight into the implementation of SPI; Peter Chilton for his assistance in the preparation of this manuscript; and Frank Davidoff, Peter Pronovost, Tim Hofer, and M Clare Derrington for comments on the manuscript.

**Contributors:** AB, MD-W, JD, NB, and RL designed the study and submitted the grant proposal. RL was chief investigator. AB, NB, RL, MG, and BDF designed the forms and methods for explicit review of case notes. AB, RL, and UN designed the semistructured forms for holistic review of case notes and methods for data extraction. AB and UN were responsible for the case note review collection. MG and BDF conducted the explicit review of case notes, and AG analysed the data. MC and TN conducted the holistic review of case notes, and KH analysed the data. UN and MG designed the database for case note review. GR and AB created the queries for data extraction. MD-W supervised the study of senior staff and qualitative study. KH and SC performed quantitative analysis of the qualitative data from the stakeholder interviews. AS carried out the ethnographic fieldwork. MD-W and AS analysed the qualitative data. JD was responsible for all aspects of the staff and patient surveys. All authors contributed to the final manuscript. RL is guarantor.
**Funding:** This study was funded by the Health Foundation and NPSA and sponsored for research governance purposes by the University of Birmingham. The study was designed independently by the researchers and was awarded after competitive tender. The researchers acted independently but worked collaboratively with the funder. The researchers independently collected, analysed, and interpreted the data and independently wrote this article. The funders were given the opportunity to provide comments before submission. All researchers had the opportunity to access participants' anonymised data.
Competing interest: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare financial support for the study by the Health Foundation, additional funding for the explicit case note review in control hospitals and the holistic case note review study from the National Patient Safety Agency, funding for KH from the National Institute for Health Research Collaborations for Leadership in Applied Health Research and Care for Birmingham and Black Country, and AG by the Engineering and Physical Sciences Research Council, Multidisciplinary Assessment Technology Centre for Healthcare programme. The Centre for Medication Safety and Service Quality is affiliated with the Centre for Patient Safety and Service Quality at Imperial College Healthcare N HS Trust, which is funded by the National Institute of Health Research; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work
**Ethical approval:** Each substudy had its own ethical approval. The staff and patient surveys were approved by the North West multi-centre research ethics committee and each site granted access to their data. The stakeholder interviews, ethnography and case note review sub-studies were approved by Trent multi-centre research ethics committee. Informed consent was elicited from interview participants and focus groups. Informed consent was not required for the case note review as there was no direct contact with participants and notes were anonymised by participating hospitals. Local research governance was followed at each site.
**Data sharing:** no additional data available, but see full report.[6]

1 Department of Health Expert Group. An organisation with a memory. Stationery Office, 2000.
2 Kohn L, Corrigan J, Donaldson M. To err is human: building a safer health system. National Academies Press, 2000.
3 Leape LL, Berwick DM. Five years after to err is human: what have we learned? *JAMA* 2005;293:2384-90.
4 Hutter B. Regulation and risk: occupational health and safety on the railways. Oxford University Press, 2001.
5 Macrae C. Learning from patient safety incidents: creating participative risk regulation in healthcare. *Health Risk Soc* 2008;10:53-67.
6 Benning A, Ghaleb M, Suokas A, Dixon-Woods M, Dawson JF, Barber N, et al. Mixed-method evaluation of a large-scale organisational intervention to improve patient safety in four UK hospitals. 2010. www.haps.bham.ac.uk/publichealth/psrp/EvalSPI.shtml.
7 Deming WE. Out of the crisis. MIT Press, 2000.
8 Dixon-Woods M, Tarrant C, Willars J, Suokas A. How will it work? A qualitative study of strategic stakeholders' accounts of a patient safety initiative. *Qual Saf Health Care* 2010;19:74-8.
9 Varkey P, Reller MK, Resar RK. Basics of quality improvement in health care. *Mayo Clin Proc* 2007;82:735-9.
10 Institute for Healthcare Improvement. 17th Annual National Forum on Quality Improvement in Health Care. 2005 December 11-14; Orlando, USA. www.ihi.org/Ihi/Files/Forum/2005/Handouts/IHI_033_ATT_GUIDE_fnl_2.pdf.

11 Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100 000 lives campaign: setting a goal and a deadline for improving health care quality. *JAMA* 2006;295:324-7.
12 Bisognano C, McCannon J, Botwinick L. A campaign for 100,000 lives. *Trustee* 2005;58:12-4.
13 Ogrinc G, Mooney SE, Estrada C, Foster T, Goldmann D, Hall LW, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care* 2008;17(suppl 1):i13-32.
14 Benning A, Dixon-Woods M, Nwulu U, Ghaleb M, Dawson J, Barber N, et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;doi:10.1136/bmj.d199.
15 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 4. One size does not fit all. *Qual Saf Health Care* 2008;17:178-81.
16 Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. *BMJ* 2008;337:a2764.
17 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 1. Conceptualising and developing interventions. *Qual Saf Health Care* 2008;17:158-62.
18 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 2. Study design. *Qual Saf Health Care* 2008;17:163-9.
19 Brown C, Hofer T, Johal A, Thomson R, Nicholl J, Franklin BD, et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 3. End points and measurement. *Qual Saf Health Care* 2008;17:170-7.
20 Lilford RJ, Chilton PJ, Hemming K, Girling AJ, Taylor CA, Barach P. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. *BMJ* 2010;341:c4413.
21 Glaser B, Thomson R. The discovery of grounded theory: Strategies for qualitative research. Aldine Publishing Company, 1967.
22 Guzzo RA, Jette D, Katzell RA. The effects of psychologically based intervention programs on worker productivity: a meta-analysis. *Pers Psychol* 1985;38:275-92.
23 West MA, Guthrie JP, Dawson JF, Borrill CS, Carter MR. Reducing patient mortality in hospitals: the role of human resource management. *J Organ Behav* 2006;27:983-1002.
24 Borrill CS, West MA, Shapiro D, Rees A. Team working and effectiveness in health. *Brit J Healthc Manag* 2000;6:364-71.
25 Michie S, West M. Managing people and performance: an evidence-based framework applied to health service organisations. *Int J Manag Rev* 2004;5-6:91-111.
26 Healthcare Commission. Making sense of your staff survey data. 2006. http://www.carequalitycommission.org.uk/_db/_documents/Making_sense_of_your_Staff_Survey_Data_200703214701.pdf.
27 Smith GB, Prytherch DR, Schmidt P, Featherstone PI, Knight D, Clements G, et al. Hospital-wide physiological surveillance-a new approach to the early identification and management of the sick patient. *Resuscitation* 2006;71:19-28.
28 Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Antecedents to hospital deaths. *Intern Med J* 2001;31:343-8.
29 McQuillan P, Pilkington S, Allan A, Taylor B, Short A, Morgan G, et al. Confidential inquiry into quality of care before admission to intensive care. *BMJ* 1998;316:1853-8.
30 Lilford R, Edwards A, Girling A, Hofer T, Di Tanna GL, Petty J, et al. Inter-rater reliability of case-note audit: a systematic review. *J Health Serv Res Policy* 2007;12:173-80.
31 BTS Guidelines for the management of community acquired pneumonia in adults. *Thorax* 2001;56:iv1-64.
32 BTS Guidelines for the management of community acquired pneumonia in adults, 2004 update. 2004. www.brit-thoracic.org.uk/Portals/0/Clinical%20Information/Pneumonia/Guidelines/MACAPrevisedApr04.pdf.
33 Barber N, Franklin BD, Cornford T, Klecun E, Savage I. Safer, faster, better? Evaluating electronic prescribing. Report to the Patient Safety Research Programme. (Policy Research Programme of the Department of Health). 2006. www.haps.bham.ac.uk/publichealth/psrp/documents/PS019_Final_Report_Barber.pdf.
34 Cousins D, Hatoum H. The development of a computerised quality assurance system for clinical pharmacy. *Int J Pharmact Pract* 1991;1:86-9.
35 Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
36 Reason JT. Human error. Cambridge University Press, 1990.
37 Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991;324:377-84.
38 Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The quality in Australian health care study. *Med J Aust* 1995;163:458-71.
39 Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard medical practice study I. *N Engl J Med* 1991;324:370-6.
40 Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* 2001;322:517-9.
41 Woloshynowych M, Neale G, Vincent C. Case record review of adverse events: a new approach. *Qual Saf Health Care* 2003;12:411-5.
42 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.
43 Hackbarth A, McCannon C, Martin L, Lloyd R, Calkins D. The hard count: calculating lives saved in the 100,000 lives campaign. 2006. www.ihi.org/NR/rdonlyres/AA1B05AD-5DFF-45D6-8A2F-05C69C9C2860/0/ihiarticle.pdf.
44 Warrack BJ, Sinha MN. Integrating safety and quality: building to achieve excellence in the workplace. *Total Qual Manag* 1999;10:779-85.
45 Murray DM, Blitstein JL. Methods to reduce the impact of intraclass correlation in group-randomized trials. *Eval Rev* 2003;27:79-103.
46 Cook R, Woods D. Operating at the sharp end: the complexity of human error. In: Bogner M, ed. Human error in medicine. Lawrence Erlbaum, 1994:255-310.
47 Dixon-Woods M, Suokas A, Pitchforth E, Tarrant C. An ethnographic study of classifying and accounting for risk at the sharp end of medical wards. *Soc Sci Med* 2009;69:362-9.
48 Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998;2:iii-274.
49 Landon BE, Hicks LS, O'Malley AJ, Lieu TA, Keegan T, McNeil BJ, et al. Improving the management of chronic disease at community health centers. *N Engl J Med* 2007;356:921-34.
50 Landon BE, Wilson IB, McInnes K, Landrum MB, Hirschhorn L, Marsden PV, et al. Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection: the EQHIV study. *Ann Intern Med* 2004;140:887-96.
51 Benn J, Burnett S, Parand A, Pinto A, Iskander A, Vincent C. Studying large-scale programmes to improve patient safety in whole care systems: challenges for research. *Soc Sci Med*. 2009;69:1767-76.


**Accepted:** 12 October 2010